



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**PENGEMBANGAN SISTEM ANALITIK DATA
BERBASIS *MULTI-AGENT* DENGAN PENDEKATAN
FUNCTION CALLING DAN *METADATA-CENTRIC*
*CONTEXT MANAGEMENT***

SKRIPSI

MUHAMMAD BINTANG RAMADHAN 2207411055

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN
KOMPUTER
POLITEKNIK NEGERI JAKARTA
2026**



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**PENGEMBANGAN SISTEM ANALITIK DATA
BERBASIS *MULTI-AGENT* DENGAN PENDEKATAN
FUNCTION CALLING DAN *METADATA-CENTRIC*
*CONTEXT MANAGEMENT***

SKRIPSI

**Dibuat untuk Melengkapi Syarat-Syarat yang Diperlukan
untuk Memperoleh Diploma Empat Politeknik**

MUHAMMAD BINTANG RAMADHAN

2207411055

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN
KOMPUTER**

POLITEKNIK NEGERI JAKARTA

2026



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

SURAT PERNYATAAN BEBAS PLAGIARISME

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Bintang Ramadhan
NIM : 2207411055
Jurusan/Program Studi : Teknik Informatika dan Komputer/Teknik Informatika
Judul Skripsi : Pengembangan Sistem Analitik Data Berbasis *Multi-Agent* Dengan Pendekatan *Function Calling* dan *Metadata-Centric Context Management*

Menyatakan dengan sebenarnya bahwa skripsi ini benar-benar merupakan hasil karya saya sendiri, bebas dari peniruan terhadap karya dari orang lain. Kutipan pendapat dan tulisan orang lain ditunjuk sesuai dengan cara-cara penulisan karya ilmiah yang berlaku.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa dalam skripsi ini terkandung ciri-ciri plagiat dan bentuk-bentuk peniruan lain yang dianggap melanggar peraturan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Jakarta, 9 Juli 2026

Yang membuat pernyataan



Muhammad Bintang Ramadhan

2207411055



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

LEMBAR PENGESAHAN

Skripsi diajukan oleh :

Nama : Muhammad Bintang Ramadhan
NIM : 2207411055
Program Studi : Teknik Informatika
Judul Skripsi : Pengembangan Sistem Analitik Data Berbasis
Multi-Agent Dengan Pendekatan Function Calling
dan Metadata-Centric Context Management

Telah diuji oleh tim penguji dalam Sidang Skripsi pada hari Senin, Tanggal 22,
Bulan Juni, Tahun 2026 dan dinyatakan **LULUS**.

Disahkan oleh

Pembimbing I : Dr. Dewi Yanti Liliana, S.Kom., M.Kom.

Penguji I : Mera Kartika Delimayanti, S.Si., M.T., Ph.D.

Penguji II : Euis Oktavianti, S.Si., M.T.I.

Penguji III : Ariawan Andi Suhandana, S.Kom., M.T.I.

Mengetahui:

Jurusan Teknik Informatika dan Komputer

Ketua



Dr. Anita Hidayati, S.Kom., M.Kom.

NIP. 197908032003122003



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah SWT atas segala rahmat dan karunia-Nya sehingga skripsi yang berjudul “Pengembangan Sistem Analitik Data Berbasis *Multi-Agent* Dengan Pendekatan *Function Calling* dan *Metadata-Centric Context Management*” ini dapat diselesaikan sebagai salah satu syarat memperoleh gelar Sarjana. Pada kesempatan ini, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

- a. Ibu Dr. Dewi Yanti Liliana, S.Kom., M.Kom. selaku dosen pembimbing yang telah meluangkan waktu serta memberikan bimbingan, arahan, dan motivasi selama proses penyusunan skripsi ini.
- b. Kedua orang tua, kakak, dan adik tercinta yang senantiasa memberikan doa, kasih sayang, dan dukungan yang tiada henti.
- c. Sahabat dan teman-teman yang telah memberikan semangat serta kebersamaan yang sangat berarti selama masa studi.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Kritik dan saran yang membangun sangat penulis harapkan. Semoga skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang sistem analitik data dan kecerdasan buatan.

Jakarta, 9 Juli 2026

Muhammad Bintang Ramadhan



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Politeknik Negeri Jakarta, saya bertanda tangan dibawah ini:

Nama : Muhammad Bintang Ramadhan
NIM : 2207411055
Jurusan/Program Studi : Teknik Informatika dan Komputer/Teknik Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Politeknik Negeri Jakarta Hak Bebas Royalti Non-Eksklusif atas karya ilmiah saya yang berjudul :

Pengembangan Sistem Analitik Data Berbasis *Multi-Agent* Dengan Pendekatan *Function Calling* dan *Metadata-Centric Context Management*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksklusif ini Politeknik Negeri Jakarta berhak menyimpan, mengalihmediakan/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 7 Juni 2026

Yang membuat pernyataan



Muhammad Bintang Ramadhan

2207411055



PENGEMBANGAN SISTEM ANALITIK DATA BERBASIS *MULTI-AGENT* DENGAN PENDEKATAN *FUNCTION CALLING* DAN *METADATA-CENTRIC CONTEXT MANAGEMENT*

ABSTRAK

Analisis data menggunakan tools konvensional menghadapi tantangan kompleksitas, waktu, dan kebutuhan pemahaman teknis yang tinggi. Pemanfaatan Large Language Model (LLM) menawarkan solusi berbasis natural language, namun terkendala keterbatasan context window yang menyebabkan degradasi reasoning pada dataset berskala besar. Penelitian ini mengembangkan sistem analitik data berbasis arsitektur multi-agent yang mengintegrasikan function calling dan metadata-centric context management untuk mengatasi permasalahan tersebut. Arsitektur yang dibangun membagi proses analisis ke dalam tiga fase analisis, yaitu Data Preprocessing, EDA, dan Explainer, yang dikoordinasikan melalui Google Agent Development Kit (ADK). Sistem tidak memasukkan data raw ke dalam context window, melainkan memanfaatkan ringkasan metadata statistik sebagai konteks utama reasoning LLM. Pengujian dilakukan menggunakan dataset penjualan sintetis berukuran 100 ribu hingga 10 juta baris dalam kondisi clean dan dirty, serta dataset NvBench sebagai baseline komparatif terhadap data analyst profesional, dengan model LLM Xiaomi MiMo V2 Flash dan OpenAI GPT OSS 120B. Hasil pengujian menunjukkan penggunaan token hanya 0,014%–1,8% dari jumlah token raw dataset dengan akurasi kalkulasi 100% pada dataset clean. Sistem menghasilkan output yang deterministik pada data clean, sementara data dirty menghasilkan konsistensi yang bervariasi antar model.

Kata Kunci: *agentic AI, data analytic, function calling, large language model, metadata-centric context management*

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



DAFTAR ISI

KATA PENGANTAR	4
ABSTRAK	6
DAFTAR ISI	7
DAFTAR GAMBAR	9
DAFTAR TABEL	11
BAB I PENDAHULUAN	13
1.1 Latar Belakang	13
1.2 Rumusan Masalah	14
1.3 Batasan Masalah	14
1.4 Tujuan dan Manfaat	15
1.4.1 Tujuan Penelitian	15
1.4.2 Manfaat Penelitian	15
1.5 Sistematika Penulisan	15
BAB II TINJAUAN PUSTAKA	17
2.1 Analitik Data Konvensional dan Jenis-Jenis Proses Analitik	17
2.2 Analitik Data dan Data-Driven Decision Making	17
2.3 Large Language Model dalam Sistem Analitik Data	18
2.4 Keterbatasan Context Window dan Penurunan Kemampuan Reasoning pada LLM	19
2.5 Agentic AI dan Multi-Agent Systems dalam Analitik Data	20
2.6 Function Calling sebagai Mekanisme Interaksi LLM	21
2.7 Metadata Statistik dan Data Profiling dalam Analitik Data	22
2.8 Metadata-Centric Context Management Berbasis Metadata Statistik	22
2.9 Evaluasi Sistem Agentic AI	23
2.10 Penelitian Sejenis	24
BAB III METODE PENELITIAN	27
3.1 Rancangan Penelitian	27
3.1.1 Identifikasi Masalah	28
3.1.2 Studi Literatur	28
3.1.3 Perancangan Sistem	28
3.1.4 Implementasi Sistem	29
3.1.5 Pengujian Sistem	29
3.1.6 Evaluasi Hasil	29
3.2 Sumber Data	29
3.3 Teknik Pengumpulan Data	30



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

3.4 Teknik Analisis Data.....	31
BAB IV HASIL DAN PEMBAHASAN	33
4.1 Analisis Kebutuhan	33
4.1.1 Kebutuhan Sistem Multi-Agent	33
4.1.2 Kebutuhan Pengembangan Web (Dashboard)	35
4.1.3 Kebutuhan Perangkat Keras	36
4.1.4 Kebutuhan Perangkat Lunak	37
4.2 Perancangan Sistem	38
4.2.1 Perancangan Sistem Multi-Agent	38
4.2.2 Perancangan Web Dashboard	43
4.3 Implementasi Sistem	62
4.3.1 Implementasi Sistem Multi-Agent	62
4.3.2 Implementasi Web Dashboard	74
4.4 Pengujian.....	81
4.4.1 Deskripsi Pengujian	81
4.4.2 Prosedur Pengujian	92
4.4.3 Data Hasil Pengujian.....	93
4.4.4 Analisis Data / Evaluasi Pengujian	108
BAB V PENUTUP.....	131
5.1 Simpulan	131
5.2 Saran.....	132
DAFTAR PUSTAKA	134
LAMPIRAN	137

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian	29
Gambar 4.1 Arsitektur Agentic AI.....	40
Gambar 4.2 Flow Metadata-centric Context Management	42
Gambar 4.3 Flow Function Calling.....	43
Gambar 4.4 Use Case Diagram	45
Gambar 4.5 Activity Diagram Sign Up.....	46
Gambar 4.6 Activity Diagram Sign In	47
Gambar 4.7 Activity Diagram Membuat Proyek Analisis	47
Gambar 4.8 Activity Diagram Mengubah Dataset.....	48
Gambar 4.9 Activity Diagram Mengubah Model LLM.....	49
Gambar 4.10 Activity Diagram Membatalkan Proses Analisis	49
Gambar 4.11 Activity Diagram Mengubah Nama Proyek Analisis.....	50
Gambar 4.12 Activity Diagram Menghapus Proyek Analisis.....	51
Gambar 4.13 Activity Diagram Melihat Proyek Analisis.....	51
Gambar 4.14 Activity Diagram Memilih Menu di Daftar Isi	52
Gambar 4.15 Activity Diagram Melihat System Logs	52
Gambar 4.16 Activity Diagram Melihat System Logs Output Details	52
Gambar 4.17 Activity Diagram Melakukan Analisis Ulang.....	53
Gambar 4.18 Activity Diagram Kembali ke Halaman Utama	53
Gambar 4.19 Activity Diagram Melihat Halaman Daftar Proyek	54
Gambar 4.20 Activity Diagram Mengubah Mode Tampilan List Proyek	54
Gambar 4.21 Activity Diagram Mencari Proyek Analisis	55
Gambar 4.22 Entity Relationship Diagram.....	56
Gambar 4.23 Wireframe Halaman Sign In	58
Gambar 4.24 Wireframe Halaman Sign Up.....	58
Gambar 4.25 Wireframe Halaman Home	59
Gambar 4.26 Wireframe Halaman Daftar Proyek Grid	59
Gambar 4.27 Wireframe Halaman Daftar Proyek List	60
Gambar 4.28 Wireframe Modal New Analysis.....	60
Gambar 4.29 Wireframe Halaman Processing.....	61
Gambar 4.30 Wireframe Tabel Logs Halaman Processing.....	61
Gambar 4.31 Wireframe Halaman Laporan Analitik.....	62
Gambar 4.32 Wireframe Tabel Logs Halaman Laporan Analitik.....	62
Gambar 4.33 Wireframe Details Tabel Logs	63
Gambar 4.34 Arsitektur Agentic AI Fase Preprocessing	64



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Gambar 4.35 Arsitektur Agentic AI Fase EDA	67
Gambar 4.36 Arsitektur Agentic AI Fase Explainer	68
Gambar 4.37 Flowchart Metadata-Centric Context Management	69
Gambar 4.38 Proses Schema Extraction pada Tool	70
Gambar 4.39 Proses Descriptive Analysis pada Tool	70
Gambar 4.40 Proses Sampling pada Tool	70
Gambar 4.41 Sequence Diagram Mekanisme Function Calling	72
Gambar 4.42 Flow Eksekusi dengan DuckDB	73
Gambar 4.43 Flow Integrasi Function Calling dan DuckDB	74
Gambar 4.44 Halaman Sign In	75
Gambar 4.45 Halaman Sign Up	76
Gambar 4.46 Halaman Utama (Home)	77
Gambar 4.47 Halaman Daftar Proyek Grid View	77
Gambar 4.48 Halaman Daftar Proyek List View	78
Gambar 4.49 Modal New Analysis	78
Gambar 4.50 Halaman Processing	79
Gambar 4.51 Tabel System Logs Pada Halaman Processing	80
Gambar 4.52 Halaman Laporan Analisis	80
Gambar 4.53 Tabel System Logs Pada Halaman Laporan Analisis	81
Gambar 4.54 Modal Log Details Pada Tabel System Logs	82
Gambar 4.55 Konfigurasi Script untuk Pembuatan Dummy Dataset	84
Gambar 4.56 Python Script untuk Noise Injection	85
Gambar 4.57 Python Script untuk Pembuatan Dummy Dataset	85
Gambar 4.58 Python Script untuk Membuat Data Ground Truth	86
Gambar 4.59 Python Script untuk Menghitung Jumlah Karakter dan Token	87
Gambar 4.60 Proses Null Value Imputation Pada Skenario Dataset NB-3	104
Gambar 4.61 Grafik Penggunaan Token dibanding Ukuran Dataset	110
Gambar 4.62 Grafik Rata-Rata Penggunaan Token per Model LLM	111
Gambar 4.63 Penggunaan token pada kondisi Clean dan Dirty	113
Gambar 4.64 Grafik Waktu Pemrosesan terhadap Ukuran Dataset	114
Gambar 4.65 Waktu Pemrosesan pada kondisi Clean dan Dirty	115
Gambar 4.66 Grafik Rata Rata Self-Correction pada Setiap Ukuran Dataset	116
Gambar 4.67 Grafik Jumlah Visualisasi yang di-generate dan yang di-embed	117
Gambar 4.68 Grafik Rata-Rata Waktu Pemrosesan pada Setiap Model	122
Gambar 4.69 Grafik Rata-Rata Gap antar Proses pada Setiap Model	123
Gambar 4.70 Grafik Volume Function Calling pada Skenario Pengujian	124
Gambar 4.71 Grafik Rata-Rata Function Calling pada Setiap Model	124
Gambar 4.72 Grafik Self-Correction pada Setiap Model	125
Gambar 4.73 Grafik Perbandingan Model LLM dari segi Token Usage	126



Gambar 4.74 Grafik Attachment Rate pada Model LLM	127
Gambar 4.75 Grafik Attach vs Discard Visualisasi pada Setiap Model LLM.....	128

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta





DAFTAR TABEL

Tabel 2.1 Penelitian Sejenis	2
Tabel 4.1 Kebutuhan Dataset	34
Tabel 4.2 Kebutuhan Large Language Model (LLM)	35
Tabel 4.3 Kebutuhan Kemampuan Agen	35
Tabel 4.4 Kebutuhan Fungsional	36
Tabel 4.5 Kebutuhan Non-Fungsional	37
Tabel 4.6 Kebutuhan Perangkat Keras	38
Tabel 4.7 Kebutuhan Perangkat Lunak	38
Tabel 4.8 Tabel Daftar Tools untuk Function Calling	72
Tabel 4.9 Spesifikasi Perangkat Environment Pengujian	83
Tabel 4.10 Daftar Business Questions untuk Pengujian	83
Tabel 4.11 Hasil Perhitungan Karakter dan Token	87
Tabel 4.12 Skenario Pengujian Skalabilitas dan Waktu Pemrosesan	88
Tabel 4.13 Metrik Pengujian Hasil Laporan Analisis	89
Tabel 4.14 Daftar Dataset dari “Is GPT-4 a Good Data Analyst”	90
Tabel 4.15 Skenario Pengujian Insight	90
Tabel 4.16 Skenario Pengujian Konsistensi	92
Tabel 4.17 Hasil Pengujian Black Box per Fungsi	94
Tabel 4.18 Hasil Pengujian Skalabilitas dan Waktu Pemrosesan	95
Tabel 4.19 Hasil Pengujian Kriteria Function Calling	96
Tabel 4.20 Hasil Pengujian Kriteria Visualisasi	97
Tabel 4.21 Hasil Pengujian Kriteria Self-Correction	98
Tabel 4.22 Hasil Pengujian Kriteria Ground Truth	99
Tabel 4.23 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-1	101
Tabel 4.24 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-2	102
Tabel 4.25 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-3	103
Tabel 4.26 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-4	104
Tabel 4.27 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-5	105
Tabel 4.28 Hasil Pengujian Kriteria Validasi Embedding Visualisasi	106
Tabel 4.29 Hasil Pengujian Kriteria Konsistensi pada Dataset Clean	107
Tabel 4.30 Hasil Pengujian Konsistensi Dataset Dirty KO-2-1	108
Tabel 4.31 Hasil Pengujian Konsistensi Dataset Dirty KO-2-2	109
Tabel 4.32 Perbandingan Token actual vs raw	111
Tabel 4.33 Nilai Korelasi terhadap Penggunaan Token	112
Tabel 4.34 Nilai Korelasi terhadap Waktu Pemrosesan	114



Tabel 4.35 Log Data Cleaning Model Xiaomi MiMo V2 Flash	118
Tabel 4.36 Log Data Cleaning Model GPT OSS 120B	119
Tabel 4.37 Perbandingan Sistem dengan Baseline Analisis Profesional.....	120
Tabel 4.38 Perbandingan Strategi Data Cleaning Antar Model.....	130



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

BAB I PENDAHULUAN

1.1 Latar Belakang

Dalam dunia bisnis saat ini, data analisis berperan penting dalam mendukung pengambilan keputusan berbasis data. Namun, proses analisis data yang digunakan saat ini masih cenderung kompleks, memakan waktu, dan sangat bergantung pada pemahaman teknis. Sebagian besar waktu analisis dihabiskan untuk proses manual seperti *data cleaning* dan *data preparation*, sehingga proses pengambilan keputusan menjadi lambat dan sulit dilakukan oleh pihak non-teknis (Begom Riipa *et al.*, 2025; Li *et al.*, 2025).

Proses analitik data pada saat ini mulai memanfaatkan *Large Language Models* (LLMs) untuk membantu proses analisis melalui bahasa alami. Penelitian terkait menunjukkan bahwa LLM mampu melakukan analisis data dengan kualitas yang mendekati analisis profesional (Cheng, Li and Bing, 2023). Namun, dalam praktiknya, sistem analitik berbasis LLM sering mengalami permasalahan ketika digunakan pada dataset berukuran besar. Model cenderung menghasilkan analisis yang tidak konsisten atau tidak sesuai dengan kondisi data, terutama ketika memproses dataset dalam jumlah besar secara langsung (Woesle, Fischer-Brandies and Buettner, 2025).

Permasalahan tersebut berkaitan dengan keterbatasan ukuran konteks yang dapat diproses oleh LLM secara optimal. Literatur menunjukkan penambahan konteks data yang terlalu besar justru menurunkan kemampuan model dalam melakukan penalaran, meskipun informasi yang dibutuhkan tersedia dalam konteks tersebut (Du *et al.*, 2025). Akibatnya, sistem analitik berbasis LLM menjadi tidak skalabel dan memiliki performa yang kurang optimal untuk analisis data berskala besar.

Sebagian besar sistem analitik berbasis LLM yang ada belum memiliki mekanisme manajemen konteks yang terstruktur dan efisien. Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan pengembangan sistem analitik data berbasis *multi-agent* yang memanfaatkan metadata statistik sebagai ringkasan informasi data utama, serta mekanisme *function calling* untuk menghubungkan model dengan



proses analitik secara terkontrol (Li *et al.*, 2025; Sapkota, Roumeliotis and Karkee, 2026). Pendekatan ini diharapkan mampu mengurangi beban konteks pada model, meningkatkan performa analisis, serta mempermudah pemanfaatan analitik data bagi pengguna tanpa pemahaman teknis.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, perumusan masalah dalam penelitian ini dinyatakan dalam bentuk pertanyaan sebagai berikut:

1. Bagaimana merancang arsitektur sistem analitik data berbasis *multi-agent* yang mampu mengkoordinasikan proses analisis data secara terstruktur dan otonom menggunakan *Large Language Model*.
2. Bagaimana penerapan pendekatan *function calling* dan *metadata-centric context management* dalam sistem analitik data berbasis *multi-agent* untuk mengatasi keterbatasan *context window* pada analisis dataset berskala besar.
3. Bagaimana performa sistem analitik data berbasis *multi-agent* yang dikembangkan ditinjau dari efektivitas, skalabilitas, dan kualitas *insight* yang dihasilkan pada kasus data penjualan.

1.3 Batasan Masalah

Agar penelitian ini tetap terfokus dan memiliki ruang lingkup yang jelas, maka batasan masalah dalam penelitian ini ditetapkan sebagai berikut:

1. Interaksi antara *Large Language Model* dan sistem eksternal dibatasi pada mekanisme *function calling*, sehingga penelitian ini tidak membahas pendekatan *python code generation* maupun *python code execution* yang dihasilkan oleh model.
2. Strategi *context management* dalam sistem dibatasi pada penggunaan metadata statistik dataset yang diperoleh melalui proses *data profiling* otomatis, seperti struktur kolom, statistik deskriptif, distribusi data, *missing values*, dan duplikasi data.
3. Metadata yang digunakan sebagai konteks utama sistem tidak mencakup metadata semantik eksternal maupun *knowledge graph* berbasis pengetahuan bisnis.
4. Evaluasi sistem dilakukan menggunakan dua jenis dataset, yaitu dataset sintetis penjualan untuk pengujian akurasi, skalabilitas, dan konsistensi; serta dataset



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

NvBench untuk perbandingan output dengan *baseline* analisis profesional. Generalisasi sistem terhadap domain di luar kedua dataset tersebut tidak diuji dalam penelitian ini.

5. Penelitian ini tidak mencakup proses pelatihan (*training*) maupun *fine-tuning* *Large Language Model*, melainkan menggunakan model yang telah tersedia sebagai komponen *reasoning* dalam sistem.
6. LLM yang digunakan dalam penelitian ini adalah *Xiaomi MiMo V2 Flash* dan *OpenAI GPT OSS 120B*.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat, tujuan dari penelitian ini adalah:

1. Merancang dan mengimplementasikan arsitektur sistem analitik data berbasis *multi-agent* yang memanfaatkan *Large Language Model* untuk mengotomatisasi proses analisis data secara terstruktur.
2. Menerapkan pendekatan *function calling* dan *metadata-centric context management* berbasis metadata statistik dataset guna mengatasi keterbatasan *context window* pada analisis data berskala besar.
3. Mengevaluasi efektivitas, skalabilitas, dan kualitas *insight* dari sistem analitik data berbasis *multi-agent* yang dikembangkan pada studi kasus data penjualan.

1.4.2 Manfaat Penelitian

Manfaat yang diharapkan dari pelaksanaan penelitian ini adalah sebagai berikut:

1. Memberikan solusi sistem analitik data yang lebih terstruktur, reliabel, dan *scalable* dengan memanfaatkan arsitektur *multi-agent* berbasis LLM.
2. Mengurangi risiko kesalahan analisis dan *hallucinations* melalui penggunaan metadata statistik sebagai *context anchor* dalam proses *reasoning* dan pengambilan keputusan analitik.
3. Mendukung proses pengambilan keputusan berbasis data secara lebih efektif dan akurat melalui sistem analitik otomatis dengan bantuan LLM.
4. Menjadi referensi dan acuan pengembangan sistem analitik data berbasis *agentic AI* pada penelitian atau implementasi selanjutnya.

1.5 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini disusun sebagai berikut:

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Bab I Pendahuluan

Berisi latar belakang penelitian terkait *Agentic AI* untuk Analitik Data, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

Bab II Tinjauan Pustaka

Membahas landasan teori dan penelitian terdahulu yang relevan terkait analitik data, *Large Language Model*, *agentic AI*, dan *multi-agent systems*.

Bab III Metode Penelitian

Menjelaskan metode penelitian, arsitektur sistem, desain *multi-agent*, alur kerja sistem, serta teknik evaluasi yang digunakan.

Bab IV Implementasi dan Pembahasan

Menguraikan proses implementasi sistem *Agentic AI*, studi kasus pada dataset penjualan (*sales*), hasil pengujian, serta analisis terhadap performa dan keterbatasan sistem.

Bab V Kesimpulan dan Saran

Memberikan kesimpulan dari hasil penelitian *Agentic AI* untuk Analitik Data serta saran untuk pengembangan dan penelitian lanjutan di masa mendatang.

**POLITEKNIK
NEGERI
JAKARTA**



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

BAB II

TINJAUAN PUSTAKA

2.1 Analitik Data Konvensional dan Jenis-Jenis Proses Analitik

Analitik data konvensional merujuk pada serangkaian proses sistematis yang dilakukan secara manual untuk mengekstrak informasi bermakna dari kumpulan data (Prasanth *et al.*, 2023). Proses ini diklasifikasikan ke dalam empat jenis utama, yaitu *descriptive analytics* yang merangkum kondisi historis data untuk menjawab “*what happened in the past?*”, *diagnostic analytics* yang mencari hubungan antar variabel untuk menjawab “*why did it happen?*”, *predictive analytics* yang membangun model statistik untuk memprediksi kondisi di masa depan, serta *prescriptive analytics* yang memberikan rekomendasi tindakan paling optimal berdasarkan hasil prediksi (Sarker, 2021).

Dalam praktiknya, keempat jenis analitik tersebut dilakukan melalui serangkaian tahapan secara berurutan, mulai dari pengumpulan data, *data cleaning* untuk menangani *missing values* dan inkonsistensi, *Exploratory Data Analysis* (EDA) untuk memahami distribusi dan pola awal, hingga transformasi dan *feature engineering* sebagai persiapan pemodelan (Chicco *et al.*, 2022; Radhi *et al.*, 2021). Seluruh tahapan ini dikerjakan secara manual menggunakan kombinasi *spreadsheet*, bahasa pemrograman, dan *tools* visualisasi, sehingga membutuhkan keahlian teknis yang tinggi dan waktu yang lama (Bilal *et al.*, 2022). Kondisi ini menjadi tantangan utama ketika berhadapan dengan dataset berskala besar, sekaligus menjadi landasan motivasi bagi pengembangan sistem analitik yang lebih terotomatisasi sebagaimana dibahas pada subbab-subbab berikutnya.

2.2 Analitik Data dan *Data-Driven Decision Making*

Analitik Data memegang peran penting dalam bisnis modern sebagai dasar pengambilan keputusan. Pendekatan *data-driven decision making* (DDDM) adalah pendekatan yang memanfaatkan data historis dan aktual untuk memahami kondisi bisnis, mengidentifikasi peluang, serta mengurangi risiko pengambilan keputusan yang bersifat subjektif (Liliana, Maulana and Setiawan, 2023; Prasanth *et al.*, 2023). Dalam praktiknya, analitik data tidak hanya berfokus pada perhitungan

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

statistik, tetapi juga mencakup eksplorasi data, visualisasi, dan penyusunan *insight* yang mudah dipahami (Cheng, Li and Bing, 2023; Silmina and Najmah, 2025). Metode *Exploratory Data Analysis* (EDA) terbukti efektif dalam mengidentifikasi pola, tren, dan anomali dalam dataset yang kompleks sebelum dilakukan pemrosesan lebih lanjut (Radhi *et al.*, 2021).

Meskipun memiliki banyak manfaat, *workflow* analitik data tradisional memiliki berbagai kendala. Proses analisis data biasanya memerlukan proses manual yang panjang, terutama pada tahap *data preprocessing* dan *data cleaning*. Studi menunjukkan bahwa sebagian besar waktu kerja *data analyst* dihabiskan untuk tahap ini, sehingga mengurangi efisiensi dan memperlambat proses pengambilan keputusan (Bilal *et al.*, 2022; Li *et al.*, 2025). Selain itu, eksplorasi data biasanya dilakukan melalui pendekatan *trial-and-error* yang membutuhkan pemahaman teknis, seperti penggunaan SQL atau bahasa pemrograman tertentu, yang tidak selalu dimiliki oleh pihak non-teknis.

Permasalahan tersebut semakin kompleks dengan meningkatnya volume dan kompleksitas data pada bisnis saat ini. Dataset kompleks berskala besar cenderung lebih sulit diproses secara menyeluruh, sehingga *insight* yang dihasilkan berpotensi tidak lengkap atau memakan waktu yang lama (Prasanth *et al.*, 2023). Kondisi ini mendorong kebutuhan akan pendekatan analitik yang lebih terotomatisasi, skalabel, dan mudah diakses oleh berbagai pihak dalam organisasi tanpa terpaku pada pemahaman teknis.

2.3 Large Language Model dalam Sistem Analitik Data

Large Language Model (LLM) memberikan pendekatan baru dalam sistem analitik data melalui kemampuan pemrosesan dan *reasoning* berbasis bahasa alami. Dengan LLM, pengguna dapat melakukan eksplorasi data menggunakan bahasa alami tanpa harus berinteraksi langsung dengan *query* atau *code script* (Cheng, Li and Bing, 2023; Nikolakopoulos *et al.*, 2025). Hal ini membuka peluang bagi pihak non-teknis untuk terlibat langsung dalam proses analisis data.

Selain sebagai *natural language interface*, LLM juga dapat dimanfaatkan untuk mendukung proses analitik secara *end-to-end*, seperti identifikasi struktur dataset, pembuatan visualisasi, serta penyusunan laporan hasil analisis (Cheng, Li and Bing,



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

2023). Dalam konteks ini, LLM berperan untuk membantu mempercepat proses eksplorasi data dan penyusunan *insight*.

Namun, pemanfaatan LLM dalam analitik data juga memiliki keterbatasan. Salah satu tantangan utamanya adalah risiko *hallucinations*, yaitu kondisi ketika model menghasilkan informasi yang terlihat benar secara linguistik, tetapi tidak sesuai dengan fakta yang sebenarnya (Woesle, Fischer-Brandies and Buettner, 2025). Selain itu, hasil analisis yang dihasilkan LLM sangat bergantung pada kualitas dan struktur dari konteks yang diberikan, sehingga penggunaan data mentah secara langsung sering kali tidak efektif dan tidak skalabel terutama untuk dataset berskala besar.

2.4 Keterbatasan *Context Window* dan Penurunan Kemampuan *Reasoning* pada LLM

Penambahan *context window* pada *Large Language Models* (LLMs) tidak selalu diikuti dengan peningkatan kemampuan yang sebanding dalam menyelesaikan permasalahan kompleks dengan konteks yang panjang (Du *et al.*, 2025). Sejumlah penelitian menunjukkan bahwa performa *reasoning* model justru mengalami penurunan yang signifikan meskipun seluruh informasi yang dibutuhkan dapat di-*retrieve* secara sempurna (Du *et al.*, 2025). Pada berbagai jenis *task*, termasuk matematika, *question answering*, dan *coding*, ditemukan penurunan performa berkisar antara 13,9% hingga 85% seiring dengan bertambahnya panjang input meskipun model mampu memberikan bukti yang relevan (Du *et al.*, 2025).

Temuan ini menunjukkan bahwa kemampuan model untuk mengakses informasi berbeda dengan kemampuannya dalam menggunakan informasi tersebut secara efektif saat melakukan penalaran. Beberapa studi menunjukkan bahwa performa model tetap menurun bahkan ketika token token yang tidak relevan digantikan dengan *whitespace* atau *dimasking* sepenuhnya (Du *et al.*, 2025). Selain itu, meskipun fenomena *Lost-in-the-Middle* menunjukkan bahwa model bekerja lebih baik ketika informasi penting berada di awal atau akhir teks, jumlah input yang terlalu besar tetap menjadi faktor utama yang menurunkan kemampuan penalaran model (Du *et al.*, 2025).

Implikasi dari temuan ini sangat penting dalam perancangan sistem analitik berbasis LLM, khususnya sistem yang menggunakan data mentah sebagai konteks



secara langsung. Pemrosesan dataset yang besar dan kompleks dalam satu *context window* sering kali tidak memungkinkan karena keterbatasan memori dan komputasi, sehingga menyebabkan hilangnya konteks dan menghasilkan *insight* yang tidak lengkap atau bahkan keliru (Nikolakopoulos *et al.*, 2025). Permasalahan serupa juga banyak ditemukan pada pendekatan *Retrieval-Augmented Generation* (RAG), di mana akurasi model cenderung mengalami penurunan ketika semakin banyak dokumen dimasukkan ke dalam konteks, hal ini tidak sejalan dengan asumsi bahwa penambahan informasi akan selalu meningkatkan kualitas output (Du *et al.*, 2025). Oleh karena itu, penelitian terkini menyarankan agar sistem tidak memasukkan seluruh data mentah ke dalam konteks, tetapi memanfaatkan ringkasan metadata untuk memungkinkan LLM berinteraksi dengan data secara lebih efisien (Nikolakopoulos *et al.*, 2025).

2.5 *Agentic AI* dan *Multi-Agent Systems* dalam Analitik Data

Paradigma kecerdasan buatan mengalami pergeseran dari sistem yang bersifat pasif dan *task-specific* menuju *Agentic AI*, yaitu sistem yang memiliki kemampuan *proactive planning*, *persistent memory*, serta *self-execution* (Abou Ali, Dornaika and Charafeddine, 2025; Kambau, 2025). Berbeda dengan *generative AI* konvensional yang sangat bergantung pada *prompt*, sistem *agentic* memiliki tingkat *autonomy*, *intentionality*, serta kemampuan untuk menyelesaikan masalah yang kompleks dalam jangka panjang tanpa intervensi manusia secara terus menerus (Sapkota, Roumeliotis and Karkee, 2026). Dalam konteks analitik data, evolusi ini memungkinkan AI dalam menggunakan *natural language processing* untuk *data exploration* dan *problem solving* (Cheng, Li and Bing, 2023).

Pendekatan *Multi-Agent Systems* (MAS) semakin memperkuat paradigma ini karena memiliki kemampuan untuk membagi *task* menjadi *subtasks* yang lebih spesifik dan kemudian didistribusikan kepada agen-agen yang saling terkoordinasi (Abou Ali, Dornaika and Charafeddine, 2025). Arsitektur ini biasanya menerapkan mekanisme koordinasi terdistribusi, di mana agen dengan peran khusus seperti *planner*, *retriever*, *executor*, dan *controller* saling berinteraksi melalui *communication protocol* serta *shared memory* (Fan *et al.*, 2021; Goyal, 2025). *Framework* ini biasanya menggunakan konsep *Role Distribution* dalam arsitekturnya, sebagai contoh: *Planner Agent* berperan untuk melakukan

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

pembagian *task*, *Executor Agent* berperan pada *analytic execution* seperti *code execution* atau *data modeling*, sementara *Orchestrator* berperan dalam mengatur *workflow*, *dependencies*, serta sinkronisasi antar agen (Zhao *et al.*, 2025).

Pendekatan *multi-agent* terdistribusi ini memiliki sejumlah keunggulan dibandingkan penggunaan *standalone Large Language Models (LLMs)*. Dengan membagi *subtask* ke dalam agen-agen yang berbeda, MAS dapat meningkatkan toleransi kesalahan dan mengurangi distraksi yang muncul ketika satu model harus memproses terlalu banyak konteks secara langsung (Li *et al.*, 2025).

2.6 Function Calling sebagai Mekanisme Interaksi LLM

Perkembangan *Large Language Models (LLMs)* dari sekadar *static text generator* menjadi alat yang digunakan secara aktif dalam analitik data sangat dipengaruhi oleh penerapan *tool-augmented reasoning*, khususnya melalui pendekatan *function calling* (Nikolakopoulos *et al.*, 2025). *Function calling* merupakan mekanisme yang memungkinkan LLM untuk berinteraksi dengan *external environment* ketika pengguna membutuhkan informasi atau aksi yang berada di luar kemampuan model (Sapkota, Roumeliotis and Karkee, 2026). Melalui mekanisme ini, model tidak hanya menghasilkan jawaban berdasarkan data *training*, tetapi juga dapat mengirimkan *request* terstruktur, yang kemudian dieksekusi oleh lingkungan eksternal. Oleh karena itu, mekanisme ini akan mengubah LLM dari sistem *generator* yang pasif menjadi *interactive problem solver* yang dapat menjalankan proses analitik secara bertahap (Sapkota, Roumeliotis and Karkee, 2026).

Integrasi *function calling* juga berperan penting dalam memitigasi risiko *hallucinations* dan *logical inaccuracies* (Sapkota, Roumeliotis and Karkee, 2026).

LLM pada umumnya hanya bergantung pada data yang ada pada saat *training*, sehingga menghasilkan output yang meyakinkan secara linguistik namun tidak sesuai fakta (Woesle, Fischer-Brandies and Buettner, 2025). Dengan mengimplementasikan *function calling* dalam sebuah *agentic loop*, agen dapat mengambil informasi dari *external environment* sebelum menghasilkan output. Pendekatan *retrieve-then-reason* ini terbukti meningkatkan reliabilitas sistem. Sebagai contoh, *framework ReAct (Reason + Act)* memungkinkan model untuk memahami hasil dari *function calling* dan menyesuaikan langkah *reasoning*



selanjutnya, sehingga risiko *error* dalam *task* analitik yang kompleks dapat berkurang (Fan *et al.*, 2021; Sapkota, Roumeliotis and Karkee, 2026).

Dari sudut pandang *systems engineering*, *function calling* tidak sekadar menjadi fitur tambahan, melainkan merupakan hal fundamental untuk reliabilitas sistem (Abou Ali, Dornaika and Charafeddine, 2025). Mekanisme ini membantu mengatasi masalah *attribution gap* pada sistem neural dengan memberikan catatan tentang setiap tindakan yang diambil oleh sistem (Sapkota, Roumeliotis and Karkee, 2026). Ketika diimplementasi ke dalam sistem, *function calling* dapat memastikan bahwa setiap *action* pada agen bersifat transparan. Dengan demikian, pendekatan ini memungkinkan pengembangan sistem cerdas yang menggabungkan fleksibilitas dari *neural model* dengan eksekusi deterministik dari pemrograman tradisional (Abou Ali, Dornaika and Charafeddine, 2025).

2.7 Metadata Statistik dan *Data Profiling* dalam Analitik Data

Data profiling merupakan proses penting dalam analitik data yang bertujuan untuk mendapatkan metadata statistik guna menggambarkan struktur dan kualitas suatu dataset (Fan *et al.*, 2021). Berbeda dengan metadata semantik yang menjelaskan makna dan konteks bisnis pada data, metadata statistik lebih menekankan pada karakteristik matematis dari data tersebut (Nikolakopoulos *et al.*, 2025). Metadata ini berisi perhitungan statistik deskriptif yang digunakan untuk merepresentasikan distribusi dan variasi pada data.

Peran metadata statistik menjadi sangat krusial untuk menganalisis kualitas data dan menilai kondisi suatu dataset (Bilal *et al.*, 2022). Melalui penggunaan *data profiling tools*, sistem dapat secara otomatis mendeteksi duplikasi data, memverifikasi tipe data setiap atribut, serta menganalisis *missing values* (Bilal *et al.*, 2022). Dengan hasil *data profiling* tersebut, sistem dapat menghasilkan informasi tentang kualitas data yang digunakan.

Dalam penelitian ini, metadata statistik diposisikan sebagai fondasi utama dari strategi *context management* untuk mengatasi keterbatasan *context window* pada *Large Language Model*. Dengan menyediakan informasi dataset yang komprehensif, LLM akan mendapat informasi penting tanpa perlu memasukkan seluruh dataset ke dalam konteks (Nikolakopoulos *et al.*, 2025).

2.8 Metadata-Centric Context Management Berbasis Metadata Statistik

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Metadata-centric context management merupakan strategi yang dirancang untuk mengatasi keterbatasan *Large Language Models* (LLMs) dalam memproses dataset berskala besar. Dibanding memasukkan seluruh dataset ke dalam konteks yang berpotensi menyebabkan *context window saturation*, pendekatan ini hanya memasukkan ringkasan informasi data kepada model (Nikolakopoulos *et al.*, 2025). Dengan ringkasan informasi tersebut, agen mendapatkan *contextual knowledge* yang penting untuk melakukan *reasoning* kompleks tanpa terbebani oleh volume dataset yang berlebihan.

Dalam penelitian ini, metadata statistik diposisikan sebagai *context anchor* yang memberikan informasi tentang data pada proses *reasoning* LLM. Meskipun LLM unggul dalam pemahaman *natural language*, model ini sering kali mengalami kesulitan dalam memahami relasi dan struktur data yang kompleks apabila hanya disajikan data mentah (Nikolakopoulos *et al.*, 2025). Dengan memanfaatkan metadata statistik, agen dapat memiliki informasi tentang kondisi dataset secara menyeluruh sehingga dapat mengidentifikasi masalah pada kualitas data, seperti *data inconsistency* atau *missing values* (Li *et al.*, 2025). Informasi statistik data ini berfungsi sebagai representasi yang reliabel, sehingga model mampu menentukan dan menjalankan operasi analitik yang sesuai dengan karakteristik dataset yang dianalisis.

2.9 Evaluasi Sistem *Agentic AI*

Evaluasi sistem *Agentic AI* membutuhkan pendekatan yang lebih komprehensif dibandingkan AI konvensional, karena melibatkan aktivitas dinamis *agent* dalam menggunakan *tools* dan berinteraksi dengan lingkungan eksternal. Berdasarkan konsep dari *AgentBench* (Liu *et al.*, 2024), evaluasi *AI Agent* tidak hanya berfokus pada tingkat *success rate*, tetapi juga harus mengukur efisiensi sistem melalui penggunaan token dan waktu pemrosesan, serta menganalisis kegagalan model saat melakukan *function call*.

Dalam prosesnya, kemampuan *agent* dapat diukur dari bagaimana *agent* memadukan proses *reasoning* dan *acting*, sebagaimana dijelaskan dalam *framework ReAct* (Yao *et al.*, 2023). Metrik utama dari kemampuan ini adalah *self-correction*, yaitu ketika *agent* mendeteksi adanya *error* atau *invalid action* saat

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

melakukan *function call*, *agent* harus mampu merevisi proses *reasoning*-nya dan mencoba kembali secara otomatis.

Proses ini didukung oleh *framework Reflexion* (Shinn et al., 2023), yang membuktikan bahwa *AI Agent* dievaluasi berdasarkan kemampuannya belajar dari kesalahan (*self-reflection*). Dengan mengevaluasi tindakan sebelumnya dan mengidentifikasi penyebab kegagalan, *agent* dapat mengambil keputusan yang lebih tepat. Hal ini secara langsung akan meningkatkan *success rate* sekaligus mengoptimalkan *token usage* dan *processing time* pada eksekusi-eksekusi berikutnya.

2.10 Penelitian Sejenis

Tabel 2.1 Penelitian Sejenis

Judul	<i>Exploring the Potential of Offline LLMs in Data Science: A Study on Code Generation for Data Analysis</i>	<i>LightVA: Lightweight Visual Analytics with LLM Agent-Based Task Planning and Execution</i>	<i>Is GPT-4 a Good Data Analyst?</i>
Metode	<i>Code Generation berbasis LLM offline untuk analisis dan data profiling menggunakan pendekatan metadata-centric untuk mengatasi limitasi context window.</i>	<i>Framework multi-agent menggunakan proses recursive yang terdiri dari planner, executor, dan controller untuk dekomposisi task dan komposisi multi-view.</i>	<i>Framework data analysis end-to-end yang terdiri dari 3 tahap utama yaitu Code Generation, Code Execution, dan Analysis/Insights Generator.</i>



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Teknologi	<i>Codestral 22B, Qwen 2.5 Coder 14B, dan Apache Spark (PySpark)</i>	<i>GPT-4, GPT-3.5, dan Vega-Lite (Altair) visualization grammar</i>	<i>GPT-4 (gpt-4-0314) dengan prompt-based instructions dan offline execution environment.</i>
Analisis	Perbandingan Efisiensi pada proses <i>Code Generation</i> . Dengan strategi menjaga keamanan data melalui <i>on-premise execution</i> dan menyediakan metadata ringkas sebagai <i>contextual knowledge</i> .	Strategi <i>task-solving</i> secara rekursif dianalisis melalui <i>human-agent collaboration</i> serta membandingkan strategi " <i>decompose on-demand</i> " vs. <i>manual system construction</i> .	Perbandingan antara LLM dan data analis profesional (<i>Senior dan Junior</i>) dengan metrik performa, waktu, dan biaya.
Hasil	<i>Codestral</i> mencapai tingkat akurasi 91% dan otomatisasi 80%; <i>Qwen 2.5 Coder</i> mencapai tingkat akurasi 90% dan otomatisasi 95,6%. Waktu respons rata-rata: 25-60 detik.	Mencapai peningkatan efisiensi 5x lipat dibandingkan metode manual (25 jam vs. 108 jam). <i>Self-reflection loop</i> mengurangi <i>initial error</i> sekitar 40%.	<i>GPT-4</i> dapat mengungguli analis junior dan menyamai analis senior dengan hanya menelan biaya sekitar 0,45% dari tarif analis senior dan dapat beroperasi lebih cepat secara signifikan.

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Penelitian terdahulu menunjukkan dua pendekatan utama dalam pengembangan sistem analitik data berbasis *Large Language Model*, yaitu penggunaan model berbasis *cloud* berskala besar dan penerapan arsitektur analitik yang terstruktur. Studi oleh *Cheng et al.* menunjukkan bahwa GPT-4 mampu berfungsi sebagai analis data *end-to-end* dengan performa yang setara analis tingkat senior, namun dengan waktu pemrosesan yang lebih cepat dan biaya yang jauh lebih rendah (Cheng, Li and Bing, 2023). Temuan ini menegaskan keunggulan model berbasis *cloud* dalam kemampuan *reasoning* dan interpretasi data secara umum.

Di sisi lain, penelitian oleh *Zhao et al.* melalui *framework LightVA* menekankan pentingnya arsitektur *multi-agent* dalam menangani proses *visual analytics* yang kompleks dan iteratif (Zhao *et al.*, 2025). Dengan membagi *tasks* analitik ke dalam *role planner*, *executor*, dan *controller*, pendekatan ini berhasil meningkatkan efisiensi proses serta menurunkan tingkat *initial error* melalui mekanisme *self-reflection*. Hasil tersebut menunjukkan bahwa meskipun model berbasis *cloud* unggul dalam *general reasoning*, *multi-agent orchestration* memiliki mekanisme yang lebih terstruktur dan *robust* untuk pemrosesan analitik yang kompleks.

Selain itu, literatur juga memperjelas *trade-off* antara kemampuan *reasoning* model berbasis *cloud* dan kebutuhan akan efisiensi konteks serta keamanan data. Penelitian oleh *Nikolakopoulos et al.* menunjukkan bahwa penggunaan LLM *offline* yang dikombinasikan dengan pendekatan *metadata-centric* mampu menjaga *functional correctness* sekaligus mengatasi keterbatasan *context window* dengan memanfaatkan informasi dataset dibanding keseluruhan dataset (Nikolakopoulos *et al.*, 2025). Berdasarkan temuan-temuan tersebut, penelitian ini menggabungkan pendekatan *multi-agent coordination*, *metadata-centric context management*, dan mekanisme *function calling* untuk menjaga kualitas *reasoning* LLM serta memastikan proses analitik berjalan secara efisien pada dataset berskala besar.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

BAB III

METODE PENELITIAN

3.1 Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode rekayasa sistem yang berfokus pada perancangan, implementasi, dan pengujian sistem analitik data berbasis *multi-agent* menggunakan *Large Language Model* (LLM). Penelitian ini dirancang untuk mengevaluasi efektivitas pendekatan *function calling* dan *metadata-centric context management* dalam mengatasi keterbatasan *context window* LLM pada proses analitik data.

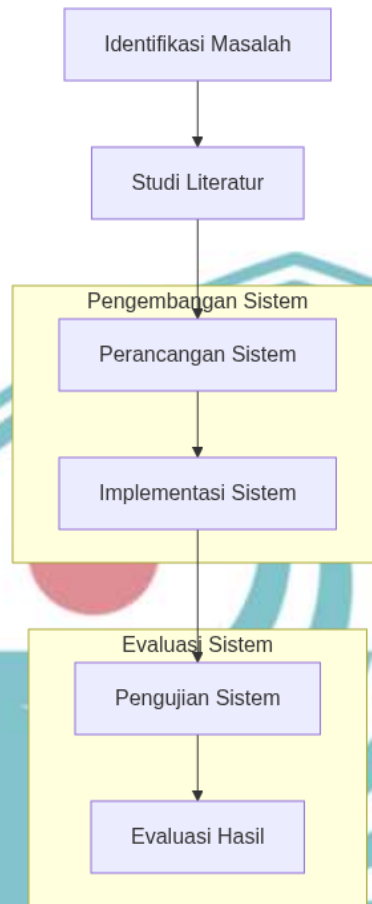
Objek penelitian berupa sistem analitik data berbasis arsitektur *multi-agent* yang mengintegrasikan mekanisme *function calling* dan *metadata-centric context management*. Sistem bekerja dengan menerima dataset berformat CSV sebagai input, yang kemudian diproses melalui tiga tahapan analisis (*Data Preprocessing*, EDA, dan *Explainer*) untuk menghasilkan laporan analitik berisi *insight* teks dan visualisasi data. *Website dashboard* dikembangkan *interface* untuk memfasilitasi pengujian sistem, namun bukan merupakan fokus utama penelitian ini.

Tahapan penelitian dilakukan secara berurutan, yaitu identifikasi masalah, studi literatur, perancangan sistem, implementasi sistem, pengujian sistem, dan evaluasi hasil.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 3.1 Tahapan Penelitian

3.1.1 Identifikasi Masalah

Tahap ini bertujuan untuk mengidentifikasi permasalahan utama dalam penggunaan LLM pada analitik data, khususnya terkait keterbatasan *context window*, degradasi kemampuan *reasoning*, serta keterbatasan pendekatan berbasis data mentah. Identifikasi dilakukan berdasarkan studi awal terhadap literatur dan praktik sistem analitik berbasis LLM.

3.1.2 Studi Literatur

Pada tahap ini dilakukan kajian terhadap penelitian terdahulu yang relevan dengan topik analitik data berbasis LLM, *multi-agent systems*, *function calling*, dan *metadata-centric context management*. Studi literatur digunakan sebagai dasar konseptual dalam perancangan arsitektur sistem yang diusulkan.

3.1.3 Perancangan Sistem

Tahap perancangan mencakup perancangan arsitektur *multi-agent*, alur kerja sistem, mekanisme *function calling*, serta strategi pengelolaan konteks berbasis



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

metadata statistik. Selain itu, dirancang pula integrasi antara *backend agentic AI system* dengan *frontend website dashboard*.

3.1.4 Implementasi Sistem

Tahap ini merealisasikan rancangan sistem ke dalam bentuk sistem yang dapat dijalankan. Implementasi difokuskan pada pengembangan *agentic AI system*, integrasi *function calling* dengan *tools* analitik, serta pembangunan *website dashboard* sebagai antarmuka pengguna.

3.1.5 Pengujian Sistem

Pengujian dilakukan untuk mengevaluasi kinerja sistem dalam memproses dataset, menghasilkan analisis data, serta menyajikan visualisasi. Aspek yang diuji meliputi uji fungsionalitas sistem, efisiensi pemrosesan, konsistensi hasil analisis, dan kemampuan sistem menangani data berskala besar.

3.1.6 Evaluasi Hasil

Tahap evaluasi bertujuan untuk menganalisis hasil pengujian dan menilai sejauh mana sistem yang dikembangkan mampu menjawab permasalahan penelitian. Evaluasi dilakukan dengan membandingkan hasil pengujian terhadap tujuan penelitian yang telah ditetapkan, mencakup efektivitas pendekatan *Metadata-Centric Context Management*, performa sistem *multi-agent*, serta kualitas dan konsistensi output analitik yang dihasilkan.

3.2 Sumber Data

Sumber data dalam penelitian ini merupakan data yang akan digunakan selama proses pengujian untuk mengevaluasi sistem *multi-agent* yang dibangun terhadap rumusan masalah dan tujuan penelitian.

Dataset yang digunakan dalam proses pengujian sistem terdiri dari dataset sintetis dan dataset dari literatur riset. Dataset sintetis digunakan untuk mensimulasikan data *sales* berskala besar yang terkontrol sehingga proses pengujian dapat sesuai dengan kebutuhan evaluasi tujuan secara objektif. Hal tersebut tidak dapat terpenuhi pada dataset publik karena kondisi data yang tidak dapat dikontrol dan secara *default* tidak memiliki *ground truth* yang terkalkulasi secara deterministik. Selain itu, dataset penjualan pada umumnya merupakan rahasia perusahaan, sehingga data penjualan asli tidak dapat ditemukan di internet secara publik. Perancangan kolom dataset sintetis mengacu pada struktur data penjualan yang

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

digunakan pada penelitian Radhi et al. (2021) dengan kolom *order_id*, *product*, *quantity_ordered*, *price_each*, *order_date*, *purchase_address*, *month*, *sales*, *city*, dan *hour*. Berdasarkan acuan tersebut, kolom dataset sintetis dalam penelitian ini diadaptasi menjadi *transaction_id*, *date*, *product_name*, *category*, *store_location*, *payment_method*, *quantity*, *unit_price*, dan *total_bill* untuk merepresentasikan karakteristik transaksi penjualan *e-commerce* di Indonesia. Dataset ini dibuat dalam dua kondisi yaitu *clean* dan *dirty* untuk menguji kemampuan sistem dalam menangani kompleksitas data. Selain itu, dataset NvBench yang diambil dari penelitian "Is GPT-4 a Good Data Analyst?" digunakan sebagai *ground truth* komparatif untuk mengevaluasi kualitas *insight* yang dihasilkan sistem terhadap analisis *data analyst* profesional.

3.3 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan melalui pembuatan dataset sintetis, dan pengambilan dataset dari literatur riset.

Dataset sintetis dihasilkan menggunakan *script* Python dengan beberapa variasi ukuran data mulai dari 100 ribu hingga 10 juta baris. Pemilihan dataset sintetis sebagai instrumen pengujian didasarkan pada tiga pertimbangan utama. Pertama, pengujian akurasi kalkulasi sistem membutuhkan *ground truth* yang dikalkulasi secara deterministik dari data yang sama sehingga tidak ada ambiguitas dalam penilaian kesesuaian hasil. Kondisi ini tidak dapat dipenuhi oleh dataset publik yang tidak menyertakan *ground truth* untuk pertanyaan bisnis spesifik yang didefinisikan dalam penelitian ini. Kedua, pengujian skalabilitas memerlukan variasi ukuran data yang sangat besar hingga 10 juta baris untuk menguji kemampuan arsitektur dalam mengatasi keterbatasan *context window* LLM, sementara dataset publik yang tersedia umumnya tidak mencukupi untuk skenario tersebut. Ketiga, pengujian membutuhkan dataset *dirty* dengan *noise* yang dimasukkan secara terkontrol agar dampak strategi *data cleaning* terhadap akurasi hasil analisis dapat diukur secara objektif. Dataset sintetis dibuat dalam dua versi, yaitu *clean* dan *dirty*, dimana data *dirty* mengandung *data quality issues* seperti *missing values*, *outliers*, nilai negatif, dan inkonsistensi teks. *Ground truth* untuk setiap dataset dihitung secara otomatis pada saat proses generasi sebagai acuan validasi akurasi hasil analisis sistem. Mekanisme pembuatan dataset *dummy* ini



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

kemudian divalidasi oleh profesional di bidang AI/Data sehingga dataset beserta *ground truth* yang dihasilkan dapat digunakan sebagai acuan yang valid dalam penelitian ini.

Dataset *NvBench* diperoleh dengan mengunduh dataset yang dilampirkan pada penelitian "*Is GPT-4 a Good Data Analyst?*" (Cheng, Li and Bing, 2023). Dataset ini mencakup lima kasus analisis bisnis dengan pertanyaan bisnis yang telah dijawab oleh *Senior Data Analyst*, *Junior Data Analyst*, dan GPT-4, sehingga dapat digunakan sebagai *baseline* komparatif untuk mengevaluasi kualitas *insight* sistem yang dibangun.

Data untuk penulisan skripsi dikumpulkan melalui studi literatur terhadap jurnal ilmiah dan *research paper* yang relevan dengan topik penelitian.

3.4 Teknik Analisis Data

Teknik analisis data dalam penelitian ini dilakukan melalui dua pendekatan utama, yaitu analisis komparatif berbasis *ground truth* dan analisis berbasis model dan *framework* sistem.

Analisis komparatif berbasis *ground truth* dilakukan melalui dua cara. Pertama, validasi akurasi numerik dengan membandingkan nilai pada laporan analisis sistem terhadap *ground truth* yang telah dikalkulasi dari dataset sintesis. Kedua, evaluasi kualitas *insight* saintifik dengan membandingkan hasil analisis sistem terhadap temuan *Senior Data Analyst*, *Junior Data Analyst*, dan GPT-4 pada dataset *NvBench* dari literatur. Perbandingan ini bersifat observasional deskriptif yang bertujuan untuk mengamati sejauh mana pendekatan dan arah rekomendasi sistem sejalan dengan analisis yang dijadikan acuan, bukan untuk menghasilkan penilaian kualitatif yang definitif. Tidak ada penilai eksternal yang dilibatkan dalam proses ini, sehingga temuan yang dihasilkan tidak diklaim sebagai evaluasi formal terhadap kualitas *insight*.

Analisis berbasis model dan *framework* dilakukan dengan mengevaluasi performa sistem menggunakan dua model LLM, yaitu Xiaomi MiMo V2 Flash dan OpenAI GPT OSS 120B, sebagai komponen *reasoning* dalam arsitektur *multi-agent* yang dikoordinasikan melalui *Google Agent Development Kit* (ADK). Analisis ini difokuskan pada efektivitas mekanisme *function calling*, konsistensi output antar

eksekusi, serta kemampuan sistem dalam memproses dataset berskala besar melalui pendekatan *metadata-centric context management*.



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta





Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisis Kebutuhan

Berdasarkan studi literatur, identifikasi masalah, dan perancangan sistem yang telah diuraikan pada bab sebelumnya, tahap pertama dalam implementasi adalah mengidentifikasi kebutuhan penelitian. Kebutuhan tersebut mencakup kebutuhan sistem analitik *multi-agent*, kebutuhan pengembangan *web dashboard*, kebutuhan perangkat keras, dan kebutuhan perangkat lunak. Semua kebutuhan ini harus dipenuhi untuk memastikan sistem dapat menjalankan proses analitik data secara otomatis dan memitigasi keterbatasan *context window*.

4.1.1 Kebutuhan Sistem *Multi-Agent*

4.1.1.1 Kebutuhan Dataset

Untuk mengevaluasi performa sistem *multi-agent* dalam menangani dataset berskala besar, diperlukan dataset uji yang merepresentasikan permasalahan analitik data riil dalam konteks bisnis. Rincian kebutuhan dataset dapat dilihat pada Tabel 4.1.

Tabel 4.1 Kebutuhan Dataset

Kode	Nama	Deskripsi
KD-1	Studi Kasus	Dataset merupakan data penjualan (<i>sales data</i>) yang merepresentasikan permasalahan bisnis.
KD-2	Format Data	Dataset berformat tabular (seperti CSV) yang strukturnya dapat dibaca dan diekstrak data statistiknya.
KD-3	Kondisi Data	Dataset terdiri dari <i>clean</i> dan <i>dirty</i> dataset.
KD-4	Ukuran Data	Dataset berukuran 100 Ribu hingga 10 Juta Baris.

Dataset yang digunakan dalam penelitian ini terdiri dari dua jenis. Pertama, dataset sintesis yang dibuat menggunakan Python *script* dengan struktur data penjualan yang diadaptasi dari penelitian Radhi et al. (2021). Dataset ini dibuat dalam lima variasi ukuran, yaitu 100 ribu, 500 ribu, 1 juta, 5 juta, dan 10 juta baris, di mana setiap ukuran memiliki dua versi *clean* dan *dirty*.



Kedua, dataset NvBench yang diperoleh dari penelitian "*Is GPT-4 a Good Data Analyst?*" (Cheng, Li and Bing, 2023) digunakan sebagai *baseline* komparatif untuk mengevaluasi kualitas *insight* sistem. Dataset yang digunakan dalam penelitian tersebut memiliki konteks yang bervariasi, yaitu *candidate_poll* berisi data hasil *polling* kandidat beserta informasi pribadinya, *student* berisi data siswa beserta dengan jam akademik dan GPA nya, *hr* berisi data karyawan beserta departemen dan gajinya, *inn* berisi data reservasi kamar beserta informasi kamarnya, dan *wine* berisi data jenis-jenis *wine* beserta informasi harga dan skornya.

4.1.1.2 Kebutuhan *Large Language Model (LLM)*

Sebagai otak dari *Agentic AI*, diperlukan model yang memadai untuk melakukan *inference*, *planning*, dan *execution*. Spesifikasi kebutuhan LLM ditunjukkan pada Tabel 4.2.

Tabel 4.2 Kebutuhan *Large Language Model (LLM)*

Kode	Nama	Deskripsi
KLLM-1	Model Spesifik	Sistem memanfaatkan LLM <i>Xiaomi MiMo V2 Flash</i> dan <i>OpenAI GPT OSS 120B</i> .
KLLM-2	Kapabilitas <i>Function Calling</i>	LLM wajib memiliki kapabilitas untuk melakukan mekanisme <i>function calling</i> guna berinteraksi dengan lingkungan eksternal secara deterministik.

Kedua model LLM yang digunakan akan mengeksekusi sejumlah kasus yang serupa untuk dibandingkan performanya. Kapabilitas *function calling* menjadi parameter yang sangat penting karena paradigma *Agentic AI* akan banyak melakukan *function calling* untuk berinteraksi dengan dataset yang diolah sistem.

4.1.1.3 Kebutuhan Kemampuan *Agentic AI*

Selain LLM, arsitektur *agent* itu sendiri harus memenuhi kebutuhan fungsionalitas yang diperlukan agar beban konteks pada proses analitik data berkurang. Rincian kebutuhan agen disajikan pada Tabel 4.3.

Tabel 4.3 Kebutuhan Kemampuan Agen

Kode	Nama	Deskripsi
KA-1	<i>Metadata-Centric</i>	Agen mampu melakukan <i>reasoning</i> dengan merujuk pada ringkasan metadata statistik (bukan memasukkan <i>raw data</i> ke <i>context window</i>).

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

KA-2	Eksekusi Analitik	Agen mampu membuat <i>query</i> analitik ke dataset yang ada.
KA-3	<i>Role Distribution</i>	Sistem mendukung pembagian subtask kepada agen dengan peran spesifik (seperti <i>planner</i> , <i>executor</i> , dan <i>orchestrator</i>).

Dalam pengembangan sistem *Agentic AI* yang skalabel, diperlukan beberapa kemampuan yang dibutuhkan oleh *agent* untuk memaksimalkan *metadata-centric context management*. *Metadata-Centric* disini akan berfungsi sebagai bahasa komunikasi antara *Agentic System* dan dataset yang diolah. Hal ini akan mengoptimalkan *context* yang dikirim ke LLM dengan hanya mengirimkan data yang diperlukan dalam proses *reasoning* LLM tanpa mengirim keseluruhan dataset yang akan menyebabkan *context window bottleneck*. Dalam proses memperoleh data yang dibutuhkan dalam *reasoning* LLM tersebut, *agent* memerlukan kemampuan untuk melakukan eksekusi analitik melalui mekanisme *function calling*. Eksekusi analitik ini dilakukan melalui *SQL query* yang akan dijalankan pada dataset yang digunakan. Hal ini akan mereduksi jumlah token yang masuk kedalam LLM sehingga LLM dapat fokus pada informasi penting yang dibutuhkan untuk proses analisis.

4.1.2 Kebutuhan Pengembangan Web (Dashboard)

Sebagai *interface* yang akan digunakan oleh *user*, *web dashboard* dirancang untuk mempermudah pemanfaatan sistem *multi-agent* yang dikembangkan dalam penelitian ini.

4.1.2.1 Kebutuhan Fungsional

Kebutuhan fungsional mendefinisikan fitur-fitur yang harus ada pada *web dashboard* analitik agar pengguna dapat dengan mudah berinteraksi dengan sistem.

Tabel 4.4 Kebutuhan Fungsional

Kode	Nama	Deskripsi
KF-1	Autentikasi	Pengguna dapat melakukan registrasi dan <i>login</i> ke dalam sistem.
KF-2	Manajemen Proyek	Pengguna dapat membuat proyek analitik dan mengunggah dataset berformat CSV.
KF-3	<i>Natural</i>	Pengguna dapat memasukkan kueri pertanyaan

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

	<i>Language Query</i>	analitik bisnis melalui bahasa alami.
KF-4	Laporan Analitik	Sistem mampu menampilkan <i>insight</i> hasil analisis <i>multi-agent</i> dalam bentuk teks serta visualisasi (<i>chart</i>).

Kebutuhan Fungsional yang diperlukan pada *web dashboard* didefinisikan pada Tabel 4.4, dimana secara umum *dashboard* yang dikembangkan harus mampu menjembatani interaksi antara *user* dan *agentic system* yang dibangun. Oleh karena itu, *dashboard* memiliki fitur untuk manajemen proyek, *natural language query*, dan juga halaman laporan analitik yang dihasilkan oleh *agentic system*.

4.1.2.1 Kebutuhan Non-Fungsional

Kebutuhan non-fungsional berkaitan dengan kualitas, performa, dan batasan dari sistem *web* yang dibangun.

Tabel 4.5 Kebutuhan Non-Fungsional

Kode	Nama	Deskripsi
KNF-1	Kecepatan Proses Analisis	Proses analisis data secara <i>end-to-end</i> harus selesai dalam waktu kurang dari 30 menit.
KNF-2	Mitigasi Halusinasi	Pemrosesan LLM dirancang untuk meminimalisir risiko halusinasi dengan eksekusi <i>query</i> secara deterministik.
KNF-3	Skalabilitas Analisis	Sistem <i>backend</i> dapat memproses analisis secara <i>asynchronous</i> agar sistem tidak <i>freeze</i> saat proses analisis berlangsung panjang.

Kebutuhan non-fungsional yang ditunjukkan pada Tabel 4.5 mendefinisikan kebutuhan performa sistem yang meliputi waktu proses analisis data *end-to-end* kurang dari 30 menit, mitigasi halusinasi LLM, serta mekanisme *asynchronous processing* untuk mencegah sistem *freeze* pada saat proses berlangsung.

4.1.3 Kebutuhan Perangkat Keras

Dalam perancangan sistem, perangkat keras dibutuhkan untuk menjalankan proses komputasi *frontend*, *backend*, dan manajemen *queue*.

Tabel 4.6 Kebutuhan Perangkat Keras

Kode	Nama	Deskripsi
KPK-1	CPU	Prosesor komputasi minimal 1 <i>core</i> untuk mengelola



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

		pemrosesan analitik.
KPK-2	RAM	Minimal 4 GB untuk melakukan pemrosesan analisis data berskala besar.
KPK-3	Internet	Koneksi internet yang stabil untuk berkomunikasi dengan layanan <i>API Large Language Model</i> yang ada di infrastruktur <i>cloud</i> .

Spesifikasi perangkat keras yang diperlukan untuk melakukan pemrosesan cenderung sederhana karena pengolahan data dioptimisasi oleh arsitektur sistem menggunakan DuckDB.

4.1.4 Kebutuhan Perangkat Lunak

Pengembangan arsitektur sistem *multi-agent* dan *dashboard* analitik membutuhkan perangkat lunak yang sesuai. Daftar perangkat lunak dipisahkan secara modul, seperti pada Tabel 4.7.

Tabel 4.7 Kebutuhan Perangkat Lunak

Kode	Nama	Deskripsi
KPL-1	<i>Library LLM & Agent</i>	Sistem <i>Agentic AI</i> dibangun menggunakan <i>library Google Agent Development Kit</i> berbasis Python untuk mengorkestrasi pemanggilan <i>function calling</i> dan LLM.
KPL-2	<i>Data Processing Engine</i>	<i>DuckDB</i> digunakan untuk mengeksekusi <i>query</i> dan memproses struktur dataset lokal secara efisien.
KPL-3	<i>Asynchronous Processing</i>	<i>Celery</i> (berbasis Python) untuk mengatur pemrosesan <i>agent</i> secara <i>asynchronous</i> .
KPL-4	<i>Framework Web & ORM</i>	<i>Next.js</i> dengan bahasa pemrograman <i>TypeScript</i> untuk pengembangan <i>frontend</i> , dipadukan dengan <i>Prisma ORM</i> untuk manajemen <i>database</i> .
KPL-5	<i>Deployment</i>	<i>Docker</i> untuk pembuatan <i>container</i> dan isolasi lingkungan aplikasi (<i>deployment</i>).

Pengembangan sistem dibagi menjadi *Frontend* dan *Backend*. *Frontend* berfokus pada pengembangan *interface* yang menjembatani *user* dalam melakukan interaksi dengan *agentic system* yang dibangun. *Framework* yang digunakan untuk membangun *web dashboard* adalah *NextJS* berbasis *TypeScript* yang dikombinasikan dengan *Prisma ORM* untuk menghubungkan antara *dashboard*

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

dengan *database* sistem. Di sisi lain, *Backend* berfokus pada pengembangan *Agentic AI System* menggunakan *framework Google Agent Development Kit* berbasis *Python* yang dibungkus dengan *FastAPI* untuk menghasilkan *API Endpoint* yang akan diintegrasikan dengan *web dashboard*. Selain itu, untuk mengisolasi *environment* dan mempermudah *deployment*, keseluruhan aplikasi dikontainerisasi menggunakan *Docker*.

4.2 Perancangan Sistem

Setelah kebutuhan penelitian didefinisikan, proses selanjutnya dalam penelitian ini adalah melakukan perancangan sistem berdasarkan kebutuhan tersebut. Hasil perancangan ini akan dijadikan sebagai acuan pada pengembangan atau implementasi sistem. Perancangan sistem pada penelitian ini dibagi menjadi dua bagian utama, yaitu perancangan sistem *multi-agent*, dan perancangan web *dashboard*.

4.2.1 Perancangan Sistem *Multi-Agent*

Sistem *multi-agent* dirancang sebagai alat analitik data yang memanfaatkan teknologi *Large Language Model (LLM)* untuk mengotomatisasi proses analisis. Perancangan pada bagian ini mencakup pembagian peran agen, *metadata-centric context management*, serta mekanisme *function calling*.

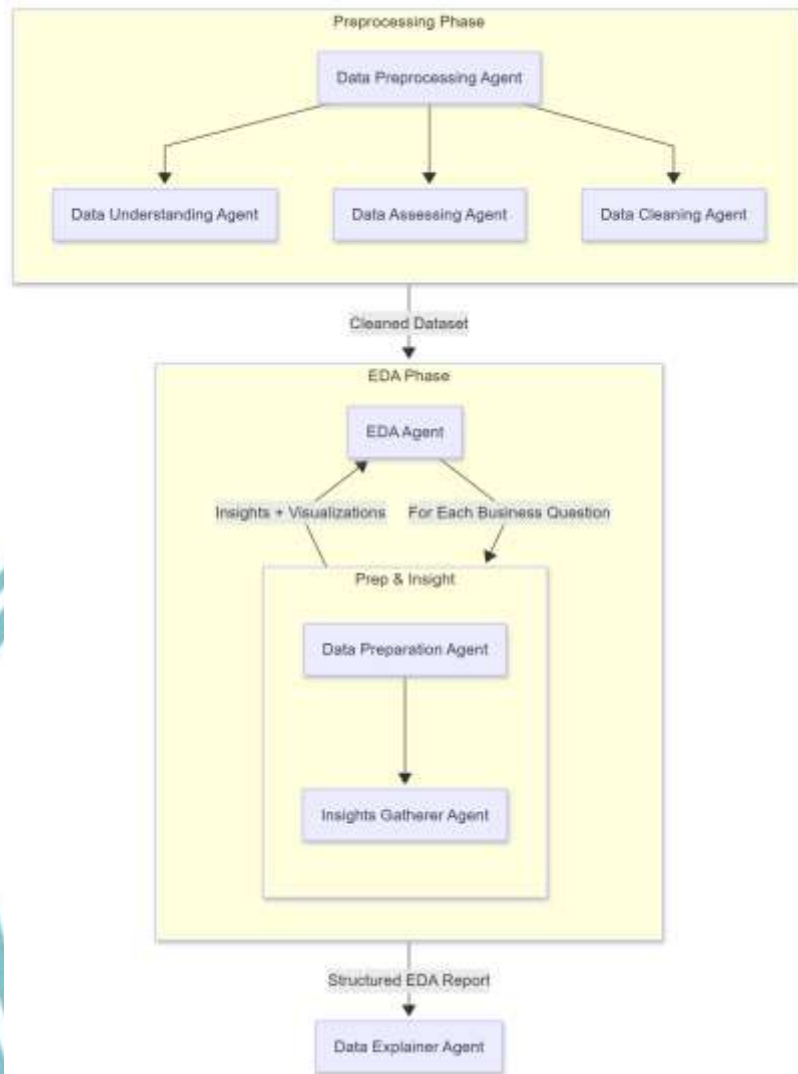
4.2.1.1 Arsitektur *Multi-Agent*

Dalam sistem *Agentic AI* ini, proses analisis data yang kompleks dipecah menjadi beberapa *sub-task* yang didistribusikan kepada *agent* dengan peran yang lebih spesifik. Pembagian peran ini bertujuan untuk mengurangi beban konteks LLM dan membuat pemrosesan menjadi lebih terarah.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.1 Arsitektur *Agentic AI*

Secara arsitektur, sistem *Multi-Agent* yang dikembangkan dibagi menjadi 3 fase utama, yaitu: *Preprocessing*, *Exploratory Data Analysis (EDA)*, dan *Explainer*. Ketiga fase tersebut digabungkan menjadi satu *pipeline* analitik data secara *sequential*, dimana dataset yang masuk kedalam sistem akan melalui tahap *Preprocessing*, *EDA*, dan *Explainer* secara bertahap.

4.2.1.1.1 Fase *Preprocessing*

Fase *Preprocessing* akan diorkestrasi dengan *Data Preprocessing Agent* yang akan membagi tugas kepada *agent* yang lebih spesifik yaitu: *Data Understanding Agent*, *Data Assessing Agent*, dan *Data Cleaning Agent*. Perancangan komposisi *agent* tersebut didasari pada rancangan *AutoDCWorkflow* dimana terdapat 3 tahap utama dalam otomatisasi *data cleaning* yaitu *Select Target Columns (Understanding)*,

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

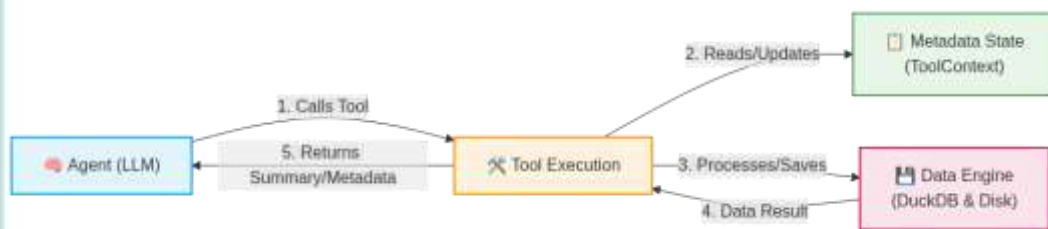
Inspect Column Quality (Assessing), dan *Generate Operations & Arguments (Cleaning)* (Li et al., 2025). Fase *Preprocessing* diharapkan mampu menghasilkan dataset yang *clean* dan sesuai dengan kebutuhan analisis mengikuti *Business Questions* yang dimasukkan oleh *user*.

4.2.1.1.2 Fase EDA dan *Explainer*

Dataset yang telah diproses pada fase *preprocessing* akan dikirim menuju fase *Exploratory Data Analysis (EDA)* untuk kemudian dilakukan *insights gathering*. Proses ini bertujuan untuk menghasilkan informasi yang dibutuhkan untuk menjawab *business questions* yang ada. Fase tersebut akan diorkestrasi oleh *EDA Agent* yang akan memanggil alur *sequential Data Preparation Agent* dan *Insights Gatherer Agent* secara bertahap untuk setiap *business question*. Proses *sequential* ini dirancang mengadaptasi penelitian yang dilakukan oleh Cheng et al. dengan alur *data collection*, *data visualization*, dan *data analysis*. Fase *EDA* akan menghasilkan narasi *insights* dan visualisasi yang akan digunakan oleh *Data Explainer Agent* dalam membuat laporan analisis yang sesuai dengan kebutuhan *user*. Laporan analisis tersebut akan disimpan dalam format *Markdown* dan akan ditampilkan kepada *user* sebagai *output* dari sistem *agentic AI* yang dikembangkan.

4.2.1.2 *Metadata-Centric Context Management*

Untuk memitigasi *context window bottleneck* saat memproses dataset berukuran besar, sistem dirancang untuk tidak memuat *raw data* ke dalam LLM. Sebagai gantinya, dirancang mekanisme *metadata-centric* seperti yang dilakukan Nikolakopoulos et al. untuk memberi *agent* pengetahuan kontekstual dari data tersebut. Sistem akan membaca dataset untuk menghasilkan ringkasan metadata yang berisi struktur kolom, tipe data, jumlah baris, dan statistik dasar. Metadata statistik inilah yang digunakan oleh *agent* sebagai *context anchor* dalam melakukan *reasoning*, menggantikan peran keseluruhan *raw data*. Selanjutnya, data statistik tersebut akan diolah menjadi *knowledge* yang menjadi acuan *agentic AI* dalam membuat *query* untuk menghasilkan *insights*. Hal ini akan mengoptimalkan *context window* dengan hanya memasukkan data-data yang memang dibutuhkan.



Gambar 4.2 Flow Metadata-centric Context Management

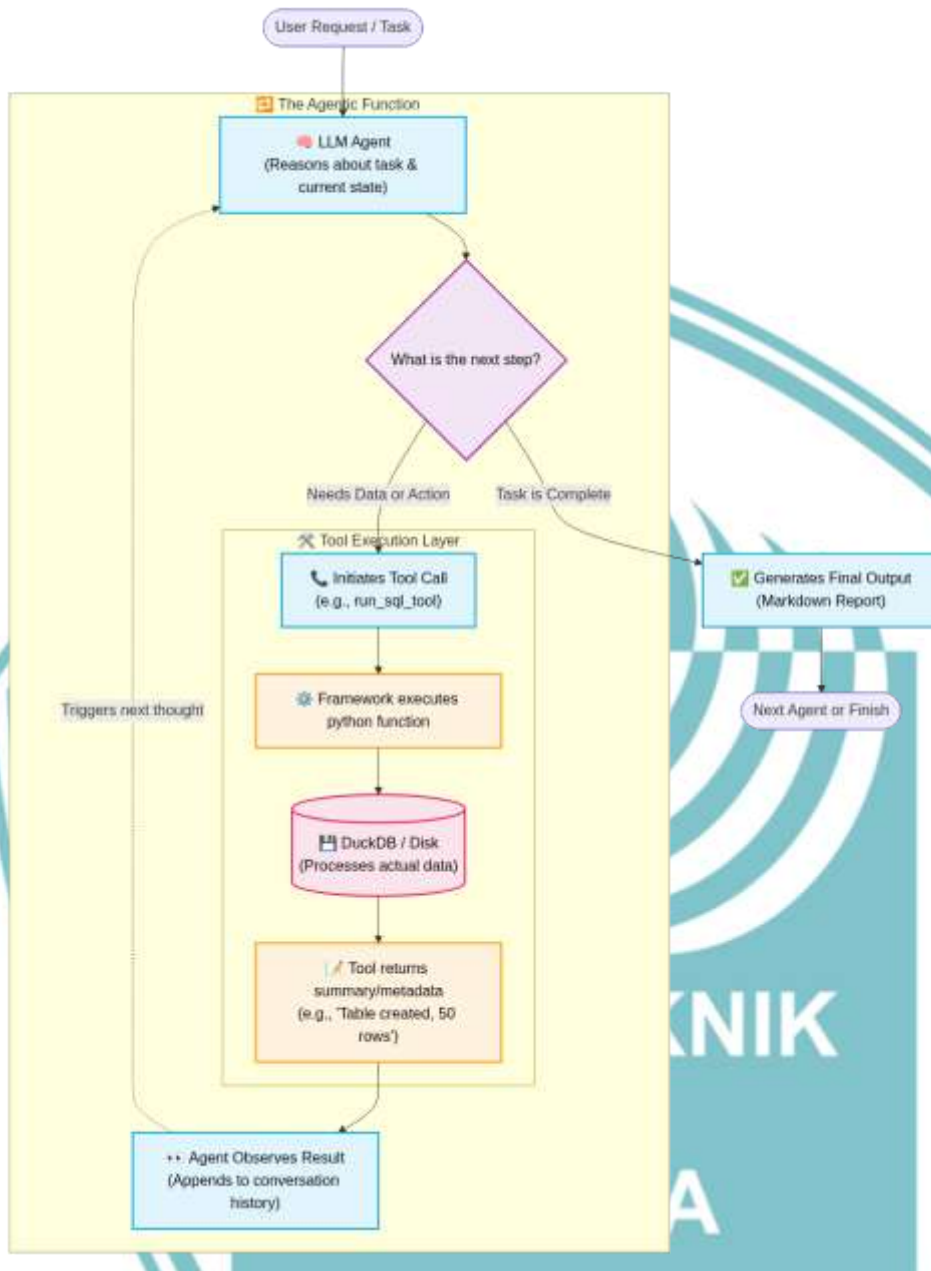
Gambar 4.2 dapat menjadi ilustrasi sederhana terkait dengan proses *Metadata-centric Context Management*. Proses ini dilakukan ketika *Agent* menjalankan *tools* yang tersedia pada *agentic system* dimana *tools* yang ada akan saling terhubung dengan *Metadata State* dan juga *Data Engine*. *Metadata State* ini akan mencatat informasi penting terkait data yang dieksekusi pada sistem seperti informasi data yang telah diolah. Selain itu, *Data Engine* akan berperan penting dalam proses pengolahan data. Proses seperti *missing values imputation*, *aggregation*, dll juga akan dilakukan pada *Data Engine* berupa *DuckDB* untuk *SQL Operation* dan juga *Disk* untuk menyimpan data yang sudah diolah. Setelah itu, *tool* yang dieksekusi oleh *agent* akan mengembalikan output berupa ringkasan data. Hal ini akan mengurangi beban konteks input *agent* tanpa menghilangkan informasi penting dari data yang diolah.

4.2.1.3 Mekanisme *Function Calling* dan Eksekusi Analitik

Agar *agent* dapat memperoleh data yang tidak ada di metadata awal, *agent* dirancang agar mampu melakukan proses analitik melalui mekanisme *function calling*. LLM akan membuat *query SQL* yang bertujuan untuk melakukan proses analitik secara deterministik. *Query* tersebut kemudian akan dieksekusi oleh *data processing engine DuckDB* pada dataset. Mekanisme *Function Calling* ini menjadi alat utama dalam mengoptimalkan interaksi antara data dengan LLM yang digunakan. Dengan ini, sistem *Agentic AI* menjadi *scalable* tanpa terbebani oleh limitasi *context window* pada LLM.

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.3 Flow Function Calling

Mekanisme *Function Calling* ini dapat digambarkan seperti pada gambar 4.3. Proses ini dimulai ketika *agent* membutuhkan informasi dari dataset. *Agent* akan melakukan *tool call* sesuai dengan kebutuhan analisis yang diperlukan. *Tool* yang dipanggil oleh *agent* kemudian akan mengeksekusi *python code* untuk melakukan operasi pada *data engine*. Hasil dari operasi tersebut akan dikirimkan kembali ke *agent* sebagai informasi tambahan dalam melakukan *reasoning* sebelum menghasilkan output.



Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

4.2.2 Perancangan *Web Dashboard*

Bagian kedua dari perancangan sistem adalah perancangan *web dashboard*. Web ini dirancang sebagai *interface* yang akan digunakan pengguna untuk mempermudah pemanfaatan sistem *multi-agent*. Perancangan web mencakup *use case diagram*, *activity diagram*, *database*, dan *wireframe* antarmuka.

4.2.2.1 *Use Case Diagram*

Use case diagram pada penelitian ini dikembangkan dari kebutuhan fungsional web yang telah disampaikan sebelumnya. Diagram ini menggambarkan hubungan antara pengguna dengan aksi yang dapat dilakukan pada web. Terdapat satu aktor utama pada sistem ini, yaitu *user*. Aksi-aksi yang dapat dilakukan meliputi registrasi dan *login*, manajemen proyek, memasukkan pertanyaan analitik, serta melihat laporan teks *insight* dan visualisasi.



-
- ```

graph TD
 User((User))
 MengaksesPlatform([Mengakses Platform])
 MelakukanSignIn([Melakukan Sign In])
 MelakukanSignUp([Melakukan Sign Up])
 MengunggahDataset([Mengunggah Dataset])
 MelihatSampelData([Melihat Sampel Data])
 MengubahDataset([Mengubah Dataset])
 MembuatProyekAnalisis([Membuat Proyek Analisis])
 MemasukkanPertanyaanBisnis([Memasukkan Pertanyaan Bisnis])
 MengubahModelLLM([Mengubah Model LLM])
 MemulaiAnalisis([Memulai Analisis])
 MelihatNamaFile([Melihat Nama File yang sedang dianalisis])
 MelihatDurasiAnalisis([Melihat Durasi Analisis])
 MelihatTimelineProsesAnalisis([Melihat Timeline Proses Analisis])
 MembatalkanProsesAnalisis([Membatalkan Proses Analisis])
 MelihatLogProsesAnalisis([Melihat Log Proses Analisis])
 MelihatDetailLogProsesAnalisis([Melihat Detail Log Proses Analisis])
 MemilihBagianAnalisis([Memilih Bagian Analisis])
 MelihatNarasiLaporanAnalisis([Melihat Narasi Laporan Analisis])
 MelihatDaftarIsiLaporanAnalisis([Melihat Daftar Isi Laporan Analisis])
 MelakukanAnalisisUlang([Melakukan Analisis Ulang])
 KembaliKeHalamanUtama([Kembali ke Halaman Utama])
 MenghapusProyekAnalisis([Menghapus Proyek Analisis])
 MengubahNamaProyekAnalisis([Mengubah Nama Proyek Analisis])
 MelihatProyekAnalisis([Melihat Proyek Analisis])
 MengelolaProyekAnalisis([Mengelola Proyek Analisis])
 MelihatHalamanDaftarProyek([Melihat Halaman Daftar Proyek])
 MelihatStatistikProsesAnalisis([Melihat Statistik Proses Analisis])
 MelihatListSeluruhProyek([Melihat List Seluruh Proyek])
 MengubahModeTampilanListProyek([Mengubah Mode Tampilan List Proyek])
 MencariProyekAnalisis([Mencari Proyek Analisis])

 MengaksesPlatform -.->|<<include>>| MelakukanSignIn
 MelakukanSignIn -.->|<<include>>| MelakukanSignUp
 MelakukanSignUp -.->|<<include>>| MengunggahDataset
 MengunggahDataset -.->|<<include>>| MelihatSampelData
 MengunggahDataset -.->|<<include>>| MengubahDataset
 MengunggahDataset -.->|<<include>>| MembuatProyekAnalisis
 MembuatProyekAnalisis -.->|<<include>>| MemasukkanPertanyaanBisnis
 MembuatProyekAnalisis -.->|<<include>>| MengubahModelLLM
 MembuatProyekAnalisis -.->|<<include>>| MemulaiAnalisis
 MemulaiAnalisis -.->|<<include>>| MelihatNamaFile
 MemulaiAnalisis -.->|<<include>>| MelihatDurasiAnalisis
 MemulaiAnalisis -.->|<<include>>| MelihatTimelineProsesAnalisis
 MemulaiAnalisis -.->|<<include>>| MembatalkanProsesAnalisis
 MembatalkanProsesAnalisis -.->|<<include>>| MelihatLogProsesAnalisis
 MelihatLogProsesAnalisis -.->|<<include>>| MelihatDetailLogProsesAnalisis
 MelihatDetailLogProsesAnalisis -.->|<<include>>| MemilihBagianAnalisis
 MemilihBagianAnalisis -.->|<<include>>| MelihatNarasiLaporanAnalisis
 MelihatNarasiLaporanAnalisis -.->|<<include>>| MelihatDaftarIsiLaporanAnalisis
 MelihatDaftarIsiLaporanAnalisis -.->|<<include>>| MelakukanAnalisisUlang
 MelakukanAnalisisUlang -.->|<<include>>| KembaliKeHalamanUtama
 KembaliKeHalamanUtama -.->|<<include>>| MenghapusProyekAnalisis
 MenghapusProyekAnalisis -.->|<<include>>| MengubahNamaProyekAnalisis
 MengubahNamaProyekAnalisis -.->|<<include>>| MelihatProyekAnalisis
 MelihatProyekAnalisis -.->|<<include>>| MengelolaProyekAnalisis
 MengelolaProyekAnalisis -.->|<<include>>| MelihatHalamanDaftarProyek
 MelihatHalamanDaftarProyek -.->|<<include>>| MelihatStatistikProsesAnalisis
 MelihatHalamanDaftarProyek -.->|<<include>>| MelihatListSeluruhProyek
 MelihatHalamanDaftarProyek -.->|<<include>>| MengubahModeTampilanListProyek
 MelihatHalamanDaftarProyek -.->|<<include>>| MencariProyekAnalisis

```

Gambar 4.4 *Use Case Diagram*



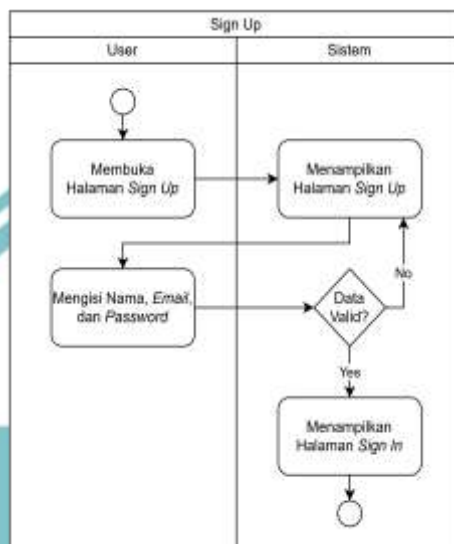
### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

#### 4.2.2.2 Activity Diagram

*Activity Diagram* dikembangkan dari *use case diagram* untuk menggambarkan alur aksi yang lebih rinci:

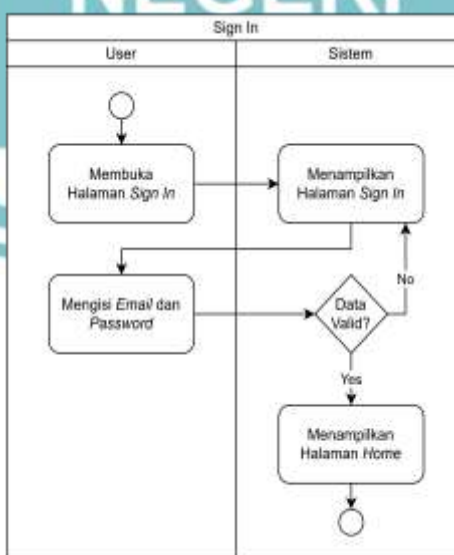
##### 4.2.2.2.1 Sign Up



Gambar 4.5 Activity Diagram Sign Up

Gambar 4.5 menunjukkan alur proses pendaftaran akun atau *Sign Up* yang dimulai dengan membuka halaman *Sign Up* lalu mengisi form data yang diperlukan dan diakhiri dengan menekan tombol “*Sign Up*” yang akan menavigasi pengguna ke halaman *Sign In* jika data yang dimasukkan valid.

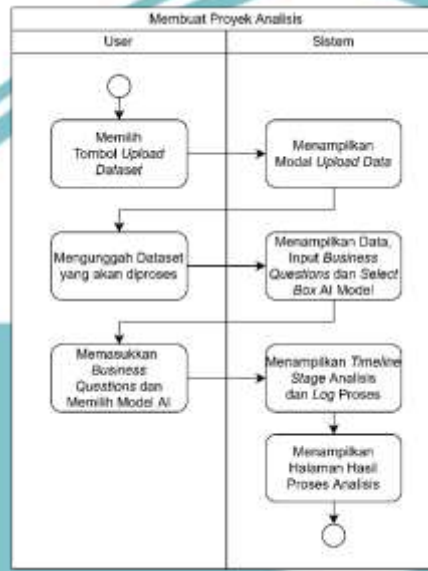
##### 4.2.2.2.2 Sign In



Gambar 4.6 Activity Diagram Sign In

Gambar 4.6 menunjukkan alur proses pendaftaran akun atau *Sign In* yang dimulai dengan membuka halaman *Sign In* lalu mengisi form kredensial dan diakhiri dengan menekan tombol “*Sign In*” yang akan menavigasi pengguna ke halaman *Home* jika data yang dimasukkan valid.

#### 4.2.2.2.3 Membuat Proyek Analisis



Gambar 4.7 Activity Diagram Membuat Proyek Analisis

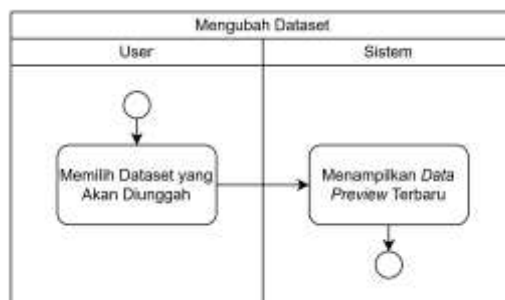
Gambar 4.7 menunjukkan alur proses pembuatan proyek analisis baru. Proses ini bermula pada saat user menekan tombol *Upload Dataset* yang akan memunculkan modal *Upload Dataset*. Setelah itu, user akan mengunggah dataset yang akan dianalisis. Ketika dataset berhasil diunggah maka akan muncul *Preview* dari data yang diunggah serta input untuk memasukkan *Business Questions* yang akan menjadi acuan sistem *Agentic AI* pada *backend* dalam melakukan analisis. Selain itu, *user* juga dapat mengganti model AI yang akan digunakan untuk analisis. *User* akan menekan tombol “*Generate Insight*” yang akan memulai proses analisis pada *backend*. Setelah itu, *website* akan melakukan *polling* ke *backend* untuk menampilkan proses yang sedang dilakukan pada *Agentic AI* beserta *log* proses sehingga seluruh proses yang dilakukan dapat dipantau user secara *real-time*. Jika proses analisis telah selesai, maka *website* akan menavigasi *user* ke halaman hasil analisis secara otomatis.



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

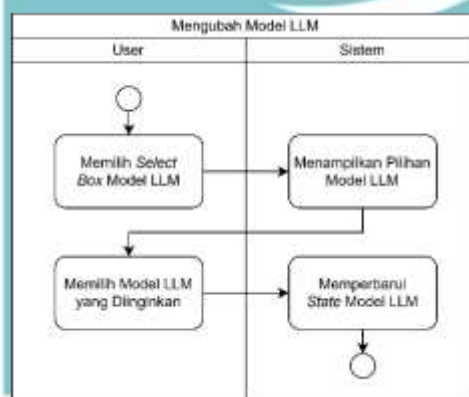
#### 4.2.2.2.4 Mengubah Dataset



Gambar 4.8 Activity Diagram Mengubah Dataset

Gambar 4.8 menunjukkan alur proses yang dilakukan ketika *user* ingin mengganti dataset yang sudah dipilih sebelumnya. Alur ini dimulai dengan *user* menekan *Dataset Upload Field* yang ada pada Modal *New Analysis*. Setelah itu *user* harus memilih dataset yang akan diunggah. Jika dataset baru sudah dipilih, maka sistem akan melakukan *Update State* dan juga *Data Preview* menyesuaikan dengan dataset yang baru diunggah. Terakhir, tampilan *Data Preview* akan diperbarui sesuai dengan dataset yang diunggah.

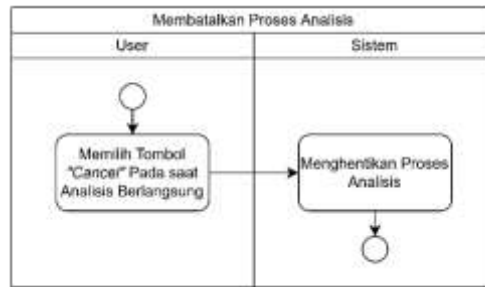
#### 4.2.2.2.5 Mengubah Model LLM



Gambar 4.9 Activity Diagram Mengubah Model LLM

Gambar 4.9 menunjukkan alur proses *user* dalam mengubah Model LLM yang akan digunakan dalam proses analisis. Pada Modal *New Analysis* terdapat *Select Box* Model LLM yang dapat diubah oleh *user* sesuai dengan model LLM yang ingin digunakan. Saat *user* menekan *Select Box* tersebut, maka sistem akan menampilkan pilihan model LLM yang tersedia pada sistem. *User* dapat langsung memilih model LLM yang diinginkan dan *state* Model LLM akan diperbarui dan dikirim ke *backend* ketika *user* menekan tombol “*Generate Insights*” pada modal.

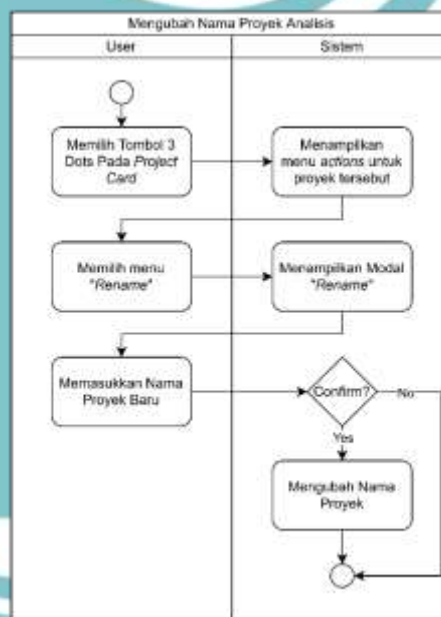
#### 4.2.2.2.6 Membatalkan Proses Analisis



Gambar 4.10 Activity Diagram Membatalkan Proses Analisis

Gambar 4.10 menunjukkan salah satu proses yang dapat dilakukan pada sistem yaitu pembatalan proses analisis. Proses ini berjalan pada saat *user* menekan tombol “Cancel” pada *Processing View* pada saat proses analisis sedang berlangsung. Tombol tersebut akan memberi sinyal ke *backend* untuk menghentikan proses analisis yang berlangsung.

#### 4.2.2.2.7 Mengubah Nama Proyek Analisis



Gambar 4.11 Activity Diagram Mengubah Nama Proyek Analisis

Gambar 4.11 menunjukkan alur proses pengubahan Nama Proyek Analisis pada *Website*. Proses ini dilakukan dengan menekan tombol *Triple Dots* pada *Project Card* yang ada pada list proyek yang akan membuka menu aksi yang berisi *Rename* dan *Delete*. Ketika user memilih *Rename* maka akan muncul Modal untuk

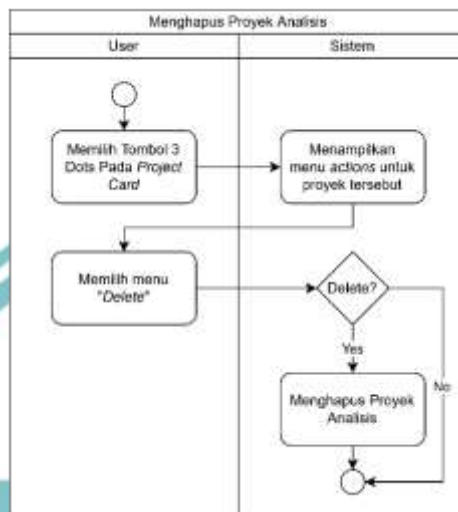


#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

mengubah nama dari proyek tersebut. Setelah itu, *user* harus menekan tombol “*Save Change*” untuk menyimpan perubahan yang dilakukan.

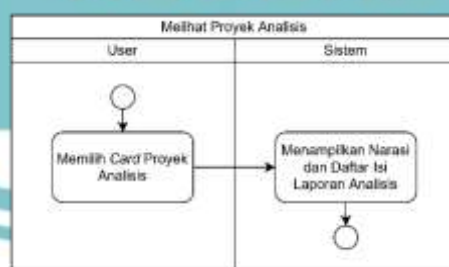
#### 4.2.2.2.8 Menghapus Proyek Analisis



Gambar 4.12 Activity Diagram Menghapus Proyek Analisis

Gambar 4.12 menunjukkan alur proses penghapusan Proyek Analisis pada *Website*. Proses ini dilakukan dengan menekan tombol *Triple Dots* pada *Project Card* yang ada pada list proyek yang akan membuka menu aksi yang berisi *Rename* dan *Delete*. Ketika user memilih *Delete* maka akan muncul Dialog untuk menghapus data proyek tersebut. Setelah itu, *user* harus menekan tombol “*Ok*” untuk benar benar menghapus proyek tersebut.

#### 4.2.2.2.9 Melihat Proyek Analisis



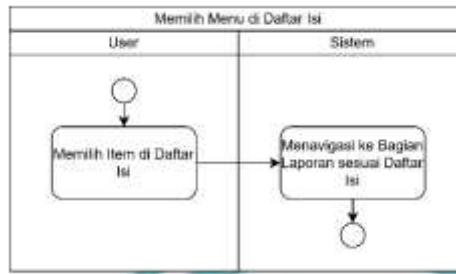
Gambar 4.13 Activity Diagram Melihat Proyek Analisis

Gambar 4.13 menunjukkan alur sederhana ketika user akan melihat proyek analisis yang sudah selesai dibuat. *User* hanya perlu menekan *Card* proyek analisis yang ada pada halaman *Home* maupun *Projects*. Setelah itu, sistem akan menampilkan narasi Laporan Analisis beserta Daftar Isinya pada halaman hasil Laporan Analisis.

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

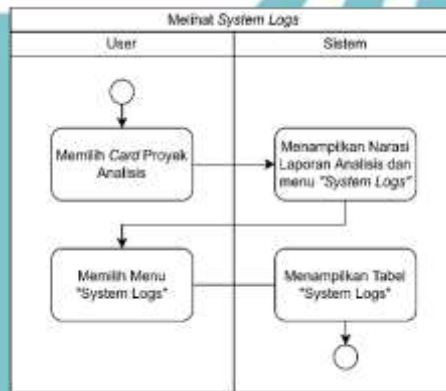
#### 4.2.2.2.10 Memilih Menu di Daftar Isi



Gambar 4.14 *Activity Diagram* Memilih Menu di Daftar Isi

Gambar 4.14 menunjukkan alur yang terjadi pada menu daftar isi di halaman hasil Laporan Analisis. Ketika *user* menekan salah satu item di daftar isi, sistem akan langsung mengalihkan *user* ke bagian laporan yang dituju.

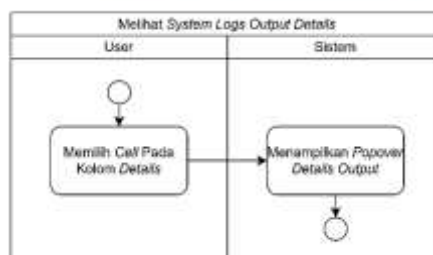
#### 4.2.2.2.11 Melihat System Logs



Gambar 4.15 *Activity Diagram* Melihat System Logs

Gambar 4.15 menunjukkan alur yang terjadi antara *user* dan sistem pada saat mengakses *system logs*. Untuk mengakses log proses analisis, *user* harus membuka halaman hasil laporan analisis lalu mengakses menu *system logs* pada bagian atas laporan analisis. Setelah itu, sistem akan menampilkan tabel berisi *process logs* dari laporan analisis tersebut.

#### 4.2.2.2.12 Melihat System Logs Output Details





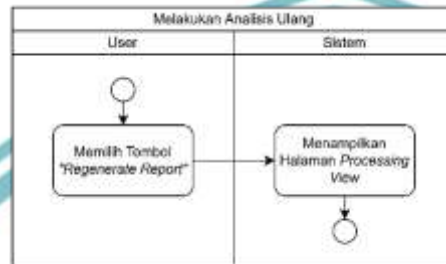
#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Gambar 4.16 *Activity Diagram* Melihat *System Logs Output Details*

Gambar 4.16 menunjukkan salah satu fitur yang dapat diakses oleh *user* pada tabel *system logs* dengan cara menekan salah satu *cell* pada kolom *details*. Sistem akan menampilkan sebuah *popover* yang berisi *details output* pada proses yang dipilih.

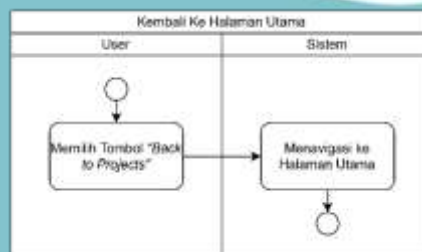
#### 4.2.2.2.13 Melakukan Analisis Ulang



Gambar 4.17 *Activity Diagram* Melakukan Analisis Ulang

Gambar 4.17 menunjukkan alur dari fitur *Regenerate Report* yang terdapat pada halaman hasil laporan analisis. Alur dimulai ketika *user* menekan tombol “*Regenerate Report*” yang akan membuat sistem menjalankan ulang proses analisis berdasarkan data yang dimasukkan *user* pada saat pertama kali membuat proyek analisis tersebut.

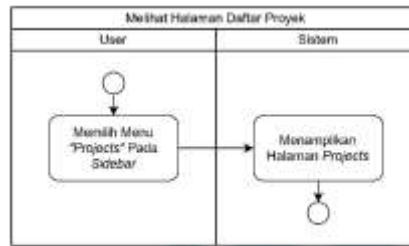
#### 4.2.2.2.14 Kembali Ke Halaman Utama



Gambar 4.18 *Activity Diagram* Kembali ke Halaman Utama

Gambar 4.18 menunjukkan alur singkat pada saat *user* menekan tombol *Back to Projects* yang terdapat pada halaman hasil laporan analisis. Sistem kemudian akan menavigasi *user* ke halaman *Home*.

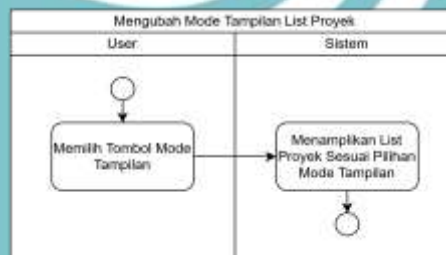
#### 4.2.2.2.15 Melihat Halaman Daftar Proyek



Gambar 4.19 *Activity Diagram* Melihat Halaman Daftar Proyek

Gambar 4.19 menunjukkan alur proses yang terjadi pada saat *user* melihat halaman daftar proyek. Alur dimulai ketika *user* menekan menu “*Projects*” pada *Sidebar* website. Proses tersebut akan membuat sistem menavigasi *user* ke halaman *Projects* yang berisi statistik dari seluruh proses analisis yang terjadi beserta list keseluruhan daftar proyek analisis yang ada.

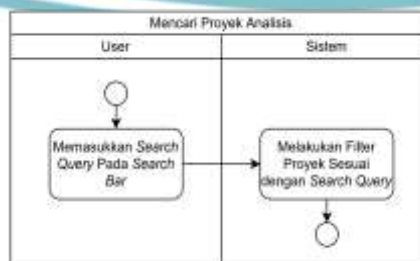
#### 4.2.2.2.16 Mengubah Mode Tampilan List Proyek



Gambar 4.20 *Activity Diagram* Mengubah Mode Tampilan List Proyek

Gambar 4.20 menunjukkan proses yang terjadi ketika *user* ingin mengubah mode tampilan list yang ada pada halaman *Projects*. Terdapat 2 mode yang tersedia, yaitu: *List* dan *Grid*. *User* hanya perlu memilih salah satu pilihan mode yang tersedia, lalu sistem akan memperbarui *state* mode list proyek dan menampilkan daftar proyek sesuai dengan mode yang dipilih.

#### 4.2.2.2.17 Mencari Proyek Analisis



Gambar 4.21 *Activity Diagram* Mencari Proyek Analisis

Gambar 4.21 menggambarkan alur yang terjadi ketika *user* melakukan pencarian proyek analisis yang ada pada halaman *Projects*. Proses ini bermula saat *user* memasukkan *search query* pada *search bar* yang tersedia di halaman *Projects*. Sistem kemudian akan memperbarui *search query state* dan melakukan filter proyek sesuai dengan *query*. Selanjutnya, list yang ditampilkan pada halaman *Projects* akan muncul sesuai dengan *search query* yang dimasukkan.



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

#### Hak Cipta :

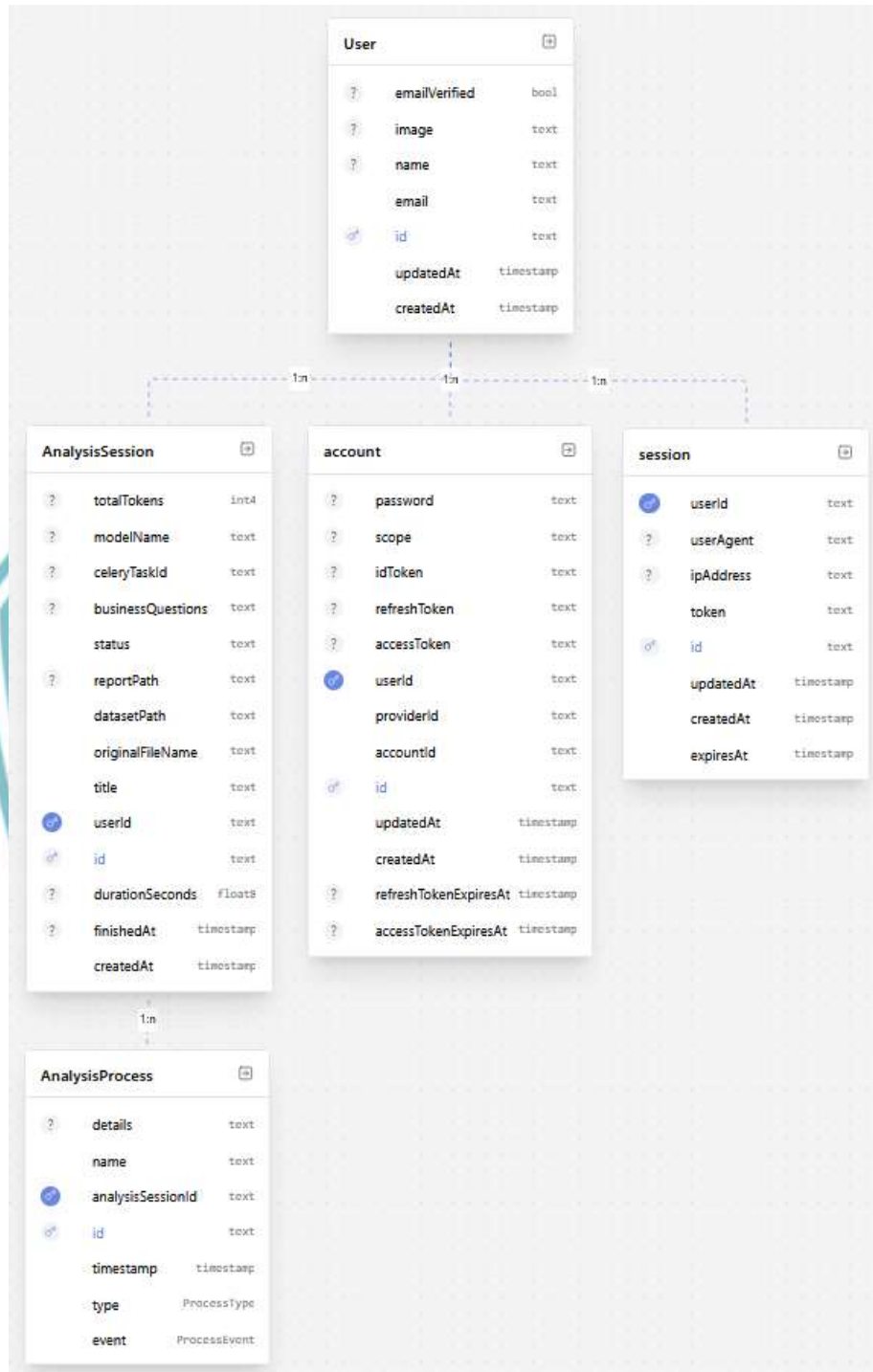
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

### 4.2.2.3 Perancangan Database



Gambar 4.22 Entity Relationship Diagram

Sistem ini dikelola menggunakan *Prisma ORM* dengan desain yang berpusat pada keamanan pengguna dan pelacakan proses analisis data secara *real-time*. Struktur datanya terbagi menjadi dua bagian utama: Autentikasi dan Analisis.



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

#### 4.2.2.3.1 Tabel untuk Mengelola Autentikasi

Sesuai dengan kebutuhan standar *Better Auth*, entitas *User* menjadi hal utama dalam manajemen pengguna. Relasi ini terbagi menjadi tiga tabel pendukung:

1. *User*: Menyimpan data dasar profil seperti nama, email, dan status verifikasi.
2. *Account*: Mengelola kredensial yang terhubung ke satu *User* via *userId*.
3. *Session*: Mencatat *session* aktif pengguna di berbagai perangkat, lengkap dengan informasi *userAgent* dan *ipAddress* untuk keamanan tambahan.

#### 4.2.2.3.2 Tabel untuk Mengelola Analisis

Bagian ini berperan dalam mengelola fungsionalitas utama dari aplikasi, yaitu pengolahan dataset CSV dan interaksi dengan *agent AI*:

1. *AnalysisSession* (1:n dengan *User*): Setiap analisis yang dibuat oleh pengguna disimpan di sini. Tabel ini berperan sebagai pusat metadata proyek, menyimpan *file path* (*datasetPath* dan *reportPath*), informasi model AI yang digunakan (*modelName*), hingga status pekerjaan yang sedang berjalan di *background worker* (*Celery*). Tabel ini juga digunakan untuk mencatat total token dan waktu pemrosesan yang dibutuhkan sistem selama melakukan analisis.
2. *AnalysisProcess* (1:n dengan *AnalysisSession*): Entitas ini adalah entitas *logs* yang mencatat setiap proses yang terjadi dalam satu proyek analisis. Di sinilah disimpan *log* proses dari *task Celery* serta output dari proses tersebut.

#### 4.2.2.4 Wireframe Antarmuka

*Wireframe* antarmuka adalah desain dari visual web yang dikembangkan menggunakan *framework Next.js* berbasis *TypeScript*. *Wireframe* pada penelitian ini mencakup beberapa halaman inti, yaitu:

1. Halaman Autentikasi: Antarmuka untuk *sign in* dan *sign up* akun.



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.23 Wireframe Halaman Sign In



Gambar 4.24 Wireframe Halaman Sign Up

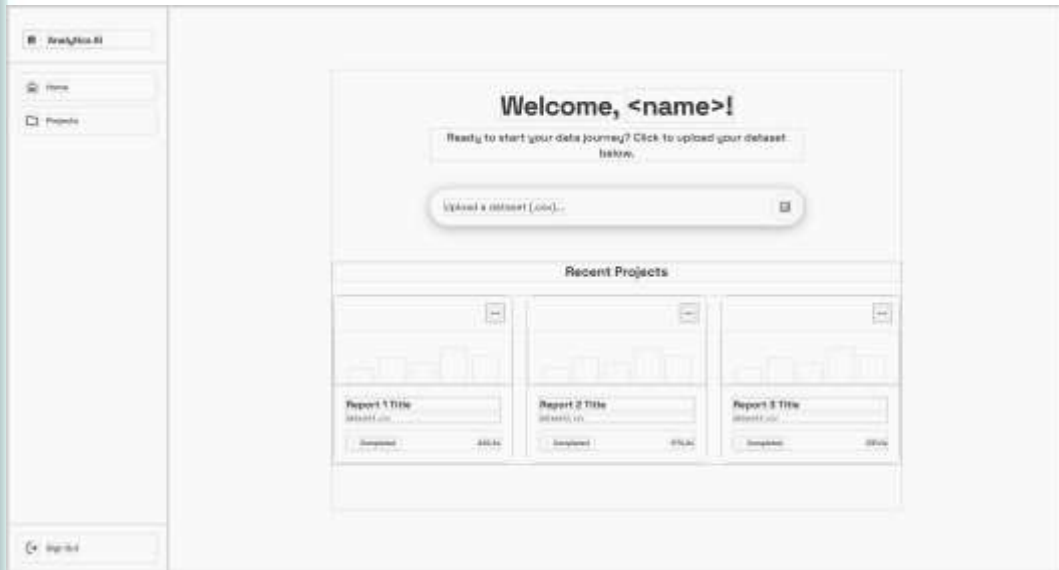
2. Halaman *Home*: Halaman ini menampilkan menu untuk *upload* dataset dan memulai analisis serta 3 hasil analisis terakhir yang telah dilakukan.



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

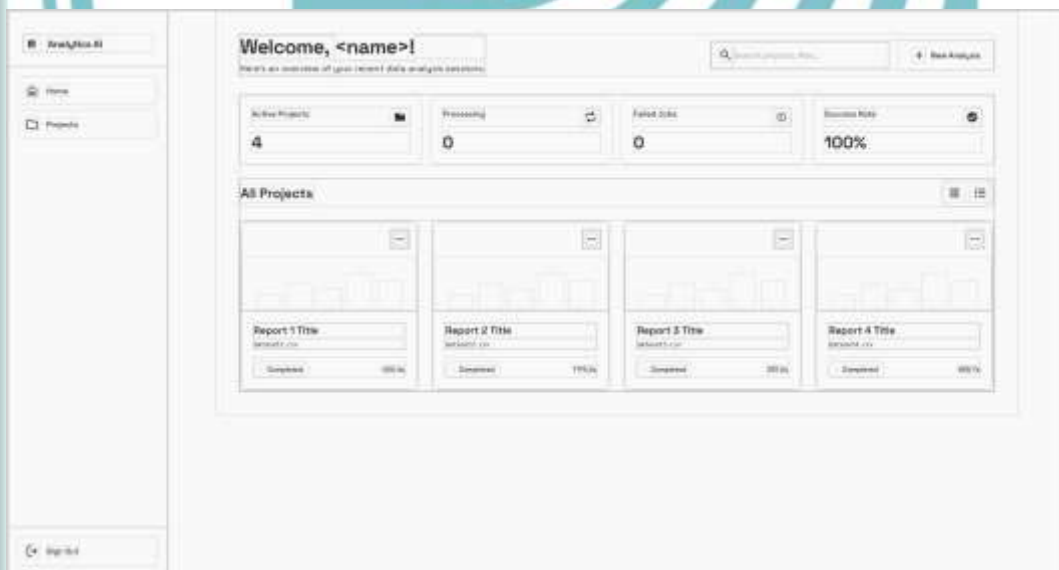
### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.25 Wireframe Halaman Home

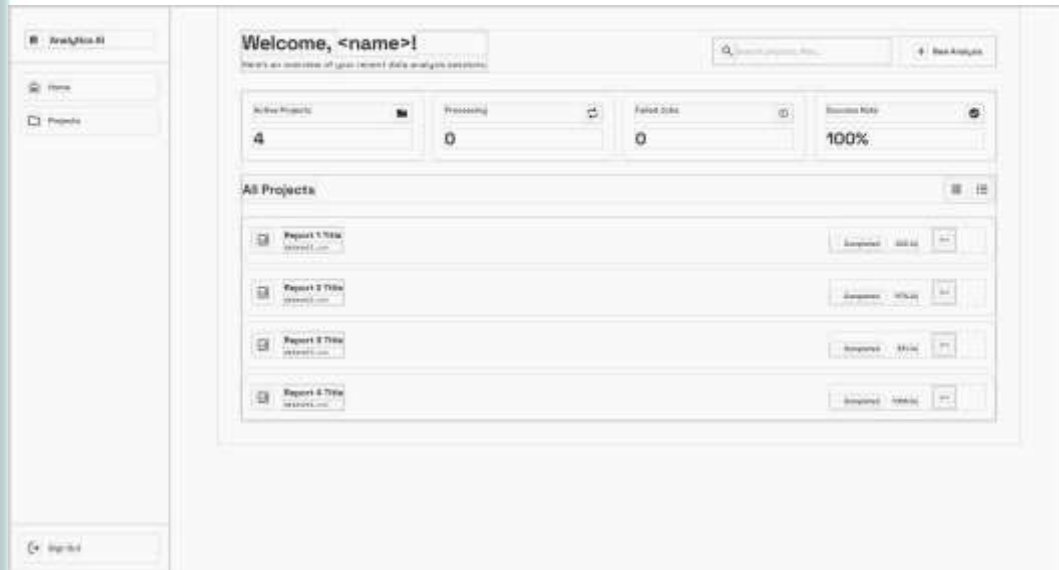
3. Halaman Daftar Proyek: Halaman yang menampilkan daftar proyek analitik pengguna dalam bentuk *grid* atau tabel, beserta status tiap proyek.



Gambar 4.26 Wireframe Halaman Daftar Proyek Grid

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.27 Wireframe Halaman Daftar Proyek List

4. *New Analysis Modal*: Halaman yang digunakan untuk mengunggah dataset penjualan dan memuat area kolom teks sebagai sarana input kueri pertanyaan analitik bisnis.



Gambar 4.28 Wireframe Modal New Analysis

5. *Halaman Processing*: Halaman yang menampilkan *timeline* proses analitik yang berjalan pada sistem saat ini beserta dengan rincian *logs* proses yang terjadi.



## Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.31 Wireframe Halaman Laporan Analitik



Gambar 4.32 Wireframe Tabel Logs Halaman Laporan Analitik



Gambar 4.33 Wireframe Details Tabel Logs

### 4.3 Implementasi Sistem

Hasil dari perancangan sistem yang dilakukan kemudian diimplementasikan ke dalam sistem yang dapat digunakan oleh *end-user*. Pada penelitian ini, implementasi sistem dibagi menjadi dua, yaitu sistem *multi-agent* dan *web dashboard*. Kedua implementasi tersebut nantinya akan saling terintegrasi melalui *Application Program Interface (API)* yang dibangun pada sistem *multi-agent* sebagai jembatan antara *Web Dashboard* dan *Agentic AI*.

#### 4.3.1 Implementasi Sistem *Multi-Agent*

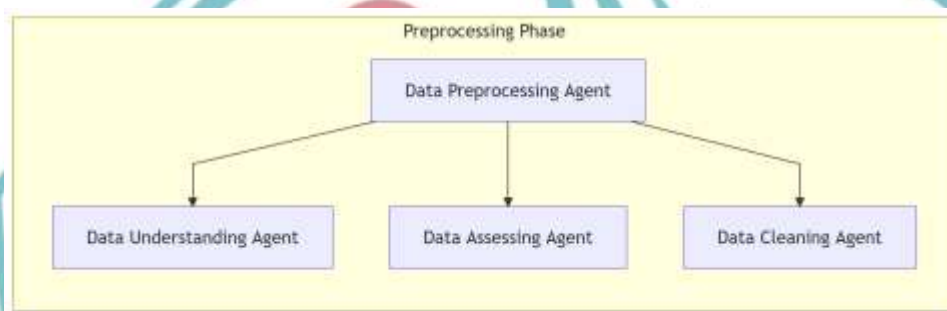
Pada penelitian ini, proses otomatisasi analitik data dilakukan menggunakan pendekatan *Agentic AI multi-agent* yang disusun menjadi suatu *pipeline* analitik data untuk menghasilkan suatu laporan analisis berdasarkan dataset dan pertanyaan yang diinputkan user. Sistem *Multi-Agent* ini dibangun menggunakan *framework Google Agent Development Kit* yang akan menunjang sistem dalam proses *Metadata-Centric Context Management* dan *Function Calling*. Dengan membangun suatu sistem *multi-agent* dengan kapabilitas tersebut, proses analitik data diharapkan dapat dilakukan secara otomatis dan efisien.

##### 4.3.1.1 Implementasi Arsitektur *Multi-Agent*

Berdasarkan arsitektur *multi-agent* yang telah dirancang pada bagian perancangan, *pipeline* yang dibangun memiliki 3 fase utama, yaitu: *Preprocessing*, *Exploratory Data Analysis*, dan *Explainer*. Ketiga fase tersebut akan dirangkai dalam suatu

*sequential pipeline* yang memastikan seluruh fase tereksekusi oleh sistem sehingga menghasilkan output yang diinginkan. Masing-masing *agent* yang terlibat dalam arsitektur *multi-agent* yang dibangun diberi instruksi sesuai dengan perannya. Instruksi tersebut dapat dilihat pada Lampiran 1, dimana instruksi diadaptasi dari struktur *prompt* yang dipaparkan oleh Nikolakopoulos et al. Struktur *prompt* tersebut meliputi *objective definition*, *context provision*, *instruction clarity*, *examples inclusion*, *error handling*, dan *message formatting* (Nikolakopoulos et al., 2025).

#### 4.3.1.1.1 Fase *Preprocessing*



Gambar 4.34 Arsitektur *Agentic AI* Fase *Preprocessing*

Pada fase *Preprocessing*, fokus utamanya adalah membersihkan dataset yang diinputkan oleh *user*. Dalam prosesnya, setiap *agent* yang terlibat diberi instruksi untuk membersihkan dataset sesuai dengan kebutuhan analisis yang didasarkan pada *query* pertanyaan *user*. Hal ini sejalan dengan prinsip “*high-quality data is fit-for-purpose*” yang ditegaskan oleh Li et al. pada penelitian *AutoDCWorkflow*. Dataset yang masuk kedalam fase ini akan secara otomatis dibersihkan dengan bantuan *function calling* yang dilakukan oleh *agent-agent* yang terlibat. Fase ini terdiri dari 1 *Orchestrator Agent* dan 3 *LLM Agents* yang akan bekerja sama dalam membersihkan dataset yang masuk.

##### a. *Data Preprocessing Agent (Orchestrator)*

Sebagai *Orchestrator Agent*, *agent* ini akan mengatur jalannya fase *Preprocessing* yang berjalan. Pada saat *user* memulai proses analisis, *agent* ini akan menerima input *user* berupa dataset dan *business questions* lalu memecah *task* yang akan dilakukan menjadi beberapa tahap. *Tasks* tersebut kemudian akan dieksekusi secara bergiliran dengan memanggil *sub-agent* yang sesuai untuk menyelesaikan *task* tersebut. Setelah *orchestrator agent* menerima informasi bahwa seluruh *task* telah



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

tereksekusi, *agent* ini akan memberi sinyal bahwa fase *preprocessing* telah selesai dan dapat dilanjut ke fase EDA.

#### b. *Data Understanding Agent*

Tahap pertama yang akan dilakukan pada fase *preprocessing* adalah *Data Understanding*. Tahap ini bertujuan agar sistem *Agentic AI* dapat memahami dataset sebelum melakukan proses analitik data lebih lanjut. Hal ini menjadi krusial karena hasil dari *data understanding* akan menjadi panduan *agent* lain dalam mengeksekusi proses analitik sesuai dengan konteks data. *Agent* ini akan memanggil *tool get\_understanding\_report\_tool()* yang menghasilkan informasi tentang nama, tipe data, dan sampel data pada setiap kolom yang ada. Informasi tersebut diadaptasi mengikuti penelitian yang dilakukan oleh Nikolakopoulos et al. yang menggunakan informasi tersebut untuk membantu LLM dalam memahami data sehingga akan memudahkan proses analisis. Pada tahap ini *agent* akan memberikan output berupa laporan *Data Understanding* yang berisi rangkuman pemahaman *agent* tentang dataset yang dimasukkan user. Contoh output dari *agent* ini dapat dilihat pada lampiran 2.

#### c. *Data Assessing Agent*

Setelah proses *Data Understanding* selesai dilakukan, *Orchestrator Agent* akan memerintahkan *Data Assessing Agent* untuk menilai kualitas dataset yang diinputkan oleh *user*. Proses *Data Assessing* ini akan dilakukan setidaknya dua kali pada saat sebelum *Data Cleaning* dan sesudah *Data Cleaning* untuk membantu *agent* menilai apakah proses *Data Cleaning* sudah cukup menuntaskan permasalahan pada dataset atau perlu dilakukan proses *Data Cleaning* lagi sehingga dataset yang dihasilkan merupakan dataset paling optimal. Dalam melakukan proses nya, *Agent* ini akan dibekali dengan 3 *tools* yaitu *get\_data\_state\_list()*, *get\_assessment\_report\_tool()*, dan *run\_sql\_tool()* sehingga proses *data assessing* dapat dilakukan secara bertahap. Pada *get\_assessment\_report\_tool()*, *agent* akan mendapat informasi terkait profil setiap kolom (tipe data, jumlah null, jumlah *unique*, sampel, min, max, mean, dan std) yang diadaptasi dari penelitian *LightVA* oleh Zhao et al. Selain itu, *tool* tersebut juga akan menghasilkan data rasio *null value* dan *duplicated row* yang akan digunakan dalam menghitung *data quality score*. Setelah rangkaian proses *Data Assessing* selesai, *agent* ini akan

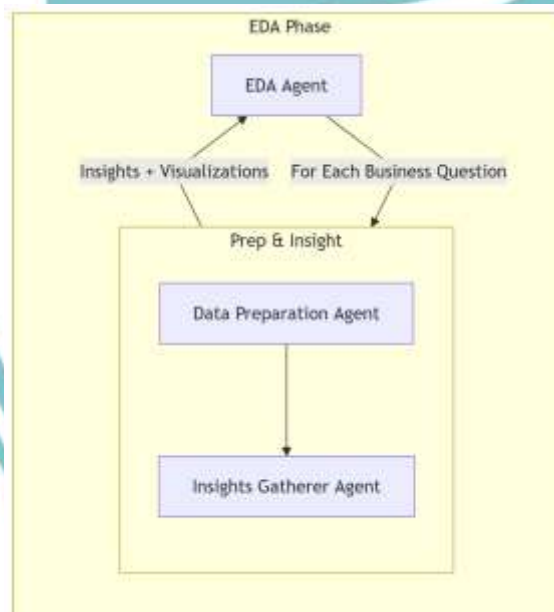


menghasilkan output *markdown* berisi list masalah yang ada pada dataset dan juga rekomendasi pembersihan dataset untuk menghilangkan permasalahan kualitas data yang ada. Contoh output dari *data assessing agent* dapat dilihat pada lampiran 3 (*pre-cleaning*) dan 5 (*post-cleaning*).

#### d. Data Cleaning Agent

Proses ini merupakan bagian yang krusial dalam fase *Preprocessing*. *Data Cleaning Agent* akan berperan dalam proses pembersihan dataset mengikuti hasil penilaian kualitas dataset yang dilakukan oleh *Data Assessing Agent*. Proses pembersihan data dilakukan dengan pemanggilan *tool run\_sql\_tool()* yang mengeksekusi *sql query* untuk meniadakan *data quality issues* yang ada pada dataset. Setelah proses pembersihan data selesai, *agent* ini akan menghasilkan output berupa *data cleaning report* yang berisi *execution logs* yang dilakukan untuk membersihkan data. Informasi ini akan berguna untuk memberikan pemahaman *orchestrator* terkait upaya apa saja yang sudah dilakukan pada proses *Data Cleaning*. Contoh output dari *data cleaning agent* dapat dilihat pada lampiran 4.

#### 4.3.1.1.2 Fase Exploratory Data Analysis (EDA)



Gambar 4.35 Arsitektur *Agentic AI* Fase *EDA*

Pada fase *Exploratory Data Analysis (EDA)*, sistem *Agentic AI* akan melakukan serangkaian proses yang bertujuan untuk menghasilkan informasi yang akan dimasukkan dalam laporan analisis. Agar proses ini tetap terarah dan sesuai dengan

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

harapan *user*, proses ini diberikan konteks berupa *business questions* yang diinputkan *user* di awal saat membuat proyek analisis baru. Fase EDA ini akan dikomandoi oleh *Orchestrator Agent* yaitu *EDA Agent*. *Agent* tersebut akan memecah *business questions* menjadi beberapa tahap yang akan dieksekusi secara bergiliran dengan memerintahkan *Prep & Insight Sequence* untuk menggali informasi pada dataset sesuai dengan pertanyaan *user*.

#### a. *EDA Agent (Orchestrator)*

Pada fase EDA, *Agent* ini akan mengatur pembagian *task* yang akan diberikan ke *Prep & Insight Sequence* sehingga proses analisis dapat terarah sesuai tujuan *user*. *EDA Agent* akan memastikan seluruh *business questions* dapat menghasilkan *insights* yang akan dimasukkan ke dalam laporan analisis. Sebagai output, *agent* ini akan menghasilkan JSON sesuai dengan skema yang telah ditentukan sehingga hasilnya akan terstruktur dan dapat dipahami oleh *Explainer Agent*.

#### b. *Data Preparation Agent*

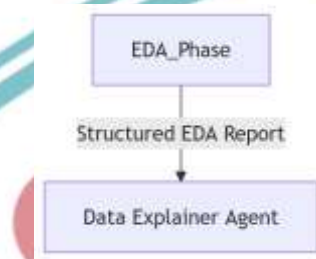
Salah satu bagian dari *Prep & Insights Sequence* adalah *Data Preparation Agent*. *Agent* ini akan berperan dalam menyiapkan tabel data pada setiap *insight* yang akan dihasilkan. Tujuannya adalah agar setiap proses pencarian informasi dapat dilakukan sefleksibel mungkin tanpa mengubah dataset awal yang akan digunakan kembali pada proses penggalian informasi untuk pertanyaan *user* lainnya. Proses *information gathering* yang dilakukan meliputi agregasi, pivot, filter, dll. Dalam melakukan prosesnya, *agent* ini dibekali dengan *tools run\_sql\_tool()* untuk melihat atau memodifikasi data, dan juga *get\_data\_state\_list()* untuk mendapatkan list tabel data yang telah dihasilkan selama proses analisis. Setelah selesai, *agent* ini akan menghasilkan *key* beserta deskripsi dari tabel data yang dihasilkan sehingga *Insights Gatherer Agent* dapat melakukan proses yang optimal sesuai dengan tujuan analisis. Contoh output *key-description* dari proses ini dapat dilihat pada lampiran 6.

#### c. *Insights Gatherer Agent*

Proses selanjutnya dari *Prep & Insights Sequence* adalah *Insights Gatherer Agent*. Pada tahap ini, *agent* akan menggali informasi yang telah disediakan oleh *Data Preparation Agent* untuk menunjang narasi analisis yang akan dicantumkan pada laporan analisis akhir. Selain itu, *agent* ini juga akan melakukan generasi plot

sebagai visualisasi dari *insights* yang dihasilkan. Oleh karena itu, *agent* ini dibekali dengan *run\_sql\_tool()* dan *plot\_tool()* untuk menghasilkan *insights* yang sesuai dengan kebutuhan *user*. Sebagai output, *agent* ini akan menghasilkan teks berisi narasi *insight*, bukti *insight*, serta nama *file* dari plot yang dihasilkan. Contoh output dari *agent* ini dapat dilihat pada lampiran 7.

#### 4.3.1.1.3 Fase *Explainer*



Gambar 4.36 Arsitektur *Agentic AI* Fase *Explainer*

Fase terakhir dalam *pipeline Agentic AI* ini adalah fase *Explainer*. Tidak seperti dua fase sebelumnya, fase ini hanya melibatkan 1 *agent* yaitu *Data Explainer Agent*. Fase ini dikhususkan untuk mengolah informasi yang dihasilkan pada fase EDA menjadi suatu laporan analisis yang siap disajikan. Untuk meminimalisir halusinasi yang biasa terjadi pada LLM, *agent* ini dibekali dengan *get\_data\_state\_list()* dan *list\_output\_files\_tool()* yang akan memberikan informasi list tabel data dan file yang dihasilkan oleh *agent agent* sebelumnya. Dengan *business questions*, list tabel, dan list file menjadi konteks, *agent* ini akan menghasilkan laporan analisis yang terarah sesuai dengan tujuan awal proses analisis.

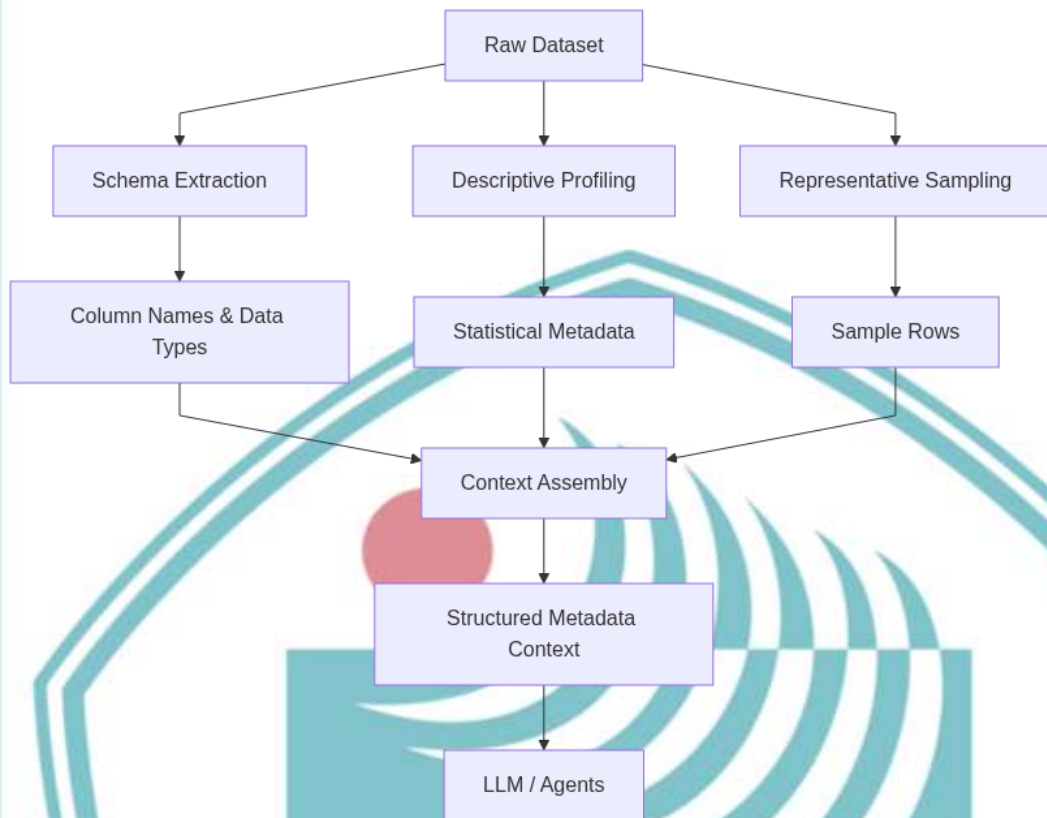
#### 4.3.1.2 Implementasi *Metadata-Centric Context Management*

Implementasi *Metadata-Centric Context Management* difokuskan pada upaya mengatasi keterbatasan *context window* pada *Large Language Model* (LLM) ketika memproses dan menganalisis dataset berukuran besar yang dapat mencapai jutaan baris. Pendekatan ini didasarkan pada prinsip bahwa LLM tidak perlu mengakses seluruh *raw data* untuk memahami konteks dari data tersebut, melainkan cukup memanfaatkan representasi data yang dapat menggambarkan karakteristik dari dataset.



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.37 Flowchart Metadata-Centric Context Management

Implementasi manajemen konteks ini dilakukan pada tahap awal *pipeline* analisis, khususnya pada *Data Understanding Agent* dan *Data Assessing Agent* melalui beberapa tahapan:

#### 1. Schema Extraction

Sistem memanfaatkan *DuckDB* untuk membaca dataset secara efisien. Informasi pertama yang diekstrak adalah nama setiap kolom beserta tipe datanya. Hal ini krusial agar *agent* mengetahui parameter apa saja yang tersedia untuk dilakukan proses analisis.

```

1 schema_res = engine.execute_query(f'DESCRIBE "{table_name}"')
2 data_info = {}
3 if schema_res.get("rows"):
4 for row in schema_res["rows"]:
5 data_info[row[0]] = {"type": row[1]}

```

Gambar 4.38 Proses Schema Extraction pada Tool

#### 2. Descriptive Profiling

Selain nama kolom, *agent* melakukan *function calling* untuk menghasilkan metadata statistik yang merangkum kondisi dataset secara keseluruhan tanpa memasukkan seluruh baris data yang ada. Informasi ini berisi Total baris dan kolom, Jumlah *missing values/nulls* pada setiap kolom, Rentang nilai (minimum dan maksimum) untuk data numerik, dan *unique values*.

```

1 summarize_res = engine.execute_query(f'SUMMARIZE "{table_name}"')
2 data_description = {
3 "count": {}, "mean": {}, "std": {}, "min": {}, "25%": {}, "50%": {},
4 "75%": {}, "max": {}, "unique": {}, "null_percentage": {}
5 }
6 if summarize_res.get("rows"):
7 for row in summarize_res["rows"]:
8 col_name = row[0]
9 data_description["count"][col_name] = row[10]
10 data_description["mean"][col_name] = row[5]
11 data_description["std"][col_name] = row[6]
12 data_description["min"][col_name] = row[2]
13 data_description["25%"][col_name] = row[7]
14 data_description["50%"][col_name] = row[8]
15 data_description["75%"][col_name] = row[9]
16 data_description["max"][col_name] = row[3]
17 data_description["unique"][col_name] = row[4]
18 data_description["null_percentage"][col_name] = row[11]

```

Gambar 4.39 Proses *Descriptive Analysis* pada Tool

### 3. Representative Sampling

Untuk memberikan gambaran mengenai *value* asli pada dataset (misalnya format tanggal YYYY-MM-DD atau penulisan nilai desimal), sistem mengambil *value* 5 baris dari dataset. Sampel yang sangat kecil ini diubah menjadi *JSON* untuk memberikan informasi pada *agent*.

```

1 sample_res = engine.execute_query(f'SELECT * FROM "{table_name}" USING SAMPLE 3')
2 sample_data = {}
3 if sample_res.get("columns") and sample_res.get("rows"):
4 columns = sample_res["columns"]
5 for idx, row in enumerate(sample_res["rows"]):
6 sample_data[str(idx)] = {col: val for col, val in zip(columns, row)}

```

Gambar 4.40 Proses *Sampling* pada Tool

### 4. Context Assembly

Setelah ketiga komponen tersebut (skema, statistik, dan sampel) selesai dikumpulkan, sistem merangkumnya menjadi sebuah teks *JSON*. Teks metadata inilah yang kemudian diproses oleh *data understanding agent* dan menghasilkan *report* yang di-*inject* ke dalam *Prompt* untuk diteruskan ke *agent* selanjutnya.

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Melalui implementasi *context management* tersebut, sistem dapat mengompresi dataset berukuran sangat besar menjadi teks yang hanya memakan beberapa ratus token. Hal ini akan secara efektif menyelesaikan dua masalah utama analitik data berbasis LLM yaitu mencegah terjadinya *out-of-memory* (melebihi batas token) dan menjaga agar *agent* tetap fokus pada tujuan analitik tanpa mengalami *lost-in-the-middle phenomenon*. Berbekal metadata ini, LLM memiliki pengetahuan yang cukup untuk menghasilkan *SQL Query* untuk proses *Function Calling* yang akan dieksekusi oleh *data engine*.

#### 4.3.1.3 Implementasi Mekanisme *Function Calling & Data Engine*

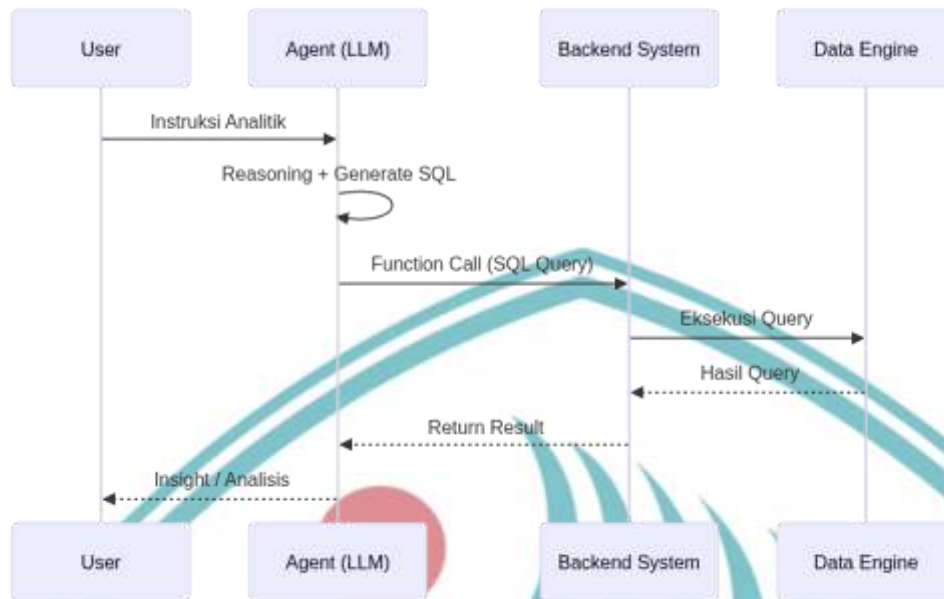
Implementasi mekanisme *Function Calling* dan *Data Engine* berfokus pada transformasi peran *Large Language Model* (LLM) dari sekadar *generator* teks menjadi *reasoning engine* yang mampu mengeksekusi dataset secara langsung melalui *tools* yang telah didefinisikan. Transformasi ini dilakukan melalui integrasi mekanisme *Function Calling* dengan *Data Engine* yang memungkinkan LLM berinteraksi secara langsung dengan proses pengolahan data.

##### 4.3.1.3.1 Implementasi Mekanisme *Function Calling*

Pada arsitektur sistem *Multi-Agent* ini, LLM tidak memproses data secara langsung di dalam memori model nya karena keterbatasan token dan potensi halusinasi. Sebagai solusi, sistem menggunakan *tools* yang dapat dieksekusi melalui mekanisme *Function Calling* sehingga *agent* pada sistem ini akan memiliki kapabilitas untuk mengeksekusi *tools* yang dapat secara langsung melakukan operasi yang dibutuhkan pada dataset.

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.41 *Sequence Diagram Mekanisme Function Calling*

Proses kerjanya adalah sebagai berikut:

1. *Agent* menerima instruksi dan informasi skema dataset.
2. Berdasarkan proses *reasoning* pada LLM, *agent* mengubah instruksi tersebut menjadi *SQL query* untuk melakukan proses analitik.
3. *Agent* selanjutnya akan memberikan respons berupa *function call request* beserta argumen *query* yang akan dieksekusi.
4. Sistem *backend* menerima *request* tersebut, mengeksekusi *function*, dan mengembalikan hasil eksekusi ke *agent* untuk dianalisis lebih lanjut.

Dalam implementasinya, sistem *agent AI* yang dikembangkan selama penelitian memiliki *tools* yang dapat digunakan oleh *agent* dalam melakukan proses analisis. Tabel 4.8 menunjukkan daftar *tools* tersebut.

Tabel 4.8 Tabel Daftar *Tools* untuk *Function Calling*

| Nama Tool                | Deskripsi                                                                               | Peran                                                                        |
|--------------------------|-----------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| <i>run_sql_tool</i>      | Mengeksekusi <i>SQL query</i> & memiliki opsi untuk menyimpan hasil ( <i>save_as</i> ). | Alat analisis utama untuk melakukan <i>SQL Query</i> selama proses analisis. |
| <i>plot_tool</i>         | Mengeksekusi <i>SQL query</i> sederhana dan merendernya menjadi grafik visualisasi.     | Alat untuk menghasilkan visualisasi data melalui <i>SQL Query</i> .          |
| <i>get_understanding</i> | Menghasilkan laporan profil                                                             | Menghasilkan metadata                                                        |



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|                                         |                                                                           |                                                                                   |
|-----------------------------------------|---------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| <code>_report_tool</code>               | data (skema, deskripsi statistik, sampel baris).                          | statistik sebagai konteks dalam memahami dataset.                                 |
| <code>get_assessment_report_tool</code> | Menganalisis kualitas kolom ( <i>null</i> , duplikasi).                   | Membantu <i>agent</i> mendapat informasi terkait kualitas data.                   |
| <code>get_data_state_list</code>        | Memberikan daftar tabel yang telah dibuat selama analisis.                | Membantu <i>agent</i> mengetahui tabel yang sudah ada dan siap di- <i>query</i> . |
| <code>list_output_files_tool</code>     | Memberikan daftar seluruh <i>files</i> yang telah dibuat selama analisis. | Membantu <i>agent</i> untuk mengetahui daftar visualisasi yang dibuat.            |

#### 4.3.1.3.2 Implementasi *Data Engine* (DuckDB)

Untuk mengeksekusi perintah yang dihasilkan oleh *agent* melalui *Function Calling*, sistem mengimplementasikan DuckDB sebagai alat untuk berinteraksi dengan data. DuckDB dipilih karena arsitektur *columnar in-process* yang dimilikinya sangat optimal untuk beban kerja analitik data dan mampu memproses dataset berukuran jutaan baris jauh lebih efisien dibandingkan Pandas.



Gambar 4.42 *Flow* Eksekusi dengan DuckDB

Implementasi DuckDB dalam sistem ini difokuskan pada:

#### 1. Koneksi dan Pembacaan Data

Sistem memanfaatkan kemampuan DuckDB untuk membaca langsung dari berbagai format file (seperti CSV atau Parquet) tanpa perlu memuat seluruh isi file ke dalam RAM komputer. Hal ini memungkinkan pemrosesan *out-of-core* untuk dataset berukuran sangat besar.

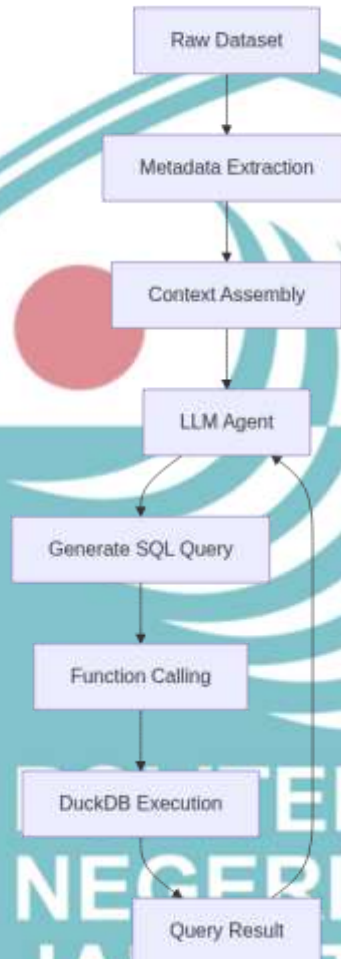
#### 2. Eksekusi *Query* LLM

*Query* SQL yang dihasilkan oleh *agent* dimasukkan ke dalam *function* yang mengimplementasikan DuckDB. *Function* ini kemudian akan menjalankan operasi analitik seperti *JOIN*, *GROUP BY*, *Window Functions*, dan *Filtering* secara efisien.

#### 3. Format Output

Hasil operasi dari DuckDB dikonversi ke dalam bentuk yang lebih hemat token, seperti *dictionary* Python, agar dapat diserap kembali oleh agen tanpa memakan terlalu banyak *context window*.

#### 4.3.1.3.3 Integrasi *Function Calling* dan DuckDB



Gambar 4.43 Flow Integrasi *Function Calling* dan DuckDB

Kombinasi antara *Function Calling* dan DuckDB dapat secara efektif menjembatani celah antara kemampuan *reasoning* dari LLM dan komputasi deterministik dari komputer. LLM bertugas mengubah tujuan analitik dari *user* untuk menghasilkan *query*, sementara DuckDB berperan sebagai alat yang akan memproses data mentah yang ada. Hasil akhirnya adalah sistem analitik yang mampu memahami bahasa alami manusia namun tetap dapat melakukan operasi secara deterministik.

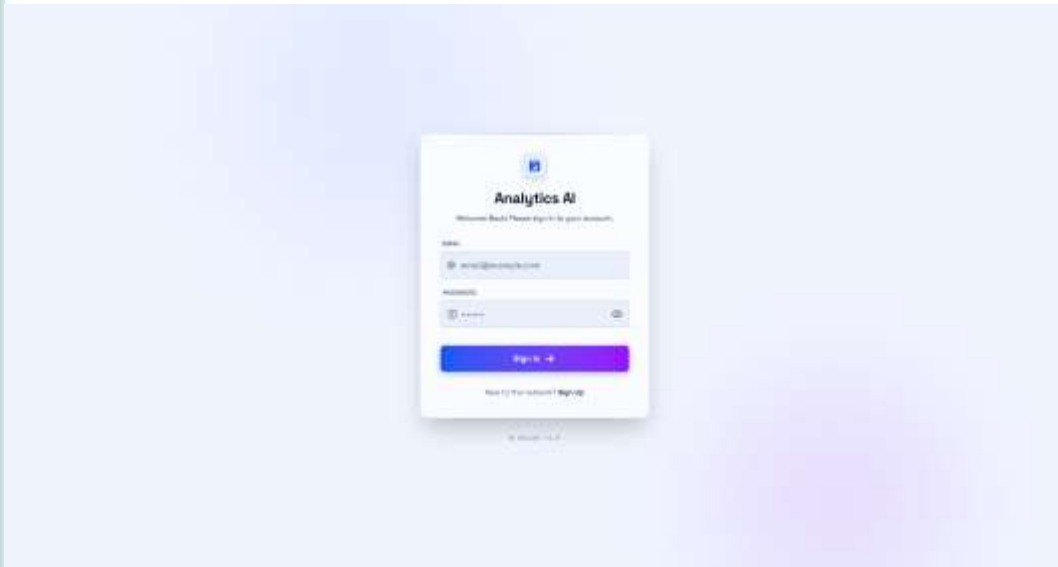
#### 4.3.2 Implementasi *Web Dashboard*

Untuk menjembatani antara *user* dengan sistem *agentic*, dikembangkan suatu *Web Dashboard* sebagai antarmuka dari sistem yang dibangun pada penelitian ini.



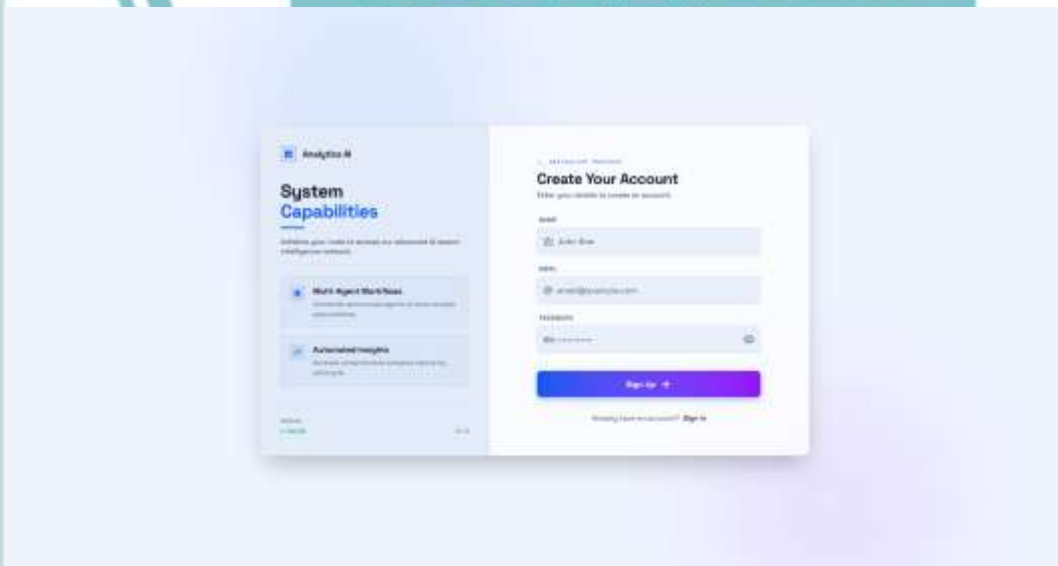
*Dashboard* yang dikembangkan ini diharapkan mampu menunjang seluruh fungsionalitas yang dapat dilakukan *user* terhadap sistem.

#### 4.3.2.1 Halaman *Sign In* dan *Sign Up*



Gambar 4.44 Halaman *Sign In*

Halaman ini menjadi gerbang utama *user* untuk mengakses *platform* data analisis ini. Pada saat *user* akan mengakses *platform*, *user* diharuskan untuk melakukan *Sign In* pada halaman *Sign In* dengan memasukkan *email* dan *password* akun yang telah dibuat. Jika *user* belum memiliki akun, maka *user* harus melakukan registrasi akun baru melalui halaman *Sign Up*.



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



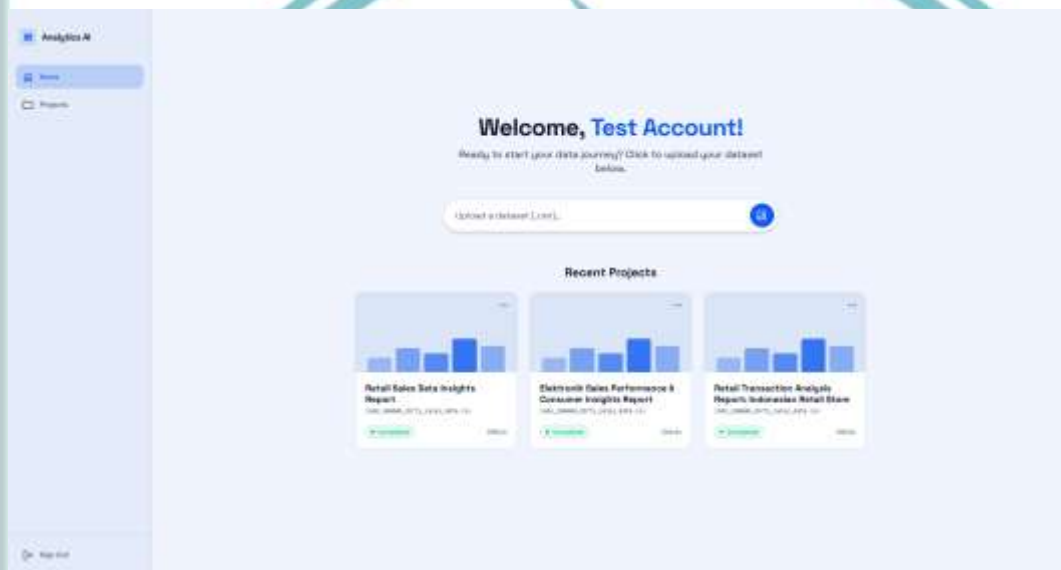
### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Gambar 4.45 Halaman *Sign Up*

Pada halaman *Sign Up*, *user* dapat melakukan pembuatan akun dengan memasukkan nama, *email*, dan *password* yang diinginkan *user*. Setelah itu, *user* harus menekan tombol “*Sign Up*” untuk menyimpan data registrasi akun yang telah dilakukan. Halaman ini juga menampilkan informasi terkait 2 kapabilitas utama sistem yaitu *Multi-Agent Workflows* dan *Automated Insights* yang menjadi keunggulan utama dari sistem yang dibangun.

#### 4.3.2.2 Halaman Utama (*Home*)

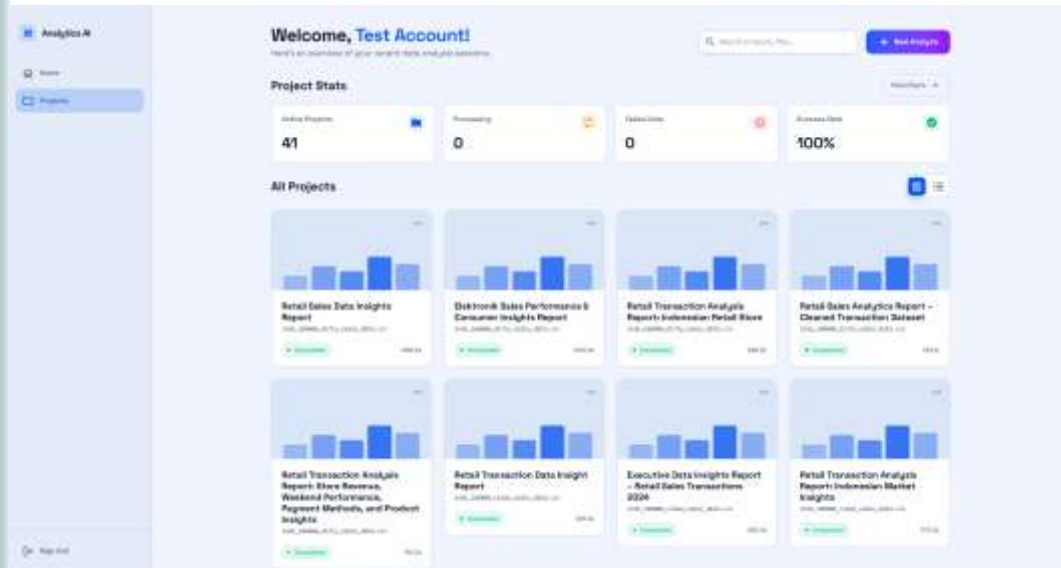


Gambar 4.46 Halaman Utama (*Home*)

Halaman ini merupakan halaman utama dari *platform* yang dibangun. Pada halaman *Home*, *platform* menampilkan tombol *Upload Dataset* yang akan memunculkan modal untuk memasukkan dataset beserta *business questions* yang akan diproses oleh sistem. Selain itu, terdapat list 3 proyek analisis terbaru yang dibuat oleh *user*. *User Interface* (UI) pada halaman *Home* dibuat sesederhana mungkin agar memudahkan *user* baru dalam menggunakan *platform* ini.

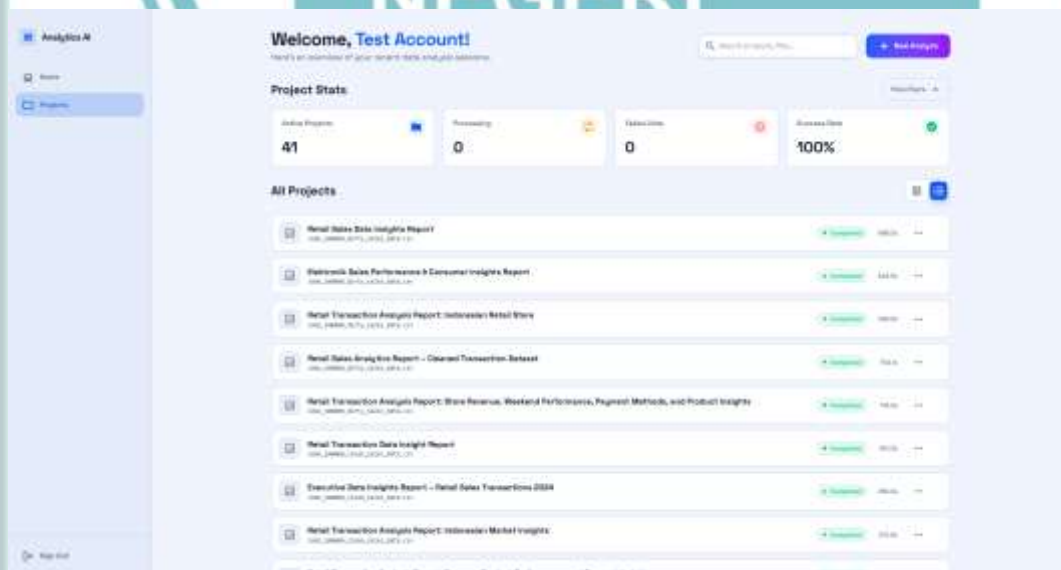


#### 4.3.2.3 Halaman Daftar Proyek



Gambar 4.47 Halaman Daftar Proyek *Grid View*

Pada halaman daftar proyek, *platform* akan menampilkan seluruh daftar proyek analisis yang dilakukan oleh *user*. Halaman ini juga akan menunjukkan statistik proyek yang telah dijalankan berdasarkan statusnya. Untuk mempermudah *user* dalam menemukan proyek analisis yang dimaksud, halaman ini juga menyediakan *search bar* yang memungkinkan *user* untuk mencari proyek analisis yang dimaksud berdasarkan judul proyek atau nama file dataset yang diunggah. Selain itu, *user* juga dapat melakukan pembuatan proyek analisis baru dengan menekan tombol “*New Analysis*” yang akan menampilkan modal *New Analysis*.

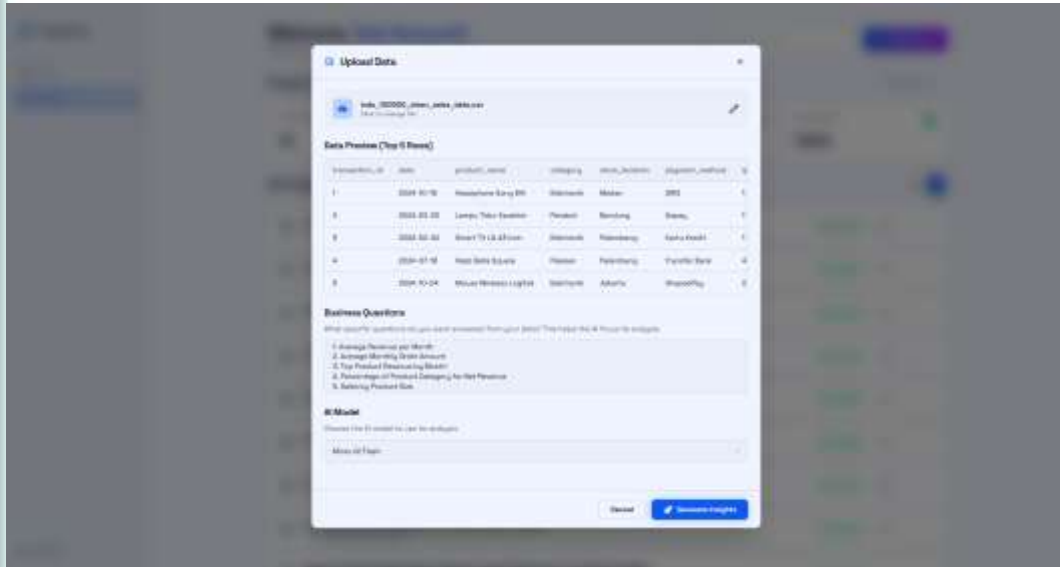




Gambar 4.48 Halaman Daftar Proyek *List View*

Halaman Daftar Proyek ini memiliki 2 jenis tampilan yaitu *List* dan *Grid* yang dapat diubah sesuai preferensi *user* dengan melakukan *switch* pada *toggle view type*.

#### 4.3.2.4 Modal *New Analysis*



Gambar 4.49 Modal *New Analysis*

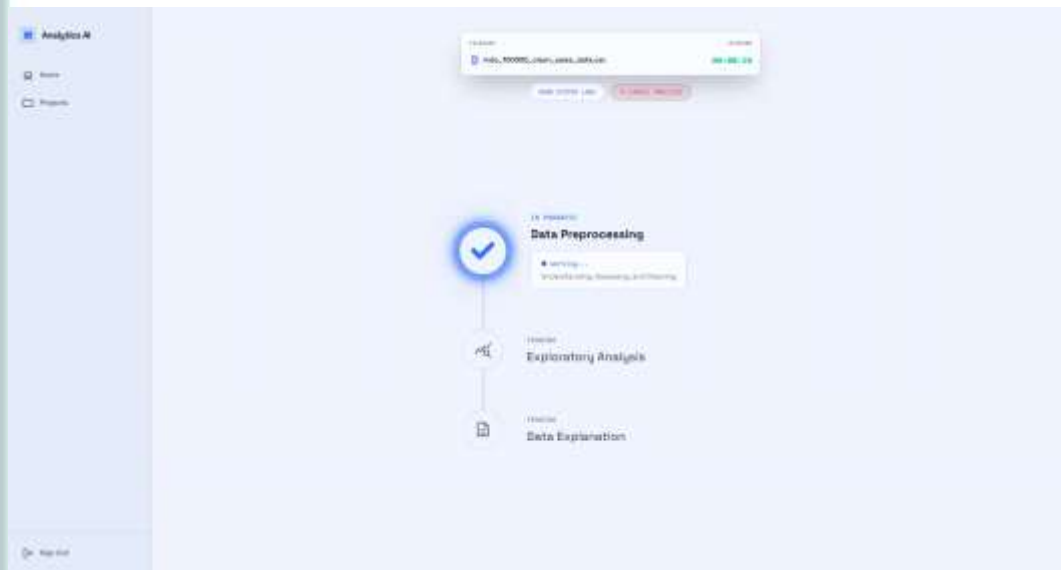
Pada saat *user* akan membuat proyek analisis baru, modal *New Analysis* ini akan ditampilkan oleh sistem. Pertama-tama, *user* harus memasukkan dataset yang akan diproses oleh sistem. Setelah dataset dimasukkan, maka akan muncul *Data Preview* yang memperlihatkan tabel berisi 5 baris teratas dari dataset yang dimasukkan. Dengan *Data Preview* ini, *user* dapat memastikan apakah data yang dimasukkan sudah benar atau belum. Selain *Data Preview*, sistem juga akan menampilkan bagian *Business Questions* yang berisi kolom untuk menuliskan pertanyaan *user* terhadap dataset yang dimasukkan. Pertanyaan inilah yang akan menjadi acuan *agentic AI* dalam melakukan proses data analisis sesuai dengan intensi dari *user*. Pada bagian bawah, terdapat *select box* yang memungkinkan *user* mengganti model LLM yang akan digunakan selama proses analisis. Fitur ini akan memberikan fleksibilitas *user* terhadap mesin pemrosesan yang digunakan selama analisis. Selanjutnya, *user* harus menekan tombol “*Generate Insights*” yang akan mengirim seluruh data analisis ke backend dan memulai proses *agentic AI*.



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

#### 4.3.2.5 Halaman *Processing & System Logs*



Gambar 4.50 Halaman *Processing*

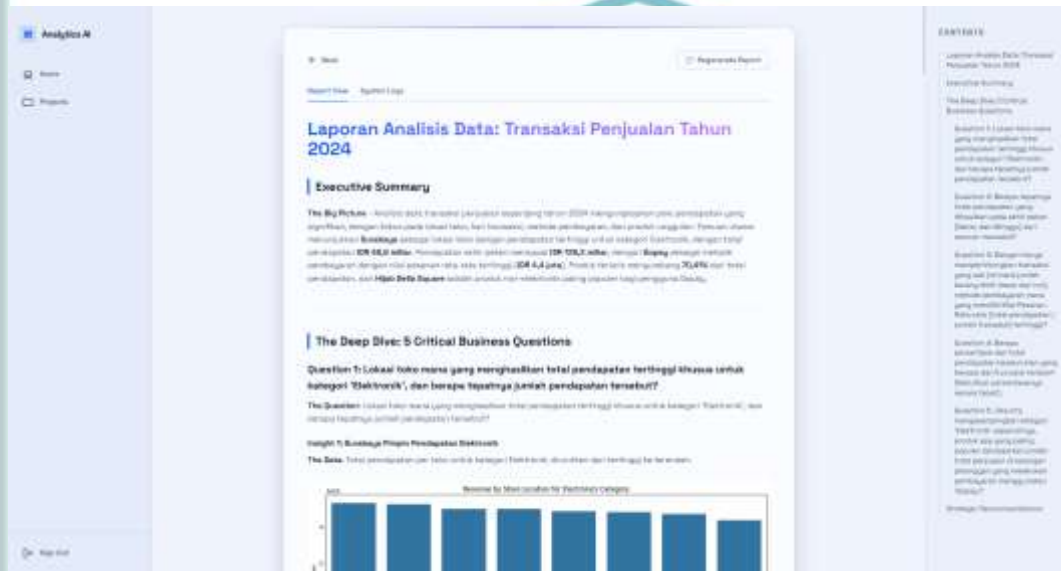
Setelah seluruh data analisis yang akan dilakukan berhasil dikirim ke *backend*, sistem akan menavigasi *user* ke halaman *Processing*. Halaman ini akan menunjukkan *real-time state* dari proses yang sedang berlangsung. Hal ini akan memberikan transparansi pada *user* sudah sejauh mana proses analisis berlangsung. Selain itu, *platform* juga memfasilitasi *user* untuk membatalkan proses analisis yang sedang berlangsung.

| Time     | Type | Name                     | Status | Message                              |
|----------|------|--------------------------|--------|--------------------------------------|
| 21-10-20 | INFO | test_agent               | OK     |                                      |
| 21-10-20 | INFO | data_preprocessing_agent | OK     |                                      |
| 21-10-20 | INFO | data_preprocessing_agent | OK     | Model's F1 Score: 0.9999999999999999 |
| 21-10-20 | INFO | data_preprocessing_agent | OK     |                                      |
| 21-10-20 | INFO | get_understanding_agent  | OK     | Concept: (True, 0.9999999999999999)  |
| 21-10-20 | INFO | get_understanding_agent  | OK     | Concept: (True, 0.9999999999999999)  |

Gambar 4.51 Tabel *System Logs* Pada Halaman *Processing*

Pada halaman *processing* ini juga terdapat menu untuk menampilkan *system logs* yang menunjukkan aktivitas *function calling* pada saat *workflow* analitik berlangsung.

#### 4.3.2.6 Halaman Hasil Laporan Analisis



Gambar 4.52 Halaman Laporan Analisis

Ketika proses analisis yang dilakukan sistem sudah selesai, maka *user* akan berpindah ke halaman Hasil Laporan Analisis. Halaman ini merupakan halaman utama dari analisis yang dilakukan. Halaman ini berisi *Markdown* narasi analisis dan *plot* yang mengilustrasikan narasi analisis yang disampaikan. Selain itu, terdapat juga bagian daftar isi yang dapat ditekan *user* untuk langsung berpindah ke bagian laporan analisis yang ingin dituju. Jika *user* merasa perlu dilakukan analisis ulang, *user* dapat menekan tombol “*Regenerate Report*” yang akan memulai proses analisis dengan data yang sama pada saat *user* pertama kali membuat proyek analisis.





Fokus pengujian tidak hanya pada apakah aplikasi berjalan dengan baik secara fungsional, tetapi juga menguji kemampuan arsitektur *agentic AI*, efisiensi token dan komputasi, serta kualitas *agent* dalam menghasilkan *insights* secara akurat berdasarkan *ground truth*.

#### 4.4.1 Deskripsi Pengujian

Sesuai dengan tujuan dan rumusan masalah penelitian yang sudah didefinisikan di awal, dalam mengukur keberhasilan sistem yang dikembangkan, akan dilakukan serangkaian pengujian untuk kemudian dilakukan analisis lebih lanjut dan memberikan narasi terkait hasil penelitian yang dilakukan. Terdapat beberapa pengujian yang akan dilakukan, dibagi menjadi 3 bagian utama yaitu *Black Box*, Skalabilitas & Waktu Pemrosesan, serta Kualitas *Insights* yang dihasilkan.

##### 4.4.1.1 Black Box Testing

Pengujian *Black Box* berfokus pada fungsionalitas sistem secara *end-to-end* tanpa melihat proses internal pada sistem *Agentic AI*. Tujuannya adalah memastikan integrasi antara antarmuka (Next.js), *task queue* (Celery), dan *database* berjalan dengan baik. Skenario pengujian dapat dilihat pada Lampiran 8.

##### 4.4.1.2 Pengujian Skalabilitas dan Waktu Pemrosesan

Pengujian ini bertujuan untuk mengukur performa sistem ketika dihadapkan pada dataset bervolume besar serta untuk mengevaluasi efisiensi penggunaan token karena pendekatan *Metadata-Centric Context Management*. Pengujian ini krusial untuk membuktikan bahwa sistem tidak akan mengalami *Token Limit Exceeded* meskipun jumlah baris data meningkat secara drastis.

Dalam proses pengujian, *environment* sistem yang digunakan memiliki spesifikasi seperti yang dicantumkan pada Tabel 4.9.

Tabel 4.9 Spesifikasi Perangkat *Environment* Pengujian

| Nama           | Spesifikasi             |
|----------------|-------------------------|
| CPU            | 2x vCPU ARM Neoverse-N1 |
| RAM            | 12 GB                   |
| Storage        | 50 GB                   |
| Sistem Operasi | Ubuntu                  |

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Untuk memastikan arsitektur yang dibangun dapat memproses dataset dengan ukuran yang sangat besar, maka didefinisikan 2 metrik utama yang akan diamati, yaitu Waktu Pemrosesan (*Processing Time*) dan Jumlah Token yang Digunakan (*Token Usage*). Karena konsep arsitektur *multi-agent* yang dibangun akan memproses dataset tanpa mengirimkan keseluruhan dataset ke dalam LLM, maka kedua metrik tersebut akan menunjukkan apakah ukuran dataset dapat mempengaruhi penggunaan token dan waktu pemrosesan.

Sebelum dilakukan pengujian skalabilitas dan kualitas, sistem dipersiapkan dengan membuat dataset sintetis (*dummy*) terkait data *sales* Indonesia menggunakan *Python script*. *Script* ini tidak hanya menghasilkan *file* CSV dengan konfigurasi jumlah baris dan tingkat *noise* (seperti *missing values*, *outliers*, dan nilai negatif pada *dirty* dataset), tetapi juga menghitung *Ground Truth*. Hasil *Ground Truth* ini di-*export* dalam bentuk JSON untuk memvalidasi akurasi hasil analisis *agent*.

Dalam proses pembuatan dataset *dummy*, karena diperlukan *ground truth* untuk mengukur akurasi hasil analisis, maka dirumuskan beberapa *business questions* yang akan dimasukkan ke dalam sistem. Tabel 4.10 menunjukkan daftar *business questions* yang akan digunakan dalam pengujian.

Tabel 4.10 Daftar *Business Questions* untuk Pengujian

| Kode | Pertanyaan Bisnis                                                                                                                                                                                            |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PB-1 | Lokasi toko mana yang menghasilkan total pendapatan tertinggi khusus untuk kategori 'Elektronik', dan berapa tepatnya jumlah pendapatan tersebut?                                                            |
| PB-2 | Berapa total pendapatan yang dihasilkan pada <i>weekend</i> (Sabtu dan Minggu) dari seluruh transaksi?                                                                                                       |
| PB-3 | Jika hanya memperhitungkan transaksi yang valid (dimana <i>quantity</i> lebih besar dari nol), metode pembayaran mana yang memiliki Nilai Pesanan Rata-Rata (total pendapatan / jumlah transaksi) tertinggi? |
| PB-4 | Berapa persentase pendapatan yang berasal dari 5 produk terlaris terhadap total pendapatan keseluruhan?                                                                                                      |
| PB-5 | Jika kita mengesampingkan kategori 'Elektronik', produk apa yang paling laris berdasarkan jumlah total penjualan di kalangan pelanggan yang melakukan pembayaran menggunakan 'Gopay'?                        |



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Selain itu, karena akan dibuat 2 jenis dataset (*clean* dan *dirty*), maka perlu dibuat mekanisme *noise injection* untuk memasukkan *noise* kedalam dataset yang dibuat. *Noise* tersebut akan dimasukkan mengikuti konfigurasi yang didefinisikan di awal *script*.

Dalam memvalidasi mekanisme pembuatan dataset *dummy* beserta data *ground truth* nya, seorang profesional di bidang AI/Data melakukan pengujian terhadap *script* pembuatan dataset dan menganalisis hasil dari *ground truth* yang diperoleh. Hasilnya, evaluator menyatakan bahwa mekanisme pembuatan dataset *dummy* dan *ground truth* nya valid sehingga dataset dapat dijadikan sampel pengujian dalam penelitian ini. Data validasi ini dapat dilihat pada Lampiran 9.

```

1 import pandas as pd
2 import numpy as np
3 import time
4 import json
5 import os
6
7 NUM_OF_ROWS = 10_000_000
8 QUALITY_TYPE = 'dirty'
9
10 MISSING_VALUES_PERCENTAGE = 0.05
11 OUTLIERS_PERCENTAGE = 0.0001
12 NEGATIVE_VALUES_PERCENTAGE = 0.005
13 INCONSISTENT_TEXT_PERCENTAGE = 0.03

```

Gambar 4.55 Konfigurasi Script untuk Pembuatan Dummy Dataset

Gambar 4.55 menunjukkan beberapa variabel yang digunakan dalam melakukan konfigurasi pada dataset yang akan dibuat. Setelah itu, proses *noise injection* akan dilakukan seperti pada gambar 4.56.

```

1 mask_null_cat = np.random.rand(NUM_OF_ROWS) < MISSING_VALUES_PERCENTAGE
2 df.loc[mask_null_cat, 'category'] = np.nan
3
4 mask_outlier_price = np.random.rand(NUM_OF_ROWS) < OUTLIERS_PERCENTAGE
5 df.loc[mask_outlier_price, 'unit_price'] = df.loc[mask_outlier_price, 'unit_price'] * 10
6
7 mask_neg_qty = np.random.rand(NUM_OF_ROWS) < NEGATIVE_VALUES_PERCENTAGE
8 df.loc[mask_neg_qty, 'quantity'] = df.loc[mask_neg_qty, 'quantity'] * -1
9
10 df['total_bill'] = df['quantity'] * df['unit_price']

```

Gambar 4.56 Python Script untuk Noise Injection

Setelah mekanisme *noise injection* telah dibuat, maka proses generasi *dummy* dataset akan dilakukan menggunakan *python script* utama. Gambar 4.57 memperlihatkan *python script* yang digunakan dalam generasi dataset *dummy*.



```

1 catalog = [
2 ...
3]
4
5 catalog_names = np.array([item[0] for item in catalog])
6 catalog_cats = np.array([item[1] for item in catalog])
7 catalog_prices = np.array([item[2] for item in catalog])
8
9 indices = np.random.choice(len(catalog), NUM_OF_ROWS)
10
11 cities = ['Jakarta', 'Surabaya', 'Bandung', 'Medan', 'Semarang', 'Makassar', 'Palembang', 'Yogyakarta']
12 payment_methods = ['Gopay', 'OVO', 'Transfer Bank', 'ShopeePay', 'QRIS', 'Tunai', 'Kartu Kredit']
13
14 data = {
15 'transaction_id': np.arange(1, NUM_OF_ROWS + 1),
16 'date': np.array('2024-01-01', dtype='datetime64[D]') + np.random.choice(np.arange(365), NUM_OF_ROWS),
17 'product_name': catalog_names[indices],
18 'category': catalog_cats[indices],
19 'store_location': np.random.choice(cities, NUM_OF_ROWS),
20 'payment_method': np.random.choice(payment_methods, NUM_OF_ROWS),
21 'quantity': np.random.randint(1, 6, NUM_OF_ROWS),
22 }
23
24 df = pd.DataFrame(data)
25
26 df['unit_price'] = (catalog_prices[indices] * np.random.uniform(0.95, 1.05, NUM_OF_ROWS))
27 df['unit_price'] = (df['unit_price'] // 100) * 100
28 df['total_bill'] = df['quantity'] * df['unit_price']

```

Gambar 4.57 Python Script untuk Pembuatan Dummy Dataset

Setelah dataset berhasil dibuat, maka akan di-generate data *ground truth* berdasarkan pertanyaan bisnis yang telah didefinisikan pada tabel 4.10. Pembuatan data *ground truth* dilakukan melalui *python script* pada gambar 4.58.

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

```

1 # Q1: Elektronik revenue by location
2 elektronik_rev = df[df['category'] == 'Elektronik'].groupby('store_location')['total_bill'].sum()
3 q1_location = str(elektronik_rev.idxmax())
4 q1_revenue = float(elektronik_rev.max())
5
6 # Q2: Weekend Revenue
7 # dayofweek: Monday=0, Sunday=6. Weekends are 5 and 6.
8 weekend_revenue = float(df[df['date'].dt.dayofweek.isin([5, 6])]['total_bill'].sum())
9
10 # Q3: Highest AOV by Payment Method (valid quantities only)
11 valid_transactions = df[df['quantity'] > 0]
12 aov_by_payment = valid_transactions.groupby('payment_method')['total_bill'].mean()
13 q3_payment_method = str(aov_by_payment.idxmax())
14
15 # Q4: Top 5 Products Revenue Percentage
16 total_revenue = float(df['total_bill'].sum())
17 top_5_revenue = float(df.groupby('product_name')['total_bill'].sum().nlargest(5).sum())
18 q4_percentage = (top_5_revenue / total_revenue) * 100
19
20 # Q5: Non-Elektronik, Gopay, Top product by Quantity
21 filtered_q5 = df[(df['category'] != 'Elektronik') & (df['payment_method'] == 'Gopay')]
22 q5_top_product = str(filtered_q5.groupby('product_name')['quantity'].sum().idxmax())
23
24 ground_truth = {
25 "Q1_Top_Elektronik_Location": q1_location,
26 "Q1_Top_Elektronik_Revenue_Amount": q1_revenue,
27 "Q2_Total_Weekend_Revenue": weekend_revenue,
28 "Q3_Highest_AOV_Payment_Method": q3_payment_method,
29 "Q4_Top_5_Revenue_Percentage": q4_percentage,
30 "Q5_Top_Product_Gopay_Non_Elektronik": q5_top_product
31 }

```

Gambar 4.58 Python Script untuk Membuat Data Ground Truth

Script python tersebut kemudian dijalankan pada setiap konfigurasi dataset yang diperlukan dan akan menghasilkan file csv beserta data *ground truth* nya. Data *ground truth* tersebut dapat dilihat pada Lampiran 10.

Dataset *dummy* yang telah dibuat tersebut kemudian dihitung jumlah karakter dan token nya menggunakan *library* dari *OpenAI* yaitu *tiktoken*. Gambar 4.59 menunjukkan *python script* yang digunakan untuk menghitung jumlah karakter dan token yang ada pada suatu dataset.

```

1 import tiktoken
2
3 def count_tokens_in_file(filename):
4 print("Loading file...")
5 # Read the file as a raw string
6 with open(filename, 'r', encoding='utf-8') as f:
7 text = f.read()
8
9 print(f"File loaded. Length: {len(text)} characters.")
10
11 # Initialize Tokenizer (GPT-4 encoding)
12 print("Counting tokens...")
13 enc = tiktoken.get_encoding("cl100k_base")
14 tokens = enc.encode(text)
15
16 print(f"Total Tokens: {len(tokens):,}")
17
18 # Run
19 count_tokens_in_file("datasets/10M_dirty_sales_data_fixed.csv")

```

Gambar 4.59 Python Script untuk Menghitung Jumlah Karakter dan Token

Python Script tersebut kemudian dijalankan pada setiap dataset *dummy* yang dibuat dan menghasilkan informasi jumlah karakter dan token yang ditunjukkan pada Tabel 4.11.

Tabel 4.11 Hasil Perhitungan Karakter dan Token

| Kode   | Ukuran         | Kondisi      | Jumlah Karakter | Jumlah Token |
|--------|----------------|--------------|-----------------|--------------|
| DD-1-1 | 100 Ribu Baris | <i>Clean</i> | 8.325.209       | 4.000.789    |
| DD-1-2 | 100 Ribu Baris | <i>Dirty</i> | 8.283.479       | 3.983.333    |
| DD-2-1 | 500 Ribu Baris | <i>Clean</i> | 42.085.295      | 20.009.397   |
| DD-2-2 | 500 Ribu Baris | <i>Dirty</i> | 41.869.423      | 19.920.214   |
| DD-3-1 | 1 Juta Baris   | <i>Clean</i> | 84.262.071      | 40.018.746   |
| DD-3-2 | 1 Juta Baris   | <i>Dirty</i> | 83.837.452      | 39.843.856   |
| DD-4-1 | 5 Juta Baris   | <i>Clean</i> | 425.788.137     | 204.082.381  |
| DD-4-2 | 5 Juta Baris   | <i>Dirty</i> | 423.602.168     | 203.199.034  |
| DD-5-1 | 10 Juta Baris  | <i>Clean</i> | 852.660.472     | 409.181.736  |
| DD-5-2 | 10 Juta Baris  | <i>Dirty</i> | 848.344.450     | 407.400.830  |

Data tersebut menunjukkan dataset yang ada memiliki rentang jumlah karakter 8,2 juta hingga 852 juta dan jumlah token 3,9 juta hingga 409 juta token. Dataset berukuran besar tersebut tidak dapat diproses langsung dalam memori LLM karena



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



keterbatasan *context window* yang ada. Oleh karena itu, hasil dataset yang dibuat tersebut dapat digunakan dalam pengujian skalabilitas dan waktu pemrosesan untuk membuktikan apakah arsitektur *Multi-Agent* yang dikembangkan mampu mengatasi keterbatasan dari LLM. Pengujian akan dilakukan pada dataset *clean* dengan seluruh ukuran *dummy* dataset yang tersedia menggunakan 2 LLM berbeda. Tabel 4.12 menunjukkan skenario pengujian yang akan dilakukan.

Tabel 4.12 Skenario Pengujian Skalabilitas dan Waktu Pemrosesan

| Kode   | Kode Dataset | Ukuran Dataset | Model LLM                   |
|--------|--------------|----------------|-----------------------------|
| SP-1-1 | DD-1-1       | 100 Ribu Baris | <i>Xiaomi MiMo V2 Flash</i> |
| SP-1-2 | DD-1-1       | 100 Ribu Baris | <i>GPT OSS 120B</i>         |
| SP-2-1 | DD-2-1       | 500 Ribu Baris | <i>Xiaomi MiMo V2 Flash</i> |
| SP-2-2 | DD-2-1       | 500 Ribu Baris | <i>GPT OSS 120B</i>         |
| SP-3-1 | DD-3-1       | 1 Juta Baris   | <i>Xiaomi MiMo V2 Flash</i> |
| SP-3-2 | DD-3-1       | 1 Juta Baris   | <i>GPT OSS 120B</i>         |
| SP-4-1 | DD-4-1       | 5 Juta Baris   | <i>Xiaomi MiMo V2 Flash</i> |
| SP-4-2 | DD-4-1       | 5 Juta Baris   | <i>GPT OSS 120B</i>         |
| SP-5-1 | DD-5-1       | 10 Juta Baris  | <i>Xiaomi MiMo V2 Flash</i> |
| SP-5-2 | DD-5-1       | 10 Juta Baris  | <i>GPT OSS 120B</i>         |

Dari skenario pengujian tersebut, akan dianalisis penggunaan token dan waktu pemrosesan keseluruhan dari awal hingga terbentuk laporan analisis. Data hasil pengujian tersebut akan membuktikan apakah arsitektur yang dibangun pada penelitian ini *scalable* pada ukuran dataset yang sangat besar.

#### 4.4.1.3 Pengujian Kualitas *Insight* dan Konsistensi

Pengujian kualitas *insight* dan konsistensi berfokus pada evaluasi utama dari kemampuan arsitektur *Multi-Agent* itu sendiri, yaitu *Function Calling* dan output analitik yang dihasilkan. Mengingat sistem berfokus pada dua *tools* utama (*run\_sql()* dan *plot\_tool()*), pengujian difokuskan pada kemampuan *agent* dalam memanfaatkan *tools* tersebut. Tujuan utamanya adalah untuk menguji apakah *agent* dapat memilih *tools* yang tepat saat melakukan *Function Calling*, mampu



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

memperbaiki *error* melalui kemampuan *Self Correction*, serta menghasilkan output yang konsisten ketika skenario yang sama dijalankan lebih dari satu kali.

Pengujian ini mencakup tiga pendekatan utama. Pertama, validasi keakuratan matematis menggunakan dataset sintetis dengan membandingkan hasil kalkulasi sistem terhadap *ground truth* yang telah dikalkulasi sebelumnya. Kedua, evaluasi saintifik komparatif menggunakan dataset NvBench dari literatur untuk mengamati sejauh mana pendekatan dan arah rekomendasi sistem sejalan dengan analisis yang dijadikan acuan. Ketiga, pengujian konsistensi output dengan menjalankan skenario yang sama secara berulang untuk mengukur konsistensi kalkulasi dan strategi *data cleaning* antar eksekusi.

Pada proses pengujian kualitas *insight* dan konsistensi ini, dirumuskan beberapa metrik yang akan digunakan untuk menganalisis hasil laporan analisis yang dilakukan oleh sistem. Daftar metrik tersebut dapat dilihat pada Tabel 4.13.

Tabel 4.13 Metrik Pengujian Hasil Laporan Analisis

| Kode  | Metrik                                | Deskripsi                                                                                                                                                                                               |
|-------|---------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MIK-1 | <i>Function Calling</i>               | Memastikan bahwa <i>agent</i> dapat melakukan <i>function calling</i> untuk melakukan eksekusi analitik yang dibutuhkan.                                                                                |
| MIK-2 | Visualisasi                           | Memastikan bahwa sistem dapat menghasilkan visualisasi yang dibutuhkan melalui mekanisme <i>function calling</i> .                                                                                      |
| MIK-3 | <i>Self-Correction</i>                | Menghitung kejadian <i>Error SQL</i> pada <i>system logs</i> serta cara <i>agent</i> dalam melakukan <i>self-correction</i> .                                                                           |
| MIK-4 | Validasi <i>Ground Truth</i>          | Mengkomparasikan hasil analisis pada laporan akhir dengan <i>Ground Truth</i> .                                                                                                                         |
| MIK-5 | Evaluasi Saintifik ( <i>NvBench</i> ) | Menganalisis dataset NvBench dan membandingkan output sistem melalui observasi. Perbandingan ini bersifat deskriptif tanpa penilai eksternal, dan tidak dimaksudkan sebagai evaluasi kualitatif formal. |
| MIK-6 | Validasi <i>Embedding</i> Visualisasi | Memeriksa laporan analitik yang ada untuk memastikan gambar grafik berhasil di- <i>embed</i> oleh <i>agent</i> .                                                                                        |
| MIK-7 | Konsistensi                           | Mengukur stabilitas hasil kalkulasi numerik dan                                                                                                                                                         |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|  |        |                                                                                   |
|--|--------|-----------------------------------------------------------------------------------|
|  | Output | strategi <i>data cleaning</i> ketika skenario yang sama dijalankan berulang kali. |
|--|--------|-----------------------------------------------------------------------------------|

Setiap skenario pengujian dirancang untuk dapat dianalisis dengan metrik yang relevan. Skenario menggunakan dataset NvBench akan dianalisis dengan metrik MIK-5. Dataset yang diambil dari *research paper* tersebut dipetakan pada Tabel 4.14.

Tabel 4.14 Daftar Dataset dari “Is GPT-4 a Good Data Analyst”

| Kode | Nama Dataset          | Pertanyaan Bisnis                                                                                     |
|------|-----------------------|-------------------------------------------------------------------------------------------------------|
| NB-1 | <i>candidate_poll</i> | <i>Which candidate should we accept?</i>                                                              |
| NB-2 | <i>college_1</i>      | <i>Choose 3 students who perform well on tests as representatives.</i>                                |
| NB-3 | <i>hr_1</i>           | <i>How to reduce human cost by shifting employees from different departments among these regions?</i> |
| NB-4 | <i>inn_1</i>          | <i>Which room is more popular?</i>                                                                    |
| NB-5 | <i>wine_1</i>         | <i>A client’s budget is 60 and he doesn’t like Zinfandel, which wines shall we recommend to him?</i>  |

Skenario dengan dataset *dummy* (DD-1-1 hingga DD-5-2) akan digunakan untuk menganalisis metrik MIK-1, MIK-2, MIK-3, MIK-4, dan MIK-6, karena dataset tersebut memiliki *ground truth* yang terkalkulasi. Skenario pengujian konsistensi (KO-1-1 hingga KO-2-2) digunakan untuk menganalisis metrik MIK-7 dengan menjalankan dataset DD-1-1 dan DD-1-2 sebanyak tiga kali per model LLM. Pemetaan lengkap skenario ditunjukkan pada Tabel 4.15 dan Tabel 4.16.

Tabel 4.15 Skenario Pengujian *Insight*

| Kode   | Dataset | Pertanyaan Bisnis                        | Model LLM                   | Metrik |
|--------|---------|------------------------------------------|-----------------------------|--------|
| IK-1-1 | NB-1    | <i>Which candidate should we accept?</i> | <i>Xiaomi MiMo V2 Flash</i> | MIK-5  |
| IK-1-2 |         |                                          | <i>GPT OSS 120B</i>         |        |
| IK-2-1 | NB-2    | <i>Choose 3 students</i>                 | <i>Xiaomi MiMo V2 Flash</i> |        |



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|         |        |                                                                                                                                                      |                             |                                               |
|---------|--------|------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|-----------------------------------------------|
| IK-2-2  |        | <i>who perform well on tests as representatives.</i>                                                                                                 | <i>GPT OSS 120B</i>         |                                               |
| IK-3-1  | NB-3   | <i>How to reduce human cost by shifting employees from different departments among these regions?</i>                                                | <i>Xiaomi MiMo V2 Flash</i> |                                               |
| IK-3-2  |        |                                                                                                                                                      | <i>GPT OSS 120B</i>         |                                               |
| IK-4-1  | NB-4   | <i>Which room is more popular?</i>                                                                                                                   | <i>Xiaomi MiMo V2 Flash</i> |                                               |
| IK-4-2  |        |                                                                                                                                                      | <i>GPT OSS 120B</i>         |                                               |
| IK-5-1  | NB-5   | <i>A client's budget is 60 and he doesn't like Zinfandel, which wines shall we recommend to him?</i>                                                 | <i>Xiaomi MiMo V2 Flash</i> |                                               |
| IK-5-2  |        |                                                                                                                                                      | <i>GPT OSS 120B</i>         |                                               |
| IK-6-1  | DD-1-1 | 1. Lokasi toko mana yang menghasilkan total pendapatan tertinggi khusus untuk kategori 'Elektronik', dan berapa tepatnya jumlah pendapatan tersebut? | <i>Xiaomi MiMo V2 Flash</i> | MIK-1,<br>MIK-2,<br>MIK-3,<br>MIK-4,<br>MIK-6 |
| IK-6-2  |        |                                                                                                                                                      | <i>GPT OSS 120B</i>         |                                               |
| IK-7-1  | DD-1-2 |                                                                                                                                                      | <i>Xiaomi MiMo V2 Flash</i> |                                               |
| IK-7-2  |        |                                                                                                                                                      | <i>GPT OSS 120B</i>         |                                               |
| IK-8-1  | DD-2-1 | 2. Berapa total pendapatan yang dihasilkan pada weekend (Sabtu dan Minggu) dari seluruh transaksi?                                                   | <i>Xiaomi MiMo V2 Flash</i> |                                               |
| IK-8-2  |        |                                                                                                                                                      | <i>GPT OSS 120B</i>         |                                               |
| IK-9-1  | DD-2-2 |                                                                                                                                                      | <i>Xiaomi MiMo V2 Flash</i> |                                               |
| IK-9-2  |        |                                                                                                                                                      | <i>GPT OSS 120B</i>         |                                               |
| IK-10-1 | DD-3-1 | 3. Jika hanya memperhitungkan transaksi yang valid (dimana quantity lebih besar                                                                      | <i>Xiaomi MiMo V2 Flash</i> |                                               |
| IK-10-2 |        |                                                                                                                                                      | <i>GPT OSS 120B</i>         |                                               |



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|         |        |                                                                                                                                                                                          |                      |  |
|---------|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|--|
| IK-11-1 | DD-3-2 | dari nol), metode pembayaran mana yang memiliki Nilai Pesanan Rata-Rata (total pendapatan / jumlah transaksi) tertinggi?                                                                 | Xiaomi MiMo V2 Flash |  |
| IK-11-2 |        |                                                                                                                                                                                          | GPT OSS 120B         |  |
| IK-12-1 | DD-4-1 |                                                                                                                                                                                          | Xiaomi MiMo V2 Flash |  |
| IK-12-2 |        |                                                                                                                                                                                          | GPT OSS 120B         |  |
| IK-13-1 | DD-4-2 | 4. Berapa persentase pendapatan yang berasal dari 5 produk terlaris terhadap total pendapatan keseluruhan?                                                                               | Xiaomi MiMo V2 Flash |  |
| IK-13-2 |        |                                                                                                                                                                                          | GPT OSS 120B         |  |
| IK-14-1 | DD-5-1 | 5. Jika kita mengesampingkan kategori 'Elektronik', produk apa yang paling laris berdasarkan jumlah total penjualan di kalangan pelanggan yang melakukan pembayaran menggunakan 'Gopay'? | Xiaomi MiMo V2 Flash |  |
| IK-14-2 |        |                                                                                                                                                                                          | GPT OSS 120B         |  |
| IK-15-1 | DD-5-2 |                                                                                                                                                                                          | Xiaomi MiMo V2 Flash |  |
| IK-15-2 |        |                                                                                                                                                                                          | GPT OSS 120B         |  |

Tabel 4.16 Skenario Pengujian Konsistensi

| Kode      | Dataset | Model                | Run   |
|-----------|---------|----------------------|-------|
| KO-1-1-R1 | DD-1-1  | Xiaomi MiMo V2 Flash | Run 1 |
| KO-1-1-R2 |         |                      | Run 2 |
| KO-1-1-R3 |         |                      | Run 3 |
| KO-1-2-R1 |         | GPT OSS 120B         | Run 1 |



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|           |        |                      |       |
|-----------|--------|----------------------|-------|
| KO-1-2-R2 |        |                      | Run 2 |
| KO-1-2-R3 |        |                      | Run 3 |
| KO-2-1-R1 | DD-1-2 | Xiaomi MiMo V2 Flash | Run 1 |
| KO-2-1-R2 |        |                      | Run 2 |
| KO-2-1-R3 |        |                      | Run 3 |
| KO-2-2-R1 |        | GPT OSS 120B         | Run 1 |
| KO-2-2-R2 |        |                      | Run 2 |
| KO-2-2-R3 |        |                      | Run 3 |

#### 4.4.2 Prosedur Pengujian

##### 4.4.2.1 Prosedur *Black Box Tesing*

Metode *Black Box Testing* akan dilakukan secara mandiri dengan mengeksekusi satu per satu *test case* yang telah didefinisikan pada skenario pengujian. Proses pengujian akan dilakukan dengan berinteraksi langsung dengan *web* layaknya *end-user*. Setiap langkah akan dilakukan sesuai dengan skenario pengujian yang telah didefinisikan sebelumnya. Setelah interaksi dilakukan, respons dari sistem akan diamati dan dicatat. Pengukuran keberhasilan pada tahap ini didasarkan pada perbandingan antara hasil aktual yang ditampilkan oleh antarmuka sistem dengan hasil yang diharapkan yang tercantum pada tabel skenario pengujian. Jika sistem berhasil memberikan respons yang tepat sesuai alur dan tidak terjadi kegagalan sistem (*crash*) yang tidak tertangani, maka *test case* tersebut dinyatakan berhasil.

##### 4.4.2.2 Prosedur Pengujian Skalabilitas dan Waktu Pemrosesan

Pengujian dilakukan dengan mengunggah setiap dataset *dummy* secara bergantian ke dalam sistem, mulai dari ukuran 100 ribu hingga 10 juta baris, menggunakan dua model LLM yang berbeda yaitu Xiaomi MiMo V2 Flash dan GPT OSS 120B. Setiap skenario dijalankan dengan *business questions* yang sama sesuai Tabel 4.10 sehingga satu-satunya variabel yang berubah adalah ukuran dataset dan model LLM yang digunakan.



Data metrik waktu pemrosesan dan penggunaan token diambil langsung dari *database sistem* yang secara otomatis mencatat durasi dan total token pada setiap proses analisis. Data ini kemudian dianalisis untuk membuktikan efektivitas pendekatan *Metadata-Centric Context Management* dalam menjaga stabilitas penggunaan token seiring meningkatnya ukuran dataset.

#### 4.4.2.3 Prosedur Pengujian Kualitas *Insight* dan Konsistensi

Seluruh skenario pengujian dilakukan menggunakan dua model LLM secara bergantian, yaitu Xiaomi MiMo V2 Flash dan GPT OSS 120B. Selama proses berlangsung, *system logs agent* dicatat untuk mengevaluasi metrik MIK-1, MIK-2, dan MIK-3.

Untuk metrik MIK-4, dataset DD-1-1 dan DD-1-2 dimasukkan ke dalam sistem bersamaan dengan 5 *Business Questions* tetap yang telah didefinisikan pada Tabel 4.10. Setelah *agent* selesai bekerja, hasil analisis dikomparasikan langsung dengan data *ground truth* untuk memastikan tidak terjadi kesalahan kalkulasi.

Untuk metrik MIK-5, dataset NvBench (NB-1 hingga NB-5) beserta *Business Questions* diinput ke dalam sistem. Hasil laporan analitik yang dihasilkan kemudian dibandingkan dengan temuan pada literatur "*Is GPT-4 a Good Data Analyst?*" untuk mengamati kesesuaian pendekatan dan arah rekomendasi sistem terhadap analisis yang dijadikan acuan.

Untuk metrik MIK-7, skenario KO dijalankan sebanyak tiga kali secara bergantian per model LLM tanpa modifikasi apapun pada sistem atau dataset di antara *run*. Pada skenario dataset *clean* (KO-1-1 dan KO-1-2), nilai numerik PB-1 hingga PB-5 dari setiap *run* dibandingkan satu sama lain dan terhadap *ground truth*. Pada skenario dataset *dirty* (KO-2-1 dan KO-2-2), selain perbandingan hasil analisis, log SQL *cleaning* dari setiap *run* juga dicatat untuk mengamati apakah strategi *data cleaning* yang dipilih model konsisten antar *run*.

Sebagai tahap akhir dari setiap skenario pengujian, dilakukan validasi visual (MIK-6) pada laporan *Markdown* yang dihasilkan untuk memastikan seluruh grafik berhasil di-*embed* dengan tepat.

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

#### 4.4.3 Data Hasil Pengujian

##### 4.4.3.1 Hasil *Black Box Testing*

Hasil pengujian *Black Box* terhadap seluruh 22 skenario yang telah didefinisikan dapat dilihat pada Lampiran 11. Hasil tersebut kemudian dirangkum pada Tabel 4.17.

Tabel 4.17 Hasil Pengujian *Black Box* per Fungsi

| Fungsi                      | Jumlah Skenario | Hasil             |
|-----------------------------|-----------------|-------------------|
| <i>Sign Up</i>              | 4               | Seluruhnya Sesuai |
| <i>Sign In</i>              | 5               | Seluruhnya Sesuai |
| Membuat Proyek Analisis     | 3               | Seluruhnya Sesuai |
| Membatalkan Proses Analisis | 1               | Sesuai            |
| Mengubah Nama Proyek        | 2               | Seluruhnya Sesuai |
| Menghapus Proyek Analisis   | 1               | Sesuai            |
| Navigasi Daftar Isi Laporan | 1               | Sesuai            |
| Analisis Ulang              | 1               | Sesuai            |
| Mengubah Mode Tampilan      | 1               | Sesuai            |
| Mencari Proyek Analisis     | 4               | Seluruhnya Sesuai |

Seluruh fungsi yang diuji menunjukkan hasil yang sesuai dengan harapan, tidak ditemukan kegagalan sistem maupun respons yang menyimpang dari alur yang telah didefinisikan.

Hasil ini mengkonfirmasi bahwa integrasi antara antarmuka Next.js, *task queue Celery*, dan *database* berjalan dengan baik pada seluruh alur interaksi *end-user*. Dengan demikian, fungsionalitas sistem telah memenuhi syarat untuk mendukung pengujian tahap berikutnya, yaitu pengujian skalabilitas dan kualitas *insight*.

##### 4.4.3.2 Hasil Pengujian Skalabilitas dan Waktu Pemrosesan

Hasil pengujian skalabilitas dan waktu pemrosesan dari seluruh skenario yang telah didefinisikan ditunjukkan pada Tabel 4.18.

Tabel 4.18 Hasil Pengujian Skalabilitas dan Waktu Pemrosesan

| Kode | Token Usage | Processing Time (s) | Status |
|------|-------------|---------------------|--------|
|------|-------------|---------------------|--------|

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |         |     |          |
|--------|---------|-----|----------|
| SP-1-1 | 73.842  | 220 | Berhasil |
| SP-1-2 | 58.581  | 577 | Berhasil |
| SP-2-1 | 107.043 | 349 | Berhasil |
| SP-2-2 | 46.749  | 394 | Berhasil |
| SP-3-1 | 75.213  | 385 | Berhasil |
| SP-3-2 | 63.170  | 346 | Berhasil |
| SP-4-1 | 62.226  | 312 | Berhasil |
| SP-4-2 | 58.889  | 649 | Berhasil |
| SP-5-1 | 74.848  | 386 | Berhasil |
| SP-5-2 | 51.373  | 533 | Berhasil |

Seluruh 10 skenario pengujian berhasil dieksekusi dengan status Berhasil, menunjukkan bahwa sistem mampu memproses dataset dari ukuran 100 ribu hingga 10 juta baris tanpa kegagalan. Data *token usage* dan *processing time* yang dihasilkan akan dianalisis lebih lanjut pada bagian 4.4.4.1.

#### 4.4.3.3 Hasil Pengujian Kualitas *Insight* dan Konsistensi

Bagian ini menyajikan data hasil pengujian kualitas *insight* dan konsistensi yang dibagi berdasarkan tujuh kriteria metrik yang telah didefinisikan pada Tabel 4.14. Seluruh data yang disajikan pada bagian ini merupakan output langsung dari sistem dan akan dianalisis lebih lanjut pada bagian 4.4.4.

##### 4.4.3.3.1 Hasil Pengujian Kriteria *Function Calling*

Seluruh 20 skenario pengujian telah dieksekusi dan hasilnya ditunjukkan pada Tabel 4.19. Keberhasilan yang dimaksud adalah ketika suatu *tool* berhasil dieksekusi tanpa menyebabkan *runtime error* yang mengakibatkan proses analisis terhenti atau sistem mengalami *crash*.

Tabel 4.19 Hasil Pengujian Kriteria *Function Calling*

| Kode   | Jumlah <i>Function Calling</i> | Persentase Keberhasilan |
|--------|--------------------------------|-------------------------|
| IK-6-1 | 135                            | 100%                    |
| IK-6-2 | 66                             | 100%                    |



### © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|         |     |      |
|---------|-----|------|
| IK-7-1  | 182 | 100% |
| IK-7-2  | 74  | 100% |
| IK-8-1  | 180 | 100% |
| IK-8-2  | 59  | 100% |
| IK-9-1  | 110 | 100% |
| IK-9-2  | 62  | 100% |
| IK-10-1 | 158 | 100% |
| IK-10-2 | 69  | 100% |
| IK-11-1 | 132 | 100% |
| IK-11-2 | 62  | 100% |
| IK-12-1 | 137 | 100% |
| IK-12-2 | 78  | 100% |
| IK-13-1 | 203 | 100% |
| IK-13-2 | 73  | 100% |
| IK-14-1 | 152 | 100% |
| IK-14-2 | 63  | 100% |
| IK-15-1 | 101 | 100% |
| IK-15-2 | 67  | 100% |

#### 4.4.3.3.2 Hasil Pengujian Kriteria Visualisasi

Salah satu bagian penting dari *function calling* adalah pemanggilan *plot\_tool()* dimana *tool* ini akan dieksekusi untuk menghasilkan *file* diagram yang akan di-



*embed* pada laporan analisis yang dibuat oleh sistem. Setelah dilakukan pengujian, jumlah pemanggilan *plot\_tool()* dapat dilihat pada Tabel 4.20.

Tabel 4.20 Hasil Pengujian Kriteria Visualisasi

| Kode    | Jumlah <i>plot_tool()</i> | Persentase Keberhasilan |
|---------|---------------------------|-------------------------|
| IK-6-1  | 13                        | 100%                    |
| IK-6-2  | 5                         | 100%                    |
| IK-7-1  | 22                        | 100%                    |
| IK-7-2  | 6                         | 100%                    |
| IK-8-1  | 19                        | 100%                    |
| IK-8-2  | 5                         | 100%                    |
| IK-9-1  | 11                        | 100%                    |
| IK-9-2  | 5                         | 100%                    |
| IK-10-1 | 11                        | 100%                    |
| IK-10-2 | 5                         | 100%                    |
| IK-11-1 | 5                         | 100%                    |
| IK-11-2 | 5                         | 100%                    |
| IK-12-1 | 13                        | 100%                    |
| IK-12-2 | 5                         | 100%                    |
| IK-13-1 | 24                        | 100%                    |
| IK-13-2 | 6                         | 100%                    |
| IK-14-1 | 16                        | 100%                    |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|         |    |      |
|---------|----|------|
| IK-14-2 | 5  | 100% |
| IK-15-1 | 10 | 100% |
| IK-15-2 | 5  | 100% |

Tabel 4.20 menunjukkan bahwa jumlah pemanggilan *plot\_tool()* bervariasi antara 5 hingga 24 kali di seluruh skenario pengujian. Data ini akan dianalisis lebih lanjut pada bagian 4.4.4.2.3.

#### 4.4.3.3.3 Hasil Pengujian Kriteria *Self-Correction*

Dalam prosesnya, sistem *Agentic AI* memiliki kapabilitas untuk melakukan mekanisme *self-correction*. Hal ini memungkinkan *agent* mengetahui *error* yang terjadi ketika melakukan *function calling* dan melakukan *self-correction* untuk mendapatkan hasil yang tidak *error*. Mekanisme *self-correction* ini telah diamati pada skenario pengujian yang telah dilakukan pada 20 *test case* berbeda. Hasil pengamatan mekanisme *self-correction* ini dapat dilihat pada Tabel 4.21 yang menunjukkan jumlah *self-correction* yang dilakukan beserta persentase keberhasilan dalam menangani *error* pada saat melakukan *function calling*.

Tabel 4.21 Hasil Pengujian Kriteria *Self-Correction*

| Kode   | Jumlah <i>Self-Correction</i> | Persentase Keberhasilan |
|--------|-------------------------------|-------------------------|
| IK-6-1 | 1                             | 100%                    |
| IK-6-2 | 1                             | 100%                    |
| IK-7-1 | 3                             | 100%                    |
| IK-7-2 | 0                             | -                       |
| IK-8-1 | 4                             | 100%                    |
| IK-8-2 | 0                             | -                       |
| IK-9-1 | 3                             | 100%                    |
| IK-9-2 | 1                             | 100%                    |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|         |   |      |
|---------|---|------|
| IK-10-1 | 2 | 100% |
| IK-10-2 | 2 | 100% |
| IK-11-1 | 6 | 100% |
| IK-11-2 | 2 | 100% |
| IK-12-1 | 3 | 100% |
| IK-12-2 | 0 | -    |
| IK-13-1 | 3 | 100% |
| IK-13-2 | 2 | 100% |
| IK-14-1 | 4 | 100% |
| IK-14-2 | 2 | 100% |
| IK-15-1 | 1 | 100% |
| IK-15-2 | 3 | 100% |

Tabel 4.21 menunjukkan variasi jumlah *self-correction* yang dilakukan *agent* dari 0 hingga 6 kali, yang menunjukkan bahwa sejumlah skenario tidak mengalami *error* sama sekali. Data ini akan dianalisis lebih lanjut pada bagian 4.4.4.2.2.

#### 4.4.3.3.4 Hasil Pengujian Validasi *Ground Truth*

Dalam menguji kebenaran hasil laporan analisis yang dibuat oleh *agent AI*, dilakukan perbandingan antara hasil pada laporan analisis dengan *ground truth*. Dimana dalam pengujian ini, diamati 2 variasi dataset yaitu 100 ribu baris *clean* dan 100 ribu baris *dirty* yang masing masing diuji menggunakan 2 LLM berbeda. Hasil pengujian tersebut dapat dilihat pada Tabel 4.22 berikut.

Tabel 4.22 Hasil Pengujian Kriteria *Ground Truth*

| Kode | Pertanyaan | Ground Truth | Hasil | Akurasi |
|------|------------|--------------|-------|---------|
|------|------------|--------------|-------|---------|



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |        |                        |                        |        |
|--------|--------|------------------------|------------------------|--------|
| IK-6-1 | PB-1-1 | Surabaya               | Surabaya               | 100%   |
|        | PB-1-2 | 46628363000            | 46628363000            | 100%   |
|        | PB-2   | 126192538400           | 126192538400           | 100%   |
|        | PB-3   | Gopay                  | Gopay                  | 100%   |
|        | PB-4   | 70.41                  | 70.41                  | 100%   |
|        | PB-5   | Hijab Bella Square     | Hijab Bella Square     | 100%   |
| IK-6-2 | PB-1-1 | Surabaya               | Surabaya               | 100%   |
|        | PB-1-2 | 46628363000            | 46628363000            | 100%   |
|        | PB-2   | 126192538400           | 126192538400           | 100%   |
|        | PB-3   | Gopay                  | Gopay                  | 100%   |
|        | PB-4   | 70.41                  | 70.41                  | 100%   |
|        | PB-5   | Hijab Bella Square     | Hijab Bella Square     | 100%   |
| IK-7-1 | PB-1-1 | Surabaya               | Surabaya               | 100%   |
|        | PB-1-2 | 46597984500            | 43812631000            | 94%    |
|        | PB-2   | 124513030800           | 123886716300           | 99.5%  |
|        | PB-3   | Gopay                  | Gopay                  | 100%   |
|        | PB-4   | 70.15                  | 70.13                  | 99.98% |
|        | PB-5   | Beras Pandan Wangi 5kg | Beras Pandan Wangi 5kg | 100%   |
| IK-7-2 | PB-1-1 | Surabaya               | Surabaya               | 100%   |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |                        |                        |         |
|--------|------------------------|------------------------|---------|
| PB-1-2 | 46597984500            | 46603337700            | 99.99%  |
| PB-2   | 124513030800           | 124518384000           | 99.996% |
| PB-3   | Gopay                  | Gopay                  | 100%    |
| PB-4   | 70.15                  | 70.16                  | 99.99%  |
| PB-5   | Beras Pandan Wangi 5kg | Beras Pandan Wangi 5kg | 100%    |

Dari tabel tersebut dapat dilihat bahwa persentase akurasi pada dataset *clean* mencapai 100% akurat yang menunjukkan bahwa sistem *agentic AI* yang dibangun mampu menghasilkan laporan yang reliabel pada kondisi dataset *clean*. Di sisi lain, pada dataset *dirty* terdapat perbedaan pada hasil analisis berupa angka namun sesuai pada hasil analisis berupa kategori. Hal ini terjadi karena pada dataset *dirty* proses perhitungan *ground truth* dilakukan sebelum *noise* dimasukkan ke dalam dataset. Oleh karena itu, proses *cleaning* yang dilakukan oleh *agent* sangat menentukan hasil kalkulasi numerik pada laporan analisis.

#### 4.4.3.3.5 Hasil Pengujian Kriteria Evaluasi Saintifik

##### a. Hasil Pengujian Dataset NB-1

Dataset NB-1 berjudul *candidate\_poll* telah diuji kedalam sistem yang dibangun menggunakan 2 LLM berbeda menghasilkan output sebagai berikut:

Tabel 4.23 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-1

| Kode   | Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| IK-1-1 | <p><i>The Big Picture: Analysis of polling data for 6 candidates reveals Vincent Montméat has the highest support rate at 33%, making him the most popular choice among voters. However, this comes with the highest opposition rate at 45%, creating a polarization effect. Loïc De Kergret presents as the alternative candidate with the lowest opposition rate (32%) and moderate support (24%).</i></p> <p><i>The "So What?" for decision-makers: While Vincent Montméat is statistically the strongest candidate based on raw polling numbers, his high opposition suggests potential polarization risk. The dataset contains no qualitative decision criteria, forcing reliance on quantitative metrics alone.</i></p> |



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| IK-1-2 | <p><i>The Big Picture: Across all candidates, Vincent Montméat leads with the highest average support rate (33 %). However, Loïc De Kergret exhibits the most favorable net support (support – oppose = -0.08), indicating the lowest opposition relative to support. Support and consideration rates move together, suggesting candidates strong in one tend to be strong in the other.</i></p> <p><i>So What?: Prioritizing a candidate with both high support and low opposition points to Loïc De Kergret as the most balanced choice, while Vincent Montméat offers the highest outright support.</i></p> |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Kedua LLM menghasilkan temuan yang serupa. Keduanya menyebut Vincent Montméat sebagai kandidat dengan *support rate* tertinggi sekaligus mempertimbangkan *opposition rate*-nya sebagai faktor risiko. Perbedaannya terletak pada ketegasan rekomendasi akhir, GPT OSS 120B secara eksplisit merekomendasikan Loïc De Kergret, sementara Xiaomi MiMo V2 Flash menyajikannya sebagai alternatif tanpa rekomendasi langsung. Perbandingan hasil ini dengan *baseline* analis profesional disajikan pada bagian 4.4.4.3.2.

#### b. Hasil Pengujian Dataset NB-2

Dataset NB-2 berjudul *college\_1* telah diuji kedalam sistem yang dibangun menggunakan 2 LLM berbeda menghasilkan output sebagai berikut:

Tabel 4.24 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-2

| Kode   | Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| IK-2-1 | <p><i>The Big Picture: This report analyzes student academic performance using GPA as a proxy for test performance, given the absence of explicit test score data in the dataset. Three students stand out as high performers: Gerald Robertson (GPA: 3.87), Anne Smithson (GPA: 3.27), and Raphael Katinga (GPA: 3.15). These students represent the top 3 performers across diverse departments (EDU, CIS, and ACCT).</i></p> <p><i>Key findings reveal that Junior students have the highest average GPA (3.18), non-transfer students outperform transfer students by 0.62 GPA points, and academic hours show a moderate positive correlation (0.389) with GPA. The data suggests that department affiliation, class standing, and transfer status are significant predictors of academic success.</i></p> |
| IK-2-2 | <p><i>The Big Picture: Using GPA as a proxy for test performance, the dataset reveals a concentration of students around a GPA of 3.0 with a modest tail of higher achievers. The top three performers—Gerald Robertson, Anne Smithson, and Raphael Katinga—stand out with</i></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

*GPA's of 3.87, 3.27, and 3.15 respectively. A positive correlation between credit hours earned and GPA suggests that experience contributes to higher academic outcomes.*

*So what? These insights enable the selection of three high-performing student representatives and highlight the value of experience-based academic growth.*

Kedua LLM menggunakan GPA sebagai parameter performa dan menyebut tiga nama yang sama secara berurutan. Perbedaan terletak pada kedalaman informasi, dimana Xiaomi MiMo V2 Flash memberikan informasi tambahan seperti korelasi jam akademik dengan GPA dan pola performa berdasarkan status transfer, sedangkan GPT OSS 120B lebih ringkas. Perbandingan jawaban langsung dengan *baseline* analis profesional disajikan pada bagian 4.4.4.3.2.

c. Hasil pengujian Dataset NB-3

Dataset NB-3 berjudul *hr\_1* telah diuji kedalam sistem yang dibangun menggunakan 2 LLM berbeda menghasilkan output sebagai berikut:

Tabel 4.25 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-3

| Kode   | Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| IK-3-1 | <p><i>The Big Picture: Analysis of 107 cleaned employee records reveals significant cost optimization opportunities through strategic employee shifting between the Americas and Europe regions. Europe's average salary is 71% higher than Americas (\$8,865 vs \$5,191), creating strong financial incentives for department consolidation. Sales department in Europe represents the highest cost reduction potential at \$311,500 in annual salary costs. The data shows clear regional concentration patterns that enable targeted workforce optimization strategies.</i></p> <p><i>Key Metrics:</i></p> <ul style="list-style-type: none"> <li>- \$3,674 average cost differential per employee between regions</li> <li>- \$639,400 total salary cost across both regions</li> <li>- 35 employees in Europe's Sales department with \$8,900 average salary</li> </ul> |
| IK-3-2 | <p><i>The Big Picture: The Americas region shoulders the bulk of salary expenses (\$363,400) and staff count (70 employees), while a smaller "NULL" region packs a high average salary (\$8,900) for only 35 employees. The Sales department in the NULL region accounts for 86 % of that region's cost. Shipping in the Americas employs many staff at a low average salary (\$3,475), presenting a clear low-cost</i></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                  |



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

*labor pool. The salary distribution is right-skewed, meaning a few high-salary roles drive most of the cost.*

*So What? Reallocating high-salary Sales staff to lower-cost departments/regions (e.g., Shipping in the Americas) can substantially cut human costs without major headcount changes.*

Hasil pengujian Dataset NB-3 tersebut menunjukkan perbedaan pada hasil analisis kedua LLM. Dimana model Xiaomi MiMo V2 Flash melakukan *missing values imputation* pada baris dengan *null value* pada kolom *region* sedangkan model GPT OSS 120B membiarkan *null values* tersebut tetap *null*.

Step 2: Resolve missing region data using CITY and STATE\_PROVINCE mapping

- SQL: `'SELECT *, CASE WHEN CITY = 'South San Francisco' THEN 2.0 WHEN CITY = 'Oxford' THEN 3.0 ELSE REGION_ID END AS REGION_ID, CASE WHEN CITY = 'South San Francisco' THEN 'Americas' WHEN CITY = 'Oxford' THEN 'Europe' ELSE REGION_NAME END AS REGION_NAME, CASE WHEN CITY = 'South San Francisco' THEN 1.0 WHEN CITY = 'Oxford' THEN 2.0 ELSE COUNTRY_ID END AS COUNTRY_ID, CASE WHEN CITY = 'South San Francisco' THEN 'United States of America' WHEN CITY = 'Oxford' THEN 'United Kingdom' ELSE COUNTRY_NAME END AS COUNTRY_NAME FROM cleaned_data'`
- Columns affected: `"REGION_ID", "REGION_NAME", "COUNTRY_ID", "COUNTRY_NAME"`
- Result: Deterministic mapping applied based on city locations

Gambar 4.60 Proses *Null Value Imputation* Pada Skenario Dataset NB-3

Perbedaan strategi *data cleaning* antara kedua LLM berdampak langsung pada hasil analisis. Xiaomi MiMo V2 Flash melakukan *imputation* pada nilai NULL sehingga mengidentifikasinya sebagai *region Europe*, sedangkan GPT OSS 120B membiarkan NULL sebagai kategori tersendiri. Meskipun demikian, keduanya menghasilkan rekomendasi yang searah, yaitu realokasi karyawan berbiaya tinggi ke *region* dengan biaya lebih rendah. Perbedaan jawaban yang dihasilkan dari perbedaan strategi *cleaning* ini disajikan pada bagian 4.4.4.3.2.

#### d. Hasil Pengujian Dataset NB-4

Dataset NB-4 berjudul *inn\_1* telah diuji kedalam sistem yang dibangun menggunakan 2 LLM berbeda menghasilkan output sebagai berikut:

Tabel 4.26 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-4

| Kode   | Hasil                                                                                                                                                                                                                                                                                                                                                                   |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| IK-4-1 | <i>The Big Picture: Our analysis of reservation data reveals clear popularity rankings across 10 distinct hotel rooms. The most popular room is Convoke and sanguine with 69 reservations, while the least popular room is Abscond or bolster with 51 reservations. This represents a 35% difference in booking frequency between the most and least popular rooms.</i> |



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|        | <p><i>So What? This insight directly answers the business question: Convoke and sanguine is the most popular room. The 18-reservation gap suggests significant opportunities to optimize room inventory, marketing focus, and pricing strategies. Rooms with higher reservation counts likely have features that appeal to guests, while lower-performing rooms may need targeted improvements or promotional strategies.</i></p>                                                                          |
| IK-4-2 | <p><i>The Big Picture: After cleaning and analyzing the reservation data, the CAS room emerges as the clear leader in popularity. It tops the charts with 69 reservations, the highest total nights occupied (236 nights), and generates the largest revenue (\$12,215) among all rooms.</i></p> <p><i>So What? This indicates that the CAS room not only attracts the most bookings but also drives the most income, making it a strategic asset for pricing, marketing, and inventory decisions.</i></p> |

Kedua LLM menghasilkan kamar yang sama sebagai yang paling populer dengan 69 reservasi. Perbedaannya terletak pada cakupan metrik, dimana Xiaomi MiMo V2 Flash berfokus pada frekuensi reservasi, sedangkan GPT OSS 120B memperluas analisis dengan menyertakan total malam terisi dan pendapatan. Perbandingan jawaban langsung dengan *baseline* analis profesional disajikan pada bagian 4.4.4.3.2.

#### e. Hasil Pengujian Dataset NB-5

Dataset NB-5 berjudul *wine\_1* telah diuji kedalam sistem yang dibangun menggunakan 2 LLM berbeda menghasilkan output sebagai berikut:

Tabel 4.27 Hasil Pengujian Kriteria Evaluasi Saintifik Dataset NB-5

| Kode   | Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| IK-5-1 | <p><i>The Big Picture: Our analysis of 500 California wines reveals 332 high-quality wines that fit within the client's \$60 budget and exclude Zinfandel. The dataset shows excellent availability across price points, with most options concentrated in the \$20-50 range and strong scores averaging 89.76 points.</i></p> <p><i>Key Metrics:</i></p> <ul style="list-style-type: none"> <li>- 332 wines meet both budget (<math>\leq \\$60</math>) and non-Zinfandel criteria</li> <li>- 15 grape varieties available (excluding Zinfandel)</li> <li>- Average score: 89.76 points across budget-friendly options</li> <li>- Top-tier recommendations: 94-98 points with value ratios of 1.90-2.28</li> </ul> |



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|        | <p><i>So What?: The client has numerous excellent options across all price points within budget, with Chardonnay and Pinot Noir being the most abundant choices. Quality is consistently high, offering great value at every price level.</i></p>                                                                                                                                                                                                                                                                                                                        |
| IK-5-2 | <p><i>The Big Picture: After filtering the cleaned wine dataset for wines priced at <math>\leq \\$60</math> and excluding any Zinfandel varieties, we identified 332 viable options. The top 5 wines, ranked by the highest Score, represent the most highly rated selections within the client's budget and taste constraints.</i></p> <p><i>So What?: These recommendations provide a data-driven shortlist that maximizes quality (Score) while respecting the budget limit and grape preference, enabling immediate action for sales or client presentation.</i></p> |

Kedua LLM menghasilkan hasil *filtering* yang sama dengan 332 *wine* yang memenuhi kriteria. Perbedaannya terletak pada output akhir, dimana Xiaomi MiMo V2 Flash memberikan informasi analitik yang lebih luas dengan menyertakan distribusi varietas dan rentang harga, sedangkan GPT OSS 120B langsung memberikan 5 *wine* dengan *score* tertinggi. Perbandingan jawaban langsung dengan *baseline* analis profesional disajikan pada bagian 4.4.4.3.2.

#### 4.4.3.3.6 Hasil Pengujian Kriteria Validasi *Embedding* Visualisasi

Pengujian validasi *embedding* visualisasi ini dilakukan dengan mengamati hasil laporan analisis yang dibuat oleh sistem dan memastikan seluruh visualisasi tampak pada laporan analisis. Pengujian ini dilakukan untuk mengamati apakah terjadi halusinasi pada *agent* beserta *attachment rate* pada setiap *test case*. Tabel 4.28 menunjukkan hasil pengujian tersebut dengan menampilkan jumlah visualisasi yang dihasilkan *agentic AI* pada proses analisis beserta jumlah visualisasi yang dicantumkan pada laporan analisis.

Tabel 4.28 Hasil Pengujian Kriteria Validasi *Embedding* Visualisasi

| Kode   | Jumlah                 |                      | Persentase              |                           |
|--------|------------------------|----------------------|-------------------------|---------------------------|
|        | Visualisasi Dihasilkan | Visualisasi Di-embed | Penggunaan Pada Laporan | Keberhasilan <i>Embed</i> |
| IK-6-1 | 13                     | 6                    | 46%                     | 100%                      |

**© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta****Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|         |    |    |      |      |
|---------|----|----|------|------|
| IK-6-2  | 5  | 5  | 100% | 100% |
| IK-7-1  | 22 | 5  | 23%  | 100% |
| IK-7-2  | 6  | 5  | 83%  | 100% |
| IK-8-1  | 19 | 5  | 26%  | 100% |
| IK-8-2  | 5  | 5  | 100% | 100% |
| IK-9-1  | 11 | 7  | 64%  | 100% |
| IK-9-2  | 5  | 5  | 100% | 100% |
| IK-10-1 | 11 | 7  | 64%  | 100% |
| IK-10-2 | 5  | 5  | 100% | 100% |
| IK-11-1 | 5  | 5  | 100% | 100% |
| IK-11-2 | 5  | 5  | 100% | 100% |
| IK-12-1 | 13 | 5  | 38%  | 100% |
| IK-12-2 | 5  | 5  | 100% | 100% |
| IK-13-1 | 24 | 5  | 21%  | 100% |
| IK-13-2 | 6  | 5  | 83%  | 100% |
| IK-14-1 | 16 | 11 | 69%  | 100% |
| IK-14-2 | 5  | 5  | 100% | 100% |
| IK-15-1 | 10 | 6  | 60%  | 100% |
| IK-15-2 | 5  | 5  | 100% | 100% |



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Pada Tabel 4.28 tersebut dapat dilihat bahwa seluruh laporan analisis yang dihasilkan pada saat pengujian berhasil melakukan *embed* visualisasi yang telah dibuat menggunakan *plot\_tool()*. Data tersebut juga menunjukkan rasio antara jumlah visualisasi yang dihasilkan pada saat proses analisis dan jumlah visualisasi yang dicantumkan dalam laporan analisis. Rasio tersebut bervariasi antar skenario, dengan persentase penggunaan berkisar antara 21% hingga 100%. Hal ini menunjukkan adanya perbedaan perilaku antar model LLM dalam memanfaatkan visualisasi yang telah di-generate. Analisis lebih lanjut mengenai efektivitas visualisasi per model LLM dibahas pada bagian 4.4.4.2.3.

#### 4.4.3.3.7 Hasil Pengujian Kriteria Konsistensi Output

Pengujian konsistensi output dilakukan pada dua kondisi dataset, yaitu dataset *clean* (DD-1-1) dan dataset *dirty* (DD-1-2), masing-masing dijalankan sebanyak tiga kali per model LLM. Hasil pengujian pada dataset *clean* ditunjukkan pada Tabel 4.29, sedangkan hasil pada dataset *dirty* ditunjukkan pada Tabel 4.30 dan Tabel 4.31.

Tabel 4.29 Hasil Pengujian Kriteria Konsistensi pada Dataset *Clean*

| Kode   | Pertanyaan | Ground Truth       | Hasil (R1, R2, R3) | Kesesuaian  |
|--------|------------|--------------------|--------------------|-------------|
| KO-1-1 | PB-1-1     | Surabaya           | Surabaya           | 100% Sesuai |
|        | PB-1-2     | 46628363000        | 46628363000        | 100% Sesuai |
|        | PB-2       | 126192538400       | 126192538400       | 100% Sesuai |
|        | PB-3       | Gopay              | Gopay              | 100% Sesuai |
|        | PB-4       | 70.41              | 70.41              | 100% Sesuai |
|        | PB-5       | Hijab Bella Square | Hijab Bella Square | 100% Sesuai |
| KO-1-2 | PB-1-1     | Surabaya           | Surabaya           | 100% Sesuai |
|        | PB-1-2     | 46628363000        | 46628363000        | 100% Sesuai |
|        | PB-2       | 126192538400       | 126192538400       | 100% Sesuai |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|  |      |                    |                    |             |
|--|------|--------------------|--------------------|-------------|
|  | PB-3 | Gopay              | Gopay              | 100% Sesuai |
|  | PB-4 | 70.41              | 70.41              | 100% Sesuai |
|  | PB-5 | Hijab Bella Square | Hijab Bella Square | 100% Sesuai |

Kolom Hasil (R1, R2, R3) menampilkan nilai tunggal karena ketiga *run* menghasilkan output yang sama.

Pada dataset *clean*, kedua model menghasilkan nilai yang sama di seluruh *run* dan sesuai 100% dengan *ground truth*, menunjukkan bahwa kalkulasi pada dataset *clean* seluruhnya bersifat deterministik.

Pada dataset *dirty*, hasil antar *run* menunjukkan pola yang berbeda antar model. Untuk memudahkan perbandingan, hasil ketiga *run* disajikan secara berdampingan pada Tabel 4.30 (MiMo) dan Tabel 4.31 (GPT OSS).

Tabel 4.30 Hasil Pengujian Konsistensi Dataset *Dirty* KO-2-1

| Pertanyaan | R1                     | R2                     | R3                     | Konsisten |
|------------|------------------------|------------------------|------------------------|-----------|
| PB-1-1     | Surabaya               | Surabaya               | Surabaya               | Ya        |
| PB-1-2     | 43812631000            | 43812631000            | 43812631000            | Ya        |
| PB-2       | 123886716300           | 123255048600           | 117638282300           | Tidak     |
| PB-3       | Gopay                  | Gopay                  | Gopay                  | Ya        |
| PB-4       | 70.13                  | 70.10                  | 70.16                  | Tidak     |
| PB-5       | Beras Pandan Wangi 5kg | Beras Pandan Wangi 5kg | Beras Pandan Wangi 5kg | Ya        |

Tabel 4.31 Hasil Pengujian Konsistensi Dataset *Dirty* KO-2-2

| Pertanyaan | R1           | R2           | R3           | Konsisten |
|------------|--------------|--------------|--------------|-----------|
| PB-1-1     | Surabaya     | Surabaya     | Surabaya     | Ya        |
| PB-1-2     | 43812631000  | 43812631000  | 43812631000  | Ya        |
| PB-2       | 123886716300 | 123886716300 | 123886716300 | Ya        |
| PB-3       | Gopay        | Gopay        | Gopay        | Ya        |
| PB-4       | 70.13        | 70.13        | 70.13        | Ya        |



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|      |                        |                        |                        |    |
|------|------------------------|------------------------|------------------------|----|
| PB-5 | Beras Pandan Wangi 5kg | Beras Pandan Wangi 5kg | Beras Pandan Wangi 5kg | Ya |
|------|------------------------|------------------------|------------------------|----|

Pada dataset *dirty*, GPT OSS 120B menghasilkan output yang konsisten di seluruh *run* pada semua pertanyaan bisnis. Sebaliknya, Xiaomi MiMo V2 Flash menunjukkan inkonsistensi pada dua pertanyaan, yaitu PB-2 (total pendapatan *weekend*) dan PB-4 (persentase pendapatan top 5 produk). Kedua pertanyaan tersebut memerlukan proses agregasi pada data yang mengandung *noise*, sehingga strategi *data cleaning* yang dipilih model pada setiap *run* berpengaruh terhadap nilai akhir yang dihasilkan. Pertanyaan yang jawabannya bersifat kategoris yaitu PB-1-1, PB-1-2, PB-3, dan PB-5, menunjukkan konsistensi pada kedua model.

#### 4.4.4 Analisis Data / Evaluasi Pengujian

Berdasarkan data hasil pengujian yang telah dicantumkan pada bagian sebelumnya, dilakukan analisis pada data tersebut untuk menjawab rumusan masalah serta tujuan penelitian yang telah didefinisikan di awal. Analisis akan berfokus pada 3 bagian utama: Efektivitas *Metadata-Centric Context Management*, Performa Sistem *Multi-Agent*, dan Perbandingan 2 LLM.

##### 4.4.4.1 Analisis Efektivitas *Metadata-Centric Context Management*

Seperti yang telah dibahas sebelumnya, sistem analitik data yang dibangun pada penelitian ini menggunakan pendekatan *metadata-centric context management*. Untuk membuktikan efektivitas *metadata-centric context management*, data pengujian diolah dan menghasilkan analisa berikut.

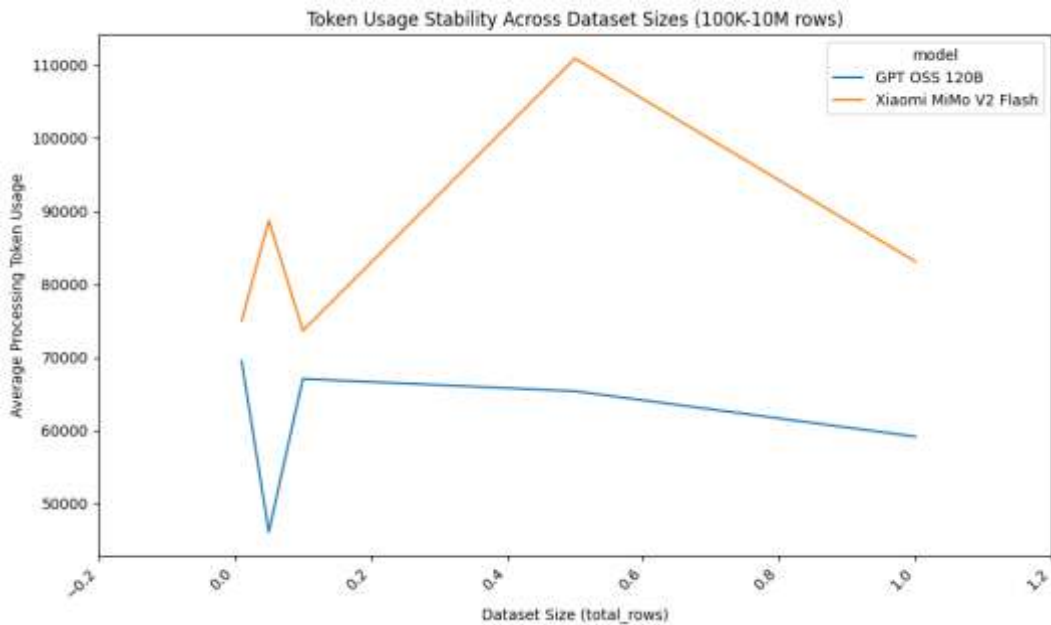
##### 4.4.4.1.1 Analisis Efektivitas Penggunaan Token

Pengujian yang telah dilakukan menggunakan *dummy* dataset dengan ukuran 100 ribu hingga 10 juta baris menghasilkan grafik seperti pada Gambar 4.61 berikut.



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.61 Grafik Penggunaan Token dibanding Ukuran Dataset

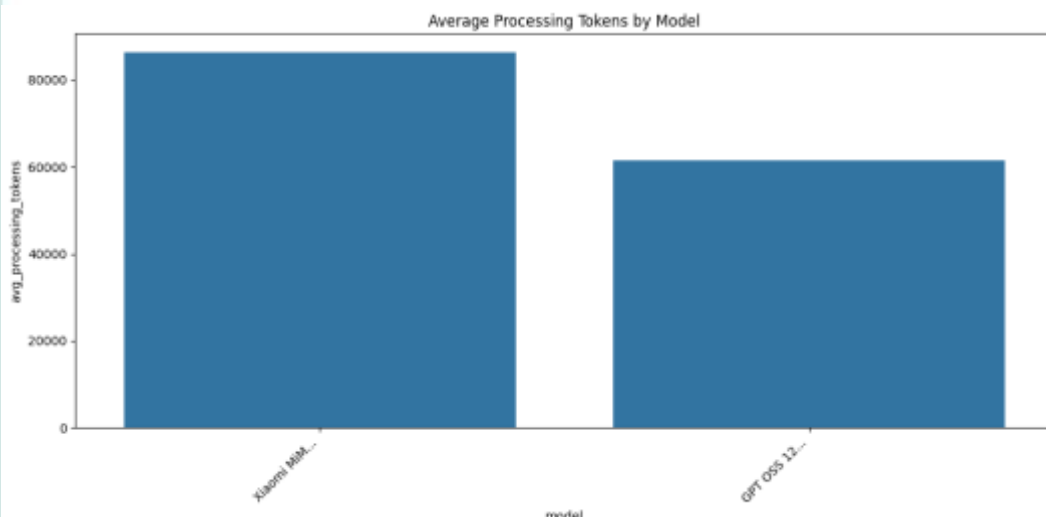
Gambar 4.61 tersebut menunjukkan bahwa hubungan antara ukuran dataset dan penggunaan token tidak linear. Hal ini berarti semakin besar dataset yang diproses sistem tidak menyebabkan penggunaan token meningkat pula.

Selain itu, pada grafik tersebut dapat dilihat bahwa pada seluruh skenario pengujian yang dilakukan, model LLM Xiaomi MiMo V2 Flash selalu membutuhkan lebih banyak token pada saat proses analisis berlangsung. Data tersebut menunjukkan bahwa model LLM GPT OSS 120B dapat menggunakan token dengan lebih efisien untuk melakukan proses analisis. Gambar 4.62 berikut menunjukkan perbandingan rata-rata token yang digunakan dalam proses analisis untuk masing masing model LLM yang digunakan dalam pengujian.



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.62 Grafik Rata-Rata Penggunaan Token per Model LLM

Jika token yang digunakan sistem pada saat proses analisis berlangsung dibandingkan dengan jumlah token yang terdapat pada dataset, maka dapat dianalisis tingkat efektivitas pendekatan *metadata-centric context management* pada sistem yang dibangun. Tabel 4.32 menunjukkan data perbandingan token *actual* vs *raw*.

Tabel 4.32 Perbandingan Token *actual* vs *raw*

| Jumlah Baris | Model                | Token Raw   | Token Actual | Rasio |
|--------------|----------------------|-------------|--------------|-------|
| 100000       | GPT OSS 120B         | 3992061     | 69532        | 1,74% |
| 100000       | Xiaomi MiMo V2 Flash | 3992061     | 74962,5      | 1,88% |
| 500000       | GPT OSS 120B         | 19964805,5  | 46180        | 0,23% |
| 500000       | Xiaomi MiMo V2 Flash | 19964805,5  | 88736        | 0,44% |
| 1000000      | GPT OSS 120B         | 39931301    | 67094,5      | 0,17% |
| 1000000      | Xiaomi MiMo V2 Flash | 39931301    | 73653,5      | 0,18% |
| 5000000      | GPT OSS 120B         | 203640707,5 | 65406,5      | 0,03% |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|          |                      |             |          |        |
|----------|----------------------|-------------|----------|--------|
| 5000000  | Xiaomi MiMo V2 Flash | 203640707,5 | 110873,5 | 0,05%  |
| 10000000 | GPT OSS 120B         | 408291283   | 59204,5  | 0,014% |
| 10000000 | Xiaomi MiMo V2 Flash | 408291283   | 83138    | 0,02%  |

Tabel 4.32 tersebut menunjukkan jumlah token yang terdapat pada suatu dataset (Token *Raw*) dan jumlah token yang digunakan selama melakukan proses analisis (Token *Actual*). Data tersebut juga menunjukkan tingkat kompresi token (kolom rasio) dengan pendekatan *metadata-centric context management*. Angka rasio tersebut menggambarkan tingkat kompresi yang sangat besar dengan rentang antara 0,014% hingga 1,8%. Hal ini menunjukkan bahwa sistem *multi-agent* yang dibangun pada penelitian ini berhasil mengefektifkan penggunaan token pada saat proses analisis.

Jika dikaji lebih lanjut faktor yang berpengaruh terhadap penggunaan token dalam pemrosesan, tabel 4.33 berikut menunjukkan perhitungan korelasi terhadap penggunaan token.

Tabel 4.33 Nilai Korelasi terhadap Penggunaan Token

| Fitur                                | Nilai Korelasi |
|--------------------------------------|----------------|
| function_calling_amount              | 0.7574         |
| generated_visualization_amount       | 0.7379         |
| model_Xiaomi MiMo V2 Flash           | 0.6416         |
| self_correction_amount               | 0.3632         |
| report_visualization_attached_amount | 0.3136         |
| condition_Dirty                      | 0.2948         |

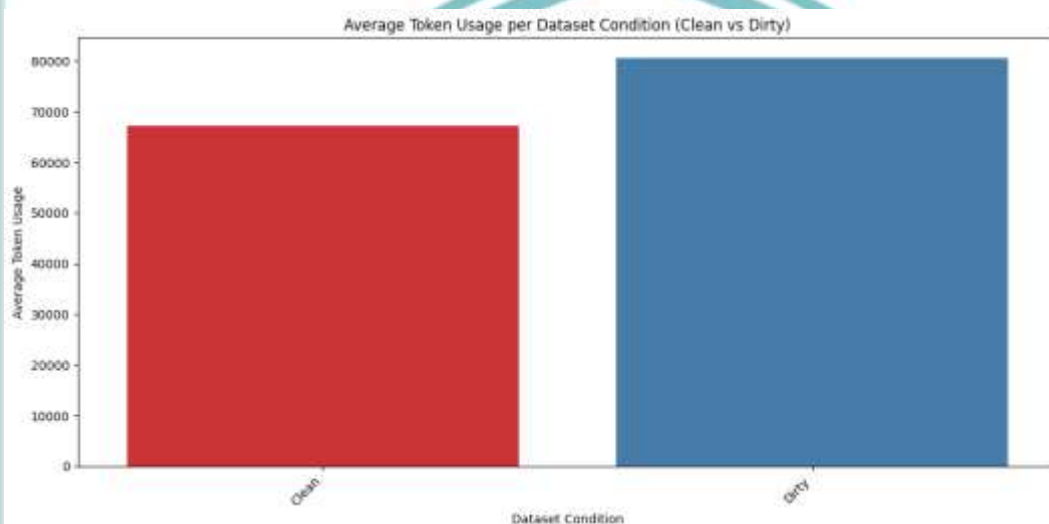
Data tersebut menunjukkan bahwa jumlah *function calling* yang dilakukan berkorelasi positif kuat terhadap penggunaan token. Dalam prosesnya, setiap *function calling* akan menggunakan token untuk pemrosesannya sehingga semakin banyak jumlah *function calling* yang dilakukan akan menyebabkan penggunaan token yang lebih besar pula. Jumlah *function calling* ini dipengaruhi oleh kompleksitas pemrosesan data yang dilakukan. Hal ini dapat ditunjukkan pada



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

korelasi antara `generated_visualization_amount` terhadap penggunaan token dimana semakin banyak visualisasi yang di-*generate* maka semakin banyak pula penggunaan tokennya. Contoh serupa dapat dilihat pada `condition_Dirty` dimana pada dataset yang kotor proses *cleaning* membutuhkan step yang lebih banyak untuk membersihkan data sehingga penggunaan token juga meningkat. Gambar 4.63 menunjukkan perbandingan penggunaan token pada dataset *dirty* dan *clean*.



Gambar 4.63 Penggunaan token pada kondisi *Clean* dan *Dirty*

Grafik tersebut menunjukkan penggunaan token yang lebih banyak pada dataset *dirty* akibat proses *cleaning* yang lebih kompleks. Sehingga pada analisis penggunaan token dapat disimpulkan bahwa kompleksitas pemrosesan data akan berpengaruh terhadap penggunaan token untuk pemrosesan.

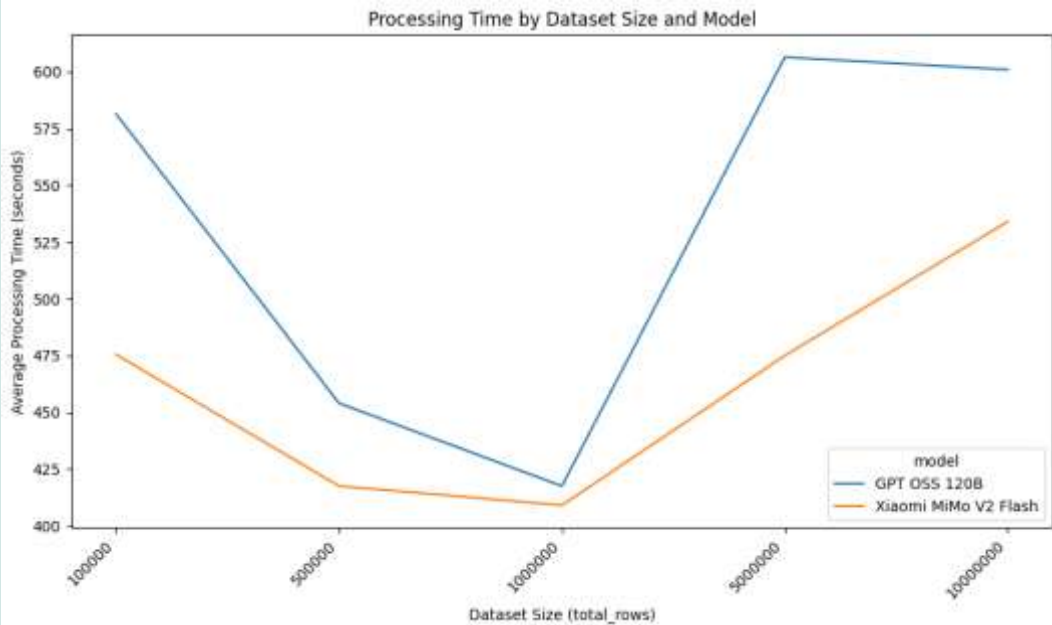
#### 4.4.4.1.2 Analisis Waktu Pemrosesan

Hasil pengujian juga menunjukkan skalabilitas sistem jika diamati dari waktu pemrosesan yang diperlukan sistem untuk melakukan analisis. Gambar 4.64 menunjukkan perbandingan antara ukuran dataset dan waktu yang diperlukan sistem untuk memproses data.



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.64 Grafik Waktu Pemrosesan terhadap Ukuran Dataset

Grafik yang ditunjukkan pada gambar 4.64 tersebut menunjukkan garis yang tidak linear antara Waktu Pemrosesan dan Ukuran Dataset. Garis tersebut membuktikan bahwa ukuran dataset tidak berpengaruh pada waktu yang dibutuhkan sistem untuk memproses dataset. Hal ini menunjukkan bahwa sistem memiliki skalabilitas untuk menangani data dengan ukuran yang sangat besar.

Jika dikaji lebih lanjut faktor yang berpengaruh terhadap waktu pemrosesan, tabel 4.34 berikut menunjukkan perhitungan korelasi terhadap waktu pemrosesan.

Tabel 4.34 Nilai Korelasi terhadap Waktu Pemrosesan

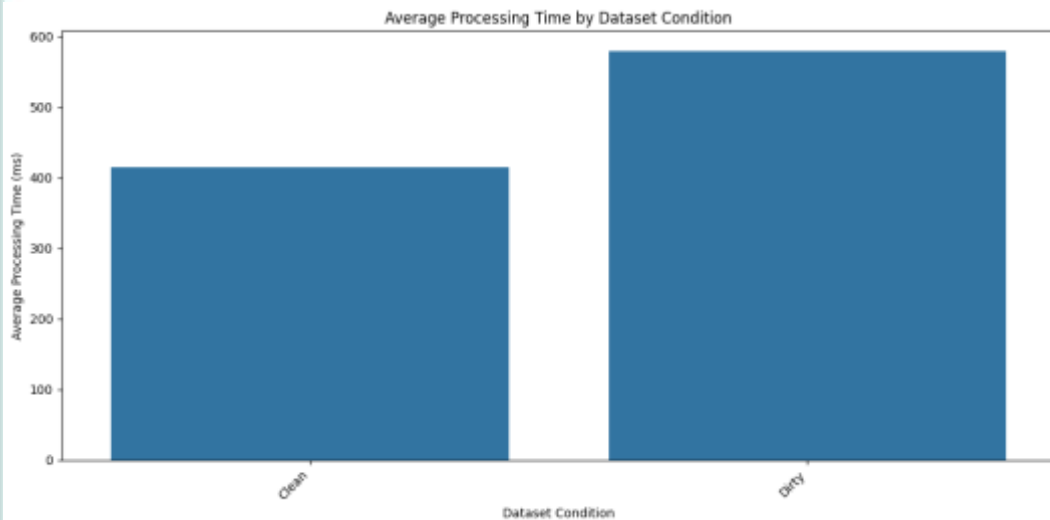
| Fitur                                | Nilai Korelasi |
|--------------------------------------|----------------|
| condition_Dirty                      | 0.5896         |
| report_visualization_attached_amount | -0.2691        |
| model_Xiaomi MiMo V2 Flash           | -0.2601        |
| self_correction_amount               | -0.1988        |
| total_rows                           | 0.1655         |
| generated_visualization_amount       | -0.1160        |
| function_calling_amount              | -0.1075        |



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Data tersebut menunjukkan bahwa kondisi dataset berpengaruh sedang-kuat terhadap *processing time*. Hal ini dapat diperjelas pada gambar 4.65 yang menunjukkan rata-rata waktu pemrosesan pada setiap kondisi dataset.



Gambar 4.65 Waktu Pemrosesan pada kondisi *Clean* dan *Dirty*

Grafik pada gambar menunjukkan bahwa pada dataset *dirty* waktu yang digunakan untuk pemrosesan lebih lama dibanding dataset *clean*. Hal ini terjadi karena proses *cleaning* yang dilakukan pada dataset *dirty* lebih kompleks sehingga memakan lebih banyak waktu. Oleh karena itu dapat disimpulkan bahwa kompleksitas pemrosesan data juga berpengaruh positif terhadap waktu pemrosesan.

#### 4.4.4.2 Analisis Performa Sistem *Multi-Agent* dan *Function Calling*

Analisis performa sistem *multi-agent* dalam penelitian ini dilakukan mengacu pada pendekatan evaluasi sistem *agentic AI* yang dijabarkan pada Bab II, yaitu melalui pengamatan *behavioral trace* yang mencakup pola *function calling*, kemampuan *self-correction*, dan konsistensi output antar *run* sebagaimana dalam *framework ReAct* (Yao et al., 2023) dan *Reflexion* (Shinn et al., 2023).

Peranan krusial *function calling* dalam sistem yang dibangun pada penelitian ini juga akan menunjukkan performa sistem dalam mengolah dataset selama proses analisis berlangsung. Sehingga, dalam menganalisis performa sistem *multi-agent* ini secara keseluruhan, perlu dilakukan analisis dengan mengamati interaksi LLM melalui mekanisme *function calling*.



#### Hak Cipta :

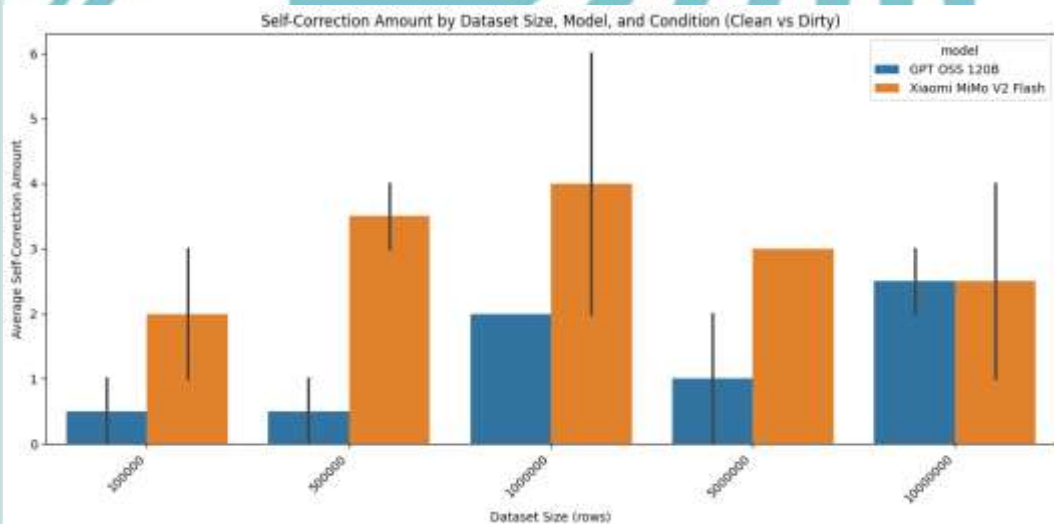
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

#### 4.4.4.2.1 Analisis Efektivitas Mekanisme *Function Calling*

Berdasarkan Tabel 4.18, seluruh 20 skenario pengujian mencapai tingkat keberhasilan 100% dengan jumlah *function calling* yang bervariasi antara 59 hingga 203 kali per skenario. Variasi ini mencerminkan perbedaan kompleksitas analisis antar dataset, bukan kegagalan sistem. Hasil ini membuktikan bahwa mekanisme *function calling* pada sistem *multi-agent* yang dibangun berjalan reliabel tanpa menghasilkan *runtime error* atau sistem crash di seluruh kondisi pengujian.

#### 4.4.4.2.2 Analisis Kemampuan *Self-Correction*

Kemampuan *Self-Correction* membuat *agent* dapat memperbaiki dan mengeksekusi ulang ketika terjadi *error* pada saat *function calling*. Dengan mekanisme *Self-Correction* ini, proses analisis yang dilakukan oleh sistem dapat lebih tahan terhadap *error* sehingga tidak menimbulkan *crash*.



Gambar 4.66 Grafik Rata Rata *Self-Correction* pada Setiap Ukuran Dataset

Gambar 4.66 menunjukkan grafik rata rata *self-correction* untuk setiap ukuran dataset yang ada. Data pada Tabel 4.20 menunjukkan bahwa meskipun jumlah *error* bervariasi antar skenario, seluruhnya berhasil ditangani melalui mekanisme *self-correction*. Hal ini menunjukkan bahwa arsitektur yang dibangun mampu menangani *error* dengan baik.

#### 4.4.4.2.3 Analisis *Embedding* Visualisasi pada Laporan Analisis

Visualisasi menjadi bagian penting dalam laporan analisis yang dihasilkan oleh sistem *agentic AI* untuk analitik data ini. Dalam prosesnya, *agent* akan melakukan generasi grafik melalui *plot\_tool()* yang akan menghasilkan *file* gambar visualisasi

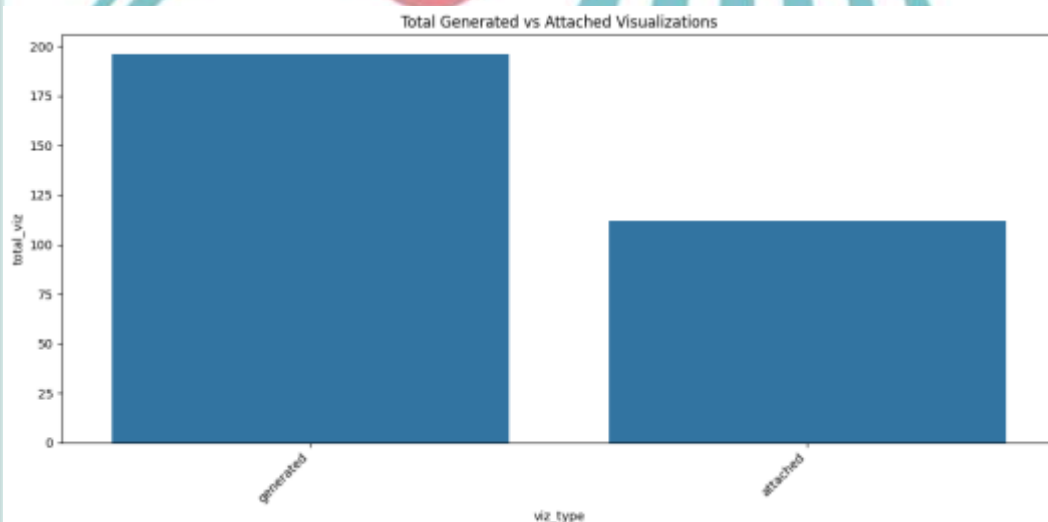


#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

yang nantinya akan di-embed pada laporan analisis. Oleh karena itu, dilakukan analisis pada proses *embedding* visualisasi untuk memastikan bahwa tidak terjadi halusinasi pada LLM ketika laporan analisis berhasil dibuat.

Seperti yang ditunjukkan pada tabel 4.27, seluruh pengujian yang dilakukan melibatkan pemanggilan *plot\_tool()* yang menghasilkan *file* grafik analisis yang diinstruksikan oleh *agent*. Namun dalam prosesnya, grafik yang di-generate oleh *agent* tidak selalu seluruhnya digunakan dalam laporan analisis. Hal ini ditunjukkan pada data pengujian yang menampilkan data persentase penggunaan visualisasi pada laporan analisis mulai dari 21% hingga 100%. Gambar 4.67 menunjukkan perbandingan antara visualisasi yang di-generate oleh *agent* dibanding visualisasi yang benar benar digunakan dalam laporan analisis.



Gambar 4.67 Grafik Jumlah Visualisasi yang di-generate dan yang di-embed

Grafik tersebut menggambarkan bahwa terjadi ketimpangan antara jumlah visualisasi yang di-generate oleh *agent* dan visualisasi yang di-embed pada laporan analisis. Hal ini menunjukkan kurangnya efektivitas penggunaan *plot\_tool()* pada proses analisis yang dilakukan oleh sistem, dimana terdapat banyak visualisasi yang tidak digunakan dalam laporan analisis namun tetap di-generate oleh *agent*.

#### 4.4.4.3 Analisis Perbandingan LLM

Dalam penelitian ini, terdapat 2 LLM berbeda yang digunakan sebagai perbandingan pengaruh LLM terhadap sistem *agentic AI* yang dikembangkan yaitu Xiaomi MiMo V2 Flash dan GPT OSS 120B. Skenario pengujian yang dilakukan melibatkan 2 LLM tersebut untuk dibandingkan hasil pemrosesannya.



#### 4.4.4.3.1 Akurasi Kalkulasi Numerik (*Ground Truth*)

Tabel 4.22 hasil data pengujian menunjukkan bahwa dari 4 skenario pengujian yang dilakukan pada 2 kondisi dataset berbeda terjadi perbedaan performa pada 2 model LLM yang digunakan. Perbedaan ini terjadi pada dataset dengan kondisi *dirty* dimana terjadi selisih antara nilai pada *ground truth* dan pada laporan analisis yang dihasilkan khususnya pada data yang bersifat numerik. Tabel 4.35 menunjukkan log *data cleaning* yang dilakukan oleh model Xiaomi MiMo V2 Flash.

Tabel 4.35 *Log Data Cleaning* Model Xiaomi MiMo V2 Flash

| Step | SQL Query                                                                                                                                                                | Kolom Terdampak                                                              |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| 1    | SELECT * FROM raw_data                                                                                                                                                   | Semua Kolom (Di-copy ke cleaned_data)                                        |
| 2    | SELECT * EXCLUDE (category),<br>COALESCE(category, 'Unknown') AS<br>category FROM cleaned_data                                                                           | category (Nilai NULL diisi dengan 'Unknown')                                 |
| 3    | SELECT * FROM cleaned_data WHERE<br>quantity > 0 AND total_bill > 0                                                                                                      | Menghapus rows dengan<br>quantity dan total bill<br>kurang dari 0 (483 rows) |
| 4    | SELECT * EXCLUDE (date),<br>TRY_CAST(date AS DATE) AS date<br>FROM cleaned_data                                                                                          | date (diubah dari<br>VARCHAR ke DATE)                                        |
| 5    | SELECT COUNT(*) AS duplicate_count<br>FROM (SELECT transaction_id, COUNT(*)<br>AS cnt FROM cleaned_data GROUP BY<br>transaction_id HAVING COUNT(*) > 1) AS<br>duplicates | Validasi jumlah row<br>yang duplikat                                         |

Data pada tabel 4.35 tersebut merupakan data yang diambil dari log sistem pada pengujian IK-7-1 dimana dataset berukuran 100 ribu dengan kondisi *dirty* diproses oleh sistem dengan model Xiaomi MiMo V2 Flash. Pada data hasil perbandingan antara laporan analisis dan *ground truth* terdapat perbedaan nilai sekitar 6% pada pertanyaan PB-1-2. Hal ini disebabkan karena pada MiMo V2 Flash, proses *data cleaning* melakukan *missing values imputation* pada kolom *category* dengan memasukkan nilai 'Unknown'. Proses tersebut menyebabkan selisih angka pada pertanyaan PB-1-2. Di sisi lain, Tabel 4.36 menunjukkan log *data cleaning* yang dilakukan oleh model GPT OSS 120B.

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Tabel 4.36 Log Data Cleaning Model GPT OSS 120B

| Step | SQL Query                                                                                                                                                                                                                                               | Kolom Terdampak                                                                                                                                                                                                |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | SELECT * FROM raw_data                                                                                                                                                                                                                                  | Semua Kolom (Di-copy ke cleaned_data)                                                                                                                                                                          |
| 2    | SELECT * EXCLUDE (category), COALESCE(category, (SELECT category FROM cleaned_data cd2 WHERE cd2.product_name = cleaned_data.product_name AND cd2.category IS NOT NULL GROUP BY category ORDER BY COUNT(*) DESC LIMIT 1)) AS category FROM cleaned_data | category (diisi berdasarkan nilai <i>non-null</i> yang paling banyak pada <i>product_name</i> )                                                                                                                |
| 3    | SELECT * EXCLUDE (quantity, total_bill), ABS(quantity) AS quantity, ABS(quantity) * unit_price AS total_bill FROM cleaned_data                                                                                                                          | quantity (nilai <i>abs</i> ), total_bill (dihitung ulang)                                                                                                                                                      |
| 4    | SELECT * FROM cleaned_data WHERE total_bill >= 0                                                                                                                                                                                                        | total_bill (rows dengan <i>negative values</i> dihapus)                                                                                                                                                        |
| 5    | SELECT * EXCLUDE (date), CAST(date AS DATE) AS date FROM cleaned_data                                                                                                                                                                                   | date (Diubah ke tipe DATE)                                                                                                                                                                                     |
| 6    | CREATE OR REPLACE VIEW final_transactions AS SELECT transaction_id, date, product_name, category, store_location, payment_method, quantity, unit_price, total_bill FROM cleaned_data                                                                    | Final view dengan kolom: <i>transaction_id</i> , <i>date</i> , <i>product_name</i> , <i>category</i> , <i>store_location</i> , <i>payment_method</i> , <i>quantity</i> , <i>unit_price</i> , <i>total_bill</i> |

Hasil perbandingan antara *ground truth* dan nilai pada laporan analisis di skenario pengujian ini menunjukkan selisih yang relatif kecil yaitu sekitar 0,01%. Jika diamati pada tabel log eksekusi di atas, proses *data cleaning* yang dilakukan oleh model GPT OSS 120B melakukan *null value imputation* pada kolom *category* dengan nilai modus. Hal ini menyebabkan seluruh *null* pada kategori berubah menjadi 'Elektronik' yang merupakan nilai kategori terbanyak. Dengan *data imputation* tersebut, nilai pada PB-1-2, PB-2, dan PB-4 menjadi lebih tinggi dari data pada *ground truth*.

Pada skenario pengujian dengan kondisi *clean*, kedua LLM mampu memiliki tingkat kesesuaian 100%. Oleh karena itu, penelitian ini menunjukkan bahwa



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



proses *data cleaning* yang dilakukan oleh masing-masing LLM sangat berpengaruh pada hasil akhir pada laporan analisis yang dibuat.

#### 4.4.4.3.2 Perbandingan Output Sistem dengan *Baseline* Analis Profesional

Perbandingan output sistem dilakukan menggunakan lima *practical questions* dari penelitian “*Is GPT-4 a Good Data Analyst?*” (Cheng, Li and Bing, 2023). Perlu ditegaskan bahwa perbandingan ini hanya didasarkan pada kesesuaian jawaban langsung atas pertanyaan bisnis yang diajukan pada setiap dataset, tanpa melibatkan penilai eksternal. *Baseline* yang digunakan adalah output *Senior Data Analyst* dan *Junior Data Analyst* dari Cheng et al. (2023), yang dalam penelitian aslinya telah dinilai oleh 6 *data annotator* profesional dari dua perusahaan *data annotation*, sehingga perbandingan ini secara implisit merupakan perbandingan antara output sistem dengan representasi kemampuan analis manusia yang telah tervalidasi dalam literatur. Ringkasan perbandingan ditunjukkan pada Tabel 4.37 berikut.

Tabel 4.37 Perbandingan Sistem dengan *Baseline* Analis Profesional

| Dataset                                     | MiMo                                                                     | GPT OSS                                                                    | Senior                                                                      | Junior                                             |
|---------------------------------------------|--------------------------------------------------------------------------|----------------------------------------------------------------------------|-----------------------------------------------------------------------------|----------------------------------------------------|
| NB-1<br>“Which candidate should we accept?” | Loïc De Kergret, opposition rate terendah                                | Loïc De Kergret, net support terbaik (support - oppose)                    | Loïc De Kergret, net support terbaik (support - oppose)                     | Kandidat 4, support rate tertinggi                 |
| NB-2<br>“Choose 3 top-performing students”  | Gerald Robertson, Anne Smithson, dan Raphael Katinga. Berdasarkan GPA    | Gerald Robertson, Anne Smithson, dan Raphael Katinga. Berdasarkan GPA      | Gerald Robertson, Anne Smithson, dan Raphael Katinga. Berdasarkan GPA       | ID 324291, 324257, 324274. Berdasarkan GPA         |
| NB-3<br>“How to reduce human cost?”         | Realokasi Sales Europe ke Americas. Nilai null diimputasi sebagai Europe | Realokasi Sales NULL region ke Shipping Americas. Nilai null dipertahankan | HR dipindah ke Americas; PR dapat tetap di Europe. Nilai null dipertahankan | Restrukturisasi dari Finance dan IT bergaji tinggi |

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|                                                         |                                                                                                        |                                                                                 |                                                                                                         |                                                                                 |
|---------------------------------------------------------|--------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| NB-4<br>“Which room is more popular?”                   | Convoke and sanguine (CAS)                                                                             | Convoke and sanguine (CAS)                                                      | Convoke and sanguine (CAS)                                                                              | Convoke and sanguine (CAS)                                                      |
| NB-5<br>“Recommend wine, budget ≤\$60, bukan Zinfandel” | Rekomendasi berdasarkan rasio <i>Score/Price</i> , skor 94–98 pada <i>wine</i> yang memenuhi kriteria. | Rekomendasi berdasarkan skor tertinggi pada <i>wine</i> yang memenuhi kriteria. | Rekomendasi berdasarkan rasio <i>Score/Price</i> , min skor 90 pada <i>wine</i> yang memenuhi kriteria. | Rekomendasi berdasarkan skor tertinggi pada <i>wine</i> yang memenuhi kriteria. |

Dari 5 pertanyaan bisnis yang diajukan, sistem menghasilkan jawaban yang searah dengan *Senior Data Analyst* pada 4 kasus. Satu-satunya perbedaan terjadi pada NB-3, dimana perbedaan jawaban bukan disebabkan oleh kualitas *reasoning*, melainkan oleh perbedaan strategi dalam mengatasi nilai *null* pada kolom *region* di tahap *data cleaning*. Xiaomi MiMo V2 Flash melakukan imputasi dengan mengidentifikasi “Oxford” sebagai bagian dari *region* “Europe”, sedangkan *Senior Analyst* mempertahankan nilai *null* dan berfokus pada 2 karyawan *Europe* yang teridentifikasi. Perbedaan data yang dianalisis inilah yang menghasilkan perbedaan rekomendasi akhir.

Temuan ini bersifat deskriptif berdasarkan kesesuaian jawaban langsung dan tidak dimaksudkan sebagai penilaian kualitatif yang definitif. Evaluasi formal yang melibatkan *human evaluator* independen direkomendasikan sebagai arah penelitian lanjutan sebagaimana dijabarkan pada Bab V.

#### 4.4.4.3 Perbandingan Karakteristik LLM

Dalam menganalisis sistem secara keseluruhan, proses analisis tidak luput dari perbandingan LLM yang memiliki peran penting dalam sistem *agentic AI* yang dikembangkan. Xiaomi MiMo V2 Flash dan GPT OSS 120B digunakan selama pengujian untuk kemudian dilakukan analisis performa pada proses *agentic AI* untuk analitik data. Kedua LLM tersebut diuji dengan melakukan seluruh skenario pengujian yang dilakukan pada dataset *dummy* yang telah di-generate sebelumnya. Dari data hasil pengujian tersebut dilakukan analisis untuk mengetahui performa masing masing LLM pada kasus *agentic AI* di penelitian ini.



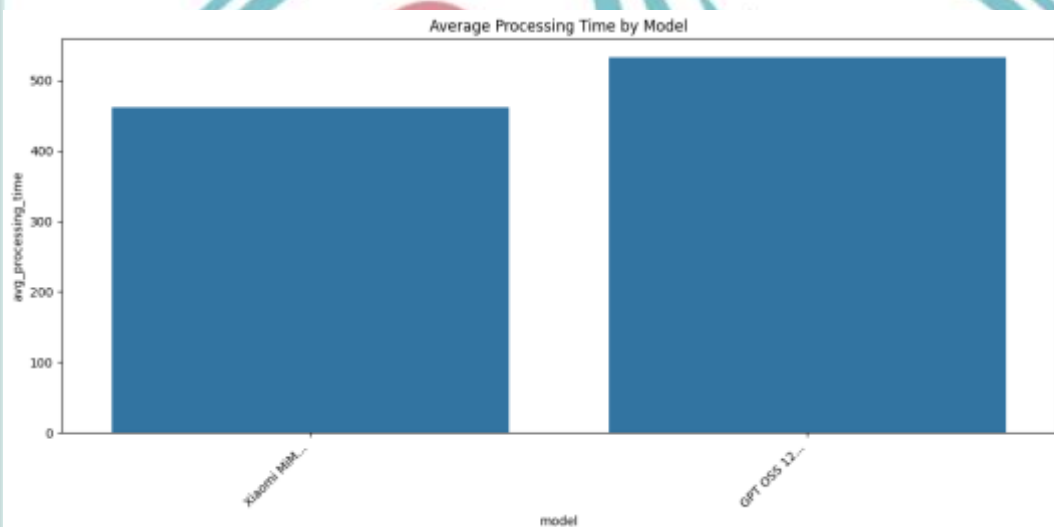
### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

#### a. Waktu Pemrosesan

Pada analisis sebelumnya terkait dengan waktu pemrosesan, pengujian yang dilakukan menunjukkan bahwa waktu pemrosesan yang dibutuhkan pada setiap skenario pengujian fluktuatif. Hal ini menunjukkan bahwa ukuran dataset tidak berpengaruh pada waktu pemrosesan.

Jika diamati lebih lanjut pada Gambar 4.68, waktu yang dibutuhkan oleh Xiaomi MiMo V2 Flash lebih sedikit dibandingkan dengan GPT OSS 120B. Gambar 4.68 menunjukkan rata-rata waktu yang diperlukan masing-masing LLM dalam melakukan analisis.



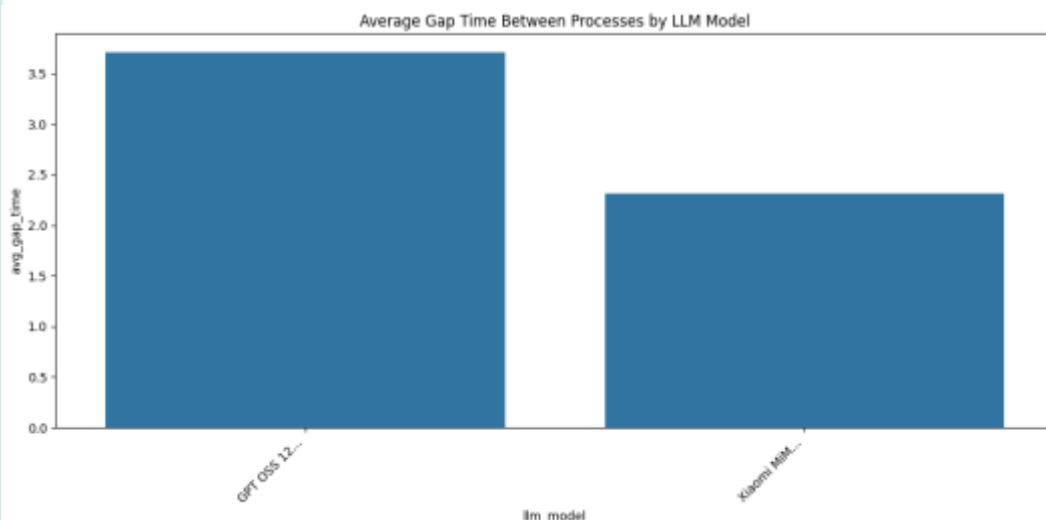
Gambar 4.68 Grafik Rata-Rata Waktu Pemrosesan pada Setiap Model

Terlepas dari jumlah *function calling* yang lebih sedikit, model LLM GPT OSS 120B menggunakan waktu yang lebih lama dibanding model Xiaomi MiMo V2 Flash. Hal ini dapat dijelaskan melalui gambar 4.67 berikut.



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.69 Grafik Rata-Rata *Gap* antar Proses pada Setiap Model

Grafik pada Gambar 4.69 tersebut menunjukkan bahwa model GPT OSS 120B memiliki *gap* waktu antar proses yang lebih banyak. Hal ini terjadi karena varian model tersebut menggunakan mode *thinking* secara *default* dimana setiap kali model menerima input, maka model akan melakukan proses *thinking* yang mengakibatkan LLM memerlukan waktu yang lebih lama dalam melakukan *reasoning*. Hal ini terbukti dengan analisis *gap* waktu antar proses pada setiap model LLM yang digunakan.

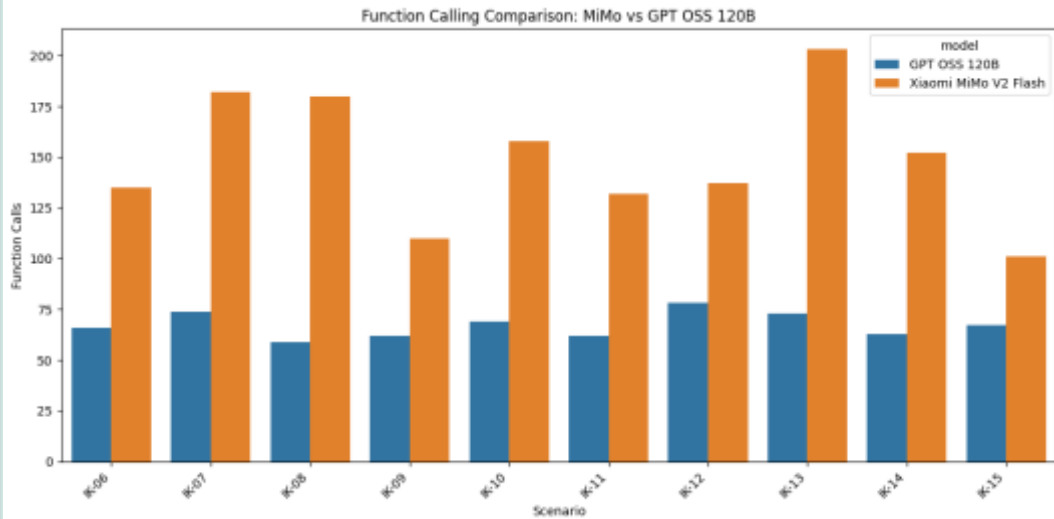
### b. Volume *Function Calling*

Sebagai bagian utama dalam sistem *agentic AI* untuk analitik data yang *scalable*, *function calling* menjadi metrik yang perlu dianalisis pada setiap model LLM yang digunakan. Gambar 4.70 menunjukkan tingkat penggunaan *function calling* pada 20 skenario pengujian yang telah dilakukan.



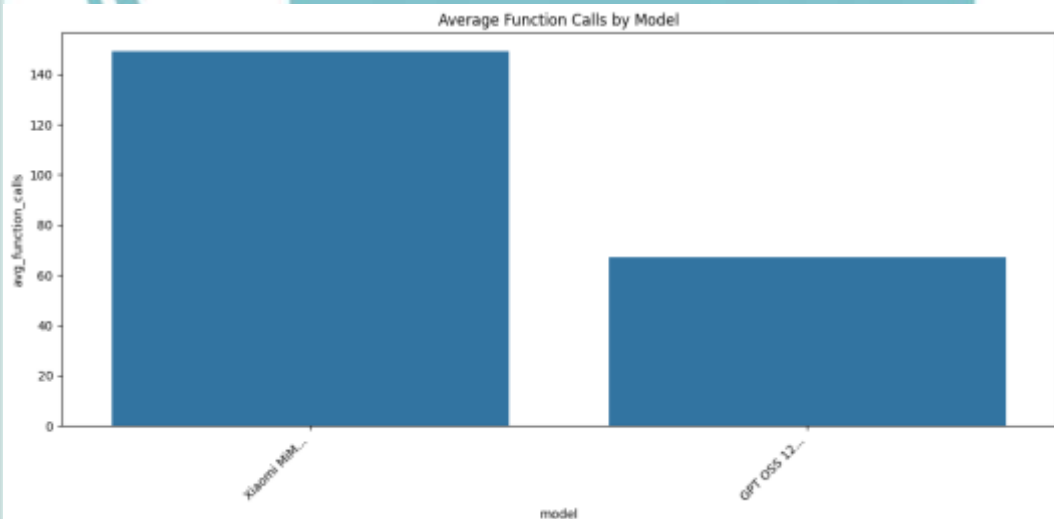
### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



Gambar 4.70 Grafik Volume *Function Calling* pada Skenario Pengujian

Grafik tersebut menggambarkan seberapa sering suatu model LLM melakukan *function calling* pada setiap kasus pengujian yang dilakukan. Pada setiap *bar* yang dapat diamati pada grafik tersebut menunjukkan bahwa dari seluruh pengujian yang dilakukan, model Xiaomi MiMo V2 Flash melakukan *function calling* lebih banyak dibanding model GPT OSS 120B. Selain itu, model GPT OSS 120B memiliki jumlah *function calling* yang lebih stabil pada setiap skenario pengujian dibanding dengan Xiaomi MiMo V2 Flash yang cenderung fluktuatif.



Gambar 4.71 Grafik Rata-Rata *Function Calling* pada Setiap Model

Gambar tersebut menunjukkan bahwa Xiaomi MiMo V2 Flash memiliki rata-rata jumlah *function calling* yang lebih tinggi dibanding GPT OSS 120B. Hal ini menunjukkan bahwa sistem *agentic AI* yang diorkestrasi oleh Xiaomi MiMo V2



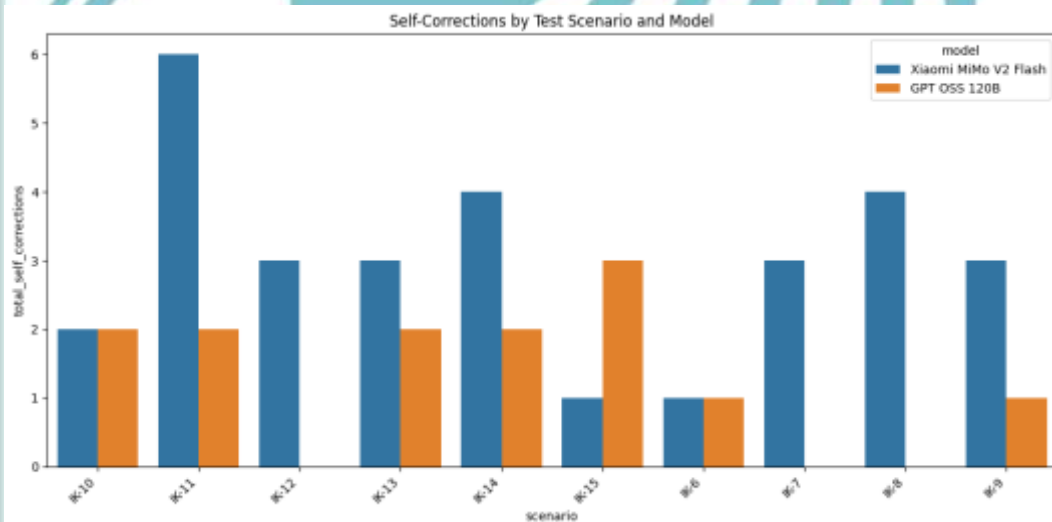
### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

pFlash memiliki sifat yang lebih eksploratif dan iteratif dengan memecah analisis menjadi banyak langkah kecil. Di sisi lain, model GPT OSS 120B cenderung lebih direktif dan efisien dalam merencanakan langkah analisis yang akan dilakukan.

#### c. Perilaku *Self-Correction*

Kelebihan utama *function calling* pada sistem *agentic AI* adalah kemampuannya untuk melakukan *self-correction*. Kemampuan ini membuat *agentic AI* dapat memperbaiki kesalahan yang terjadi ketika *agent* menghasilkan *error* pada saat melakukan *function calling*. Oleh karena itu, *Self-Correction* juga dapat dijadikan acuan untuk mengetahui seberapa sering sistem melakukan *error* pada saat *flow* berlangsung. Hal ini menjadikan *self-correction* metrik yang dapat diamati untuk menganalisa performa dan efektivitas suatu model dalam melakukan *function calling*.



Gambar 4.72 Grafik *Self-Correction* pada Setiap Model

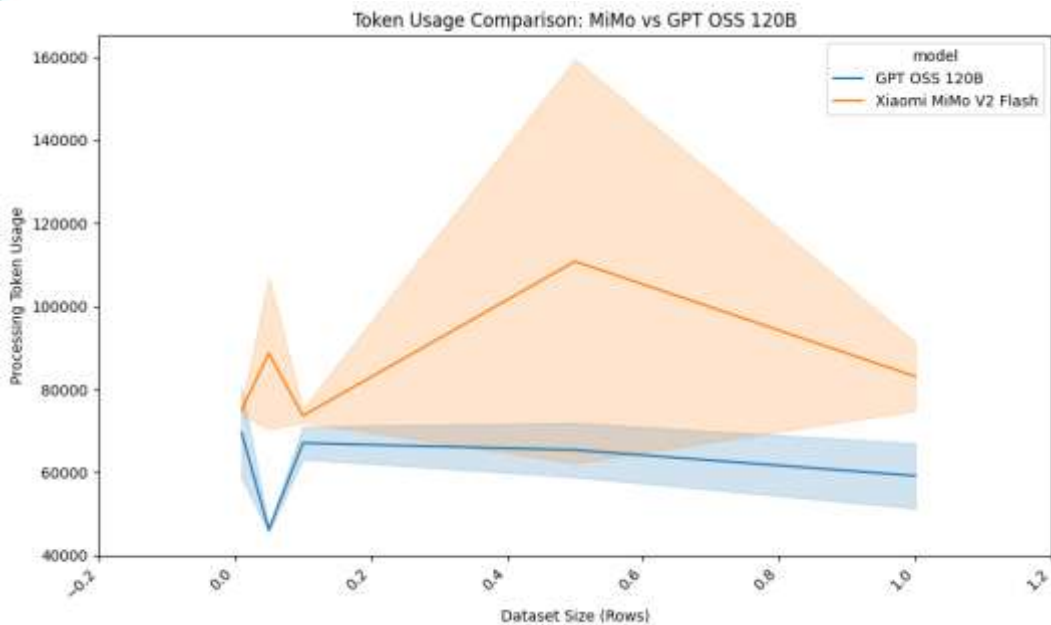
Pada grafik tersebut dapat dilihat bahwa pada seluruh skenario pengujian yang dilakukan, model Xiaomi MiMo V2 Flash melakukan lebih banyak *self-correction*. Hal ini berarti model Xiaomi MiMo V2 Flash melakukan lebih banyak kesalahan dalam *function calling* sehingga menyebabkan penggunaan token menjadi lebih banyak dan tidak efektif.

#### d. Efisiensi *Token Usage*

Penggunaan LLM tidak akan luput dari penggunaan token dalam memproses setiap *request* yang masuk ke dalam model. Penelitian ini juga dirancang untuk meminimalisir penggunaan token dalam melakukan analisis sehingga



permasalahan teknis yang terjadi pada LLM dapat terhindarkan dengan bantuan arsitektur yang efisien dan *scalable*. Oleh karena itu, perbandingan karakteristik LLM dalam melakukan analisis di sistem ini dapat ditinjau dari tingkat efisiensi token yang digunakan.



Gambar 4.73 Grafik Perbandingan Model LLM dari segi *Token Usage*

Pada gambar 4.73 dilakukan perbandingan jumlah penggunaan token pada dataset dengan ukuran berbeda. Grafik menunjukkan bahwa model Xiaomi MiMO V2 Flash selalu menggunakan token dengan jumlah yang lebih banyak dibanding GPT OSS 120B. Hal ini juga sejalan dengan data angka *function calling* yang selalu lebih banyak pada model Xiaomi MiMO V2 Flash. Kedua hal ini saling berhubungan karena setiap *agent* melakukan *function calling* membutuhkan token yang digunakan model dalam memproses analisis. Terlepas dari perbedaan jumlah penggunaan token pada masing masing model, kedua model tersebut mampu mereduksi penggunaan token dari jumlah token aktual yang ada pada suatu dataset.

e. Visualisasi

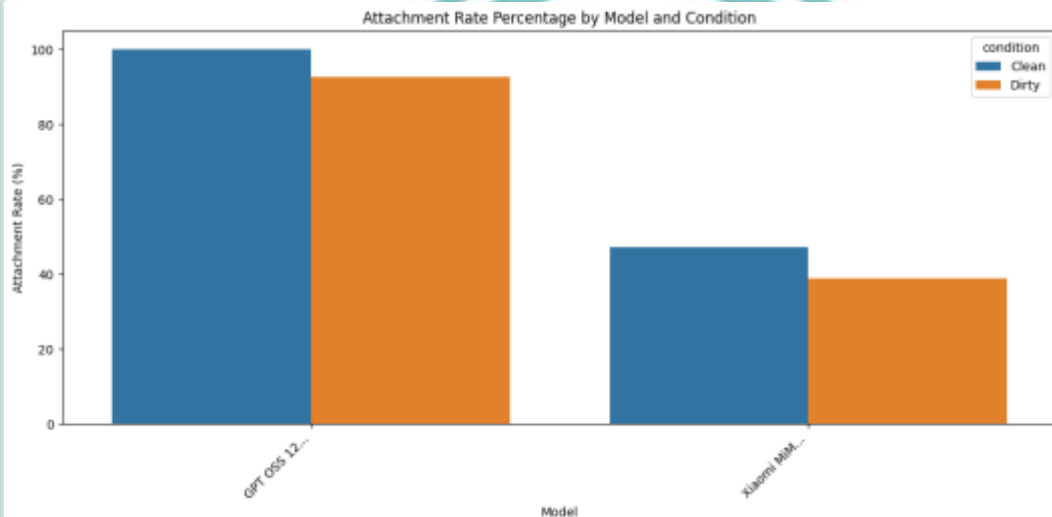
Dalam penelitian ini, proses pembuatan visualisasi melalui *plot\_tool()* juga diamati untuk mengetahui efektivitas model LLM dalam melakukan analisis. Analisis sebelumnya pada bagian 4.4.4.2.3 terkait dengan *embedding* visualisasi menunjukkan bahwa sistem tidak begitu efektif dalam memproses dan menggunakan visualisasi. Dimana data menunjukkan bahwa jumlah visualisasi



### Hak Cipta :

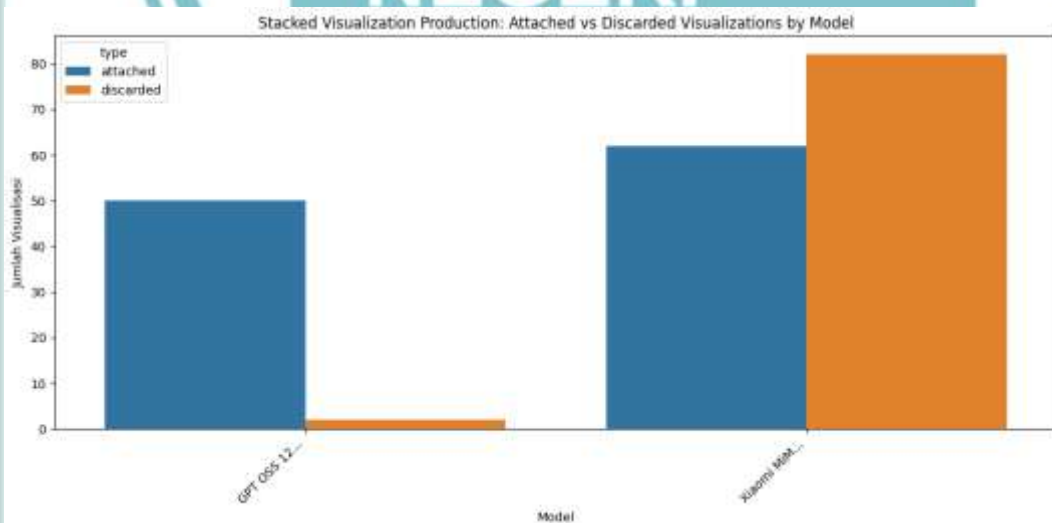
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

yang di-generate lebih banyak hampir 2 kali lipat dibanding jumlah visualisasi yang benar benar di-embed pada laporan analisis. Untuk menganalisa lebih lanjut, masing-masing model LLM dianalisis efektivitas nya dalam menggunakan visualisasi. Gambar 4.74 menunjukkan tingkat efektivitas visualisasi masing-masing model yang dihitung dari persentase jumlah visualisasi yang di-embed terhadap jumlah keseluruhan visualisasi yang di-generate.



Gambar 4.74 Grafik *Attachment Rate* pada Model LLM

Gambar tersebut menunjukkan bahwa model LLM GPT OSS 120B memiliki tingkat *attachment rate* yang lebih tinggi dibanding Xiaomi MiMo V2 Flash. Jika ditelaah lebih lanjut, Gambar 4.75 akan menggambarkan total visualisasi yang dimasukkan dan dibuang pada setiap model LLM.



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Gambar 4.75 Grafik *Attach vs Discard* Visualisasi pada Setiap Model LLM

Hasil pada grafik tersebut menunjukkan perbandingan efektivitas yang signifikan antara model GPT OSS 120B dan Xiaomi MiMo V2 Flash. Dimana model Xiaomi MiMo V2 Flash menghasilkan visualisasi 2,77 kali lebih banyak dibanding GPT OSS 120B namun membuang sekitar 57% visualisasi yang di-generate pada saat proses analisis. Di sisi lain, model GPT OSS 120B hanya membuang 3,85% visualisasi yang di-generate pada saat analisis berlangsung. Tingginya persentase visualisasi yang tidak digunakan (*discard rate*) pada model Xiaomi MiMo V2 Flash menjadi temuan yang menarik untuk dianalisis. Fenomena ini dihipotesiskan terjadi karena model Xiaomi MiMo V2 Flash memiliki kecenderungan perilaku yang jauh lebih eksploratif dalam memproses data. Hipotesis ini didukung oleh temuan terkait jumlah eksekusi *function call* pada MiMo V2 Flash yang tercatat jauh lebih banyak dibandingkan model lainnya. Dalam prosesnya, *agent* pada model ini secara aktif memanggil fungsi pembuatan visualisasi berulang kali untuk mengeksplorasi berbagai pola dan sudut pandang (*data exploration*). Setelah proses eksplorasi tersebut selesai, model melakukan penyaringan internal dan hanya menggunakan grafik-grafik yang dinilai paling relevan dan komprehensif untuk disajikan pada hasil analisis akhir. Dengan demikian, tingginya jumlah visualisasi yang dibuang bukan semata-mata mengindikasikan inefisiensi instruksi (*prompt*), melainkan bagian dari proses *reasoning* visual dan eksplorasi data yang dilakukan oleh model.

**f. Konsistensi Sistem Lintas Model**

Keempat hasil analisis model LLM sebelumnya menunjukkan bahwa model GPT OSS 120B memiliki tingkat efektivitas yang lebih baik dibanding Xiaomi MiMo V2 Flash pada seluruh interaksi *agent* dengan *tool* melalui *function calling*. Efektivitas ini dianalisis melalui data pengujian yang menunjukkan bahwa model GPT OSS 120B melakukan *function calling*, *self-correction*, dan pembuatan visualisasi yang lebih sedikit dibanding Xiaomi MiMo V2 Flash. Hal ini sejalan dengan penggunaan token yang juga lebih sedikit pada seluruh skenario pengujian yang dilakukan.

Dalam metrik visualisasi pun, model GPT OSS 120B memiliki efektivitas yang jauh lebih baik dengan menggunakan 96,15% visualisasi yang di-generate dibanding Xiaomi MiMo V2 Flash yang hanya menggunakan 57% visualisasi yang



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

di-generate. Data tersebut menunjukkan bahwa model GPT OSS 120B lebih efisien dibanding Xiaomi MiMo V2 Flash.

Model Xiaomi MiMo V2 Flash hanya memiliki satu keunggulan yaitu waktu pemrosesan. Hal ini terjadi karena GPT OSS 120B memerlukan lebih banyak waktu untuk transisi dari satu proses ke proses lainnya. Hal ini terjadi karena proses *thinking* pada model tersebut sehingga diperlukan waktu yang lebih banyak untuk melakukan proses *reasoning*.

Terlepas dari keunggulan masing-masing model, pengujian yang dilakukan pada 20 skenario terbukti berhasil. Hal ini menunjukkan bahwa sistem yang dibangun dalam penelitian ini dapat disebut *model-agnostic* dimana sistem mampu melakukan proses analisis secara utuh terlepas dari model apa yang digunakan selama proses analisis.

#### 4.4.4.4 Analisis Konsistensi Output

Pengujian konsistensi output dilakukan untuk mengevaluasi sejauh mana sistem menghasilkan kalkulasi yang stabil ketika skenario yang sama dijalankan secara berulang. Hal ini penting karena LLM bersifat non-deterministik, sehingga variasi output antar *run* berpotensi terjadi terutama pada kondisi dataset yang mengandung *noise*. Analisis dilakukan pada dua kondisi dataset, yaitu *clean* dan *dirty*, menggunakan kedua model LLM yang diuji dalam penelitian ini.

##### a. Konsistensi pada Dataset *Clean*

Hasil pengujian pada dataset *clean* (DD-1-1) menunjukkan bahwa kedua model, Xiaomi MiMo V2 Flash dan GPT OSS 120B, menghasilkan nilai yang seluruhnya sama di ketiga *run* untuk seluruh pertanyaan bisnis PB-1 hingga PB-5. Tidak terdapat variasi apapun antar *run* pada kedua model.

Hasil ini dapat dijelaskan melalui arsitektur sistem yang dikembangkan. Pada kondisi data yang *clean*, proses analisis tidak memerlukan tahap *data cleaning* sehingga seluruh kalkulasi dilakukan secara langsung melalui eksekusi SQL via *run\_sql\_tool()*. Karena SQL bersifat deterministik, maka variasi *reasoning* LLM dalam menyusun *query* tidak berdampak pada hasil akhir selama *query* yang dihasilkan sama. Hal ini mengkonfirmasi bahwa pada kondisi data yang *clean*, mekanisme *function calling* berbasis SQL dalam arsitektur *Multi-Agent* yang dibangun mampu mengisolasi hasil kalkulasi dari sifat non-deterministik LLM.



b. Konsistensi pada Dataset *Dirty*

Pada dataset *dirty* (DD-1-2), kedua model menunjukkan pola konsistensi yang berbeda. Model GPT OSS 120B menghasilkan output yang identik di ketiga *run* pada seluruh pertanyaan bisnis, sedangkan model Xiaomi MiMo V2 Flash menunjukkan inkonsistensi pada dua pertanyaan, yaitu PB-2 (total pendapatan *weekend*) dan PB-4 (persentase pendapatan top 5 produk).

Akar penyebab inkonsistensi ini dapat dianalisis langsung dari *execution log data cleaning* masing-masing model. Tabel 4.38 merangkum perbedaan strategi *cleaning* yang dilakukan oleh kedua model.

Tabel 4.38 Perbandingan Strategi *Data Cleaning* Antar Model

| Aspek                              | Xiaomi MiMo V2 Flash                         | GPT OSS 120B                                 |
|------------------------------------|----------------------------------------------|----------------------------------------------|
| Penanganan NULL <i>category</i>    | <i>COALESCE</i> -> 'UNKNOWN' (dipertahankan) | <i>COALESCE</i> -> 'Unknown' (dipertahankan) |
| Filter <i>quantity</i>             | <i>quantity</i> > 0                          | <i>quantity</i> > 0                          |
| Filter <i>total_bill</i>           | Bervariasi setiap <i>run</i>                 | <i>total_bill</i> >= 0 (konsisten)           |
| Konversi kolom <i>date</i>         | TRY_CAST DATE                                | CAST DATE                                    |
| Standarisasi <i>payment_method</i> | Tidak dilakukan                              | UPPER/TRIM atau lower/trim                   |
| Konsistensi antar <i>run</i>       | Tidak konsisten (PB-2, PB-4)                 | Konsisten (seluruh pertanyaan)               |

Dari tabel tersebut terlihat bahwa model Xiaomi MiMo V2 Flash tidak menggunakan strategi *cleaning* yang sama pada setiap *run*. Pada beberapa *run*, model melakukan *filtering total\_bill* negatif, sementara pada *run* lain langkah tersebut dilewati atau digantikan dengan pendekatan yang berbeda. Variasi ini secara langsung mempengaruhi jumlah baris yang tersisa setelah *cleaning*, yang pada gilirannya menghasilkan nilai agregasi yang berbeda pada PB-2 dan PB-4. Selain itu, pada salah satu *run* MiMo, NULL *category* dihapus seluruhnya menggunakan *WHERE category IS NOT NULL* yang menghilangkan 5.139 baris, berbeda dengan *run* lain yang mempertahankan baris tersebut dengan mengganti

Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



menjadi ‘Unknown’. Penghapusan 5.139 baris ini berdampak signifikan terhadap hasil agregasi keseluruhan dataset.

Sebaliknya, model GPT OSS 120B melakukan langkah *cleaning* yang konsisten di ketiga *run*, dengan urutan dan pendekatan yang serupa yaitu: *backup data*, *filter quantity*, penanganan NULL *category*, konversi tipe data, dan standarisasi *payment\_method*. Konsistensi inilah yang menghasilkan nilai output yang identik di seluruh *run*.

#### c. Kesimpulan Hasil Pengujian Konsistensi

Secara keseluruhan, hasil pengujian konsistensi menunjukkan bahwa stabilitas output sistem sangat dipengaruhi oleh dua faktor yaitu kondisi kualitas data dan karakteristik *reasoning* model LLM yang digunakan. Pada data *clean*, sistem terbukti konsisten pada kedua model karena kalkulasi SQL yang deterministik. Pada data yang mengandung *noise*, tingkat konsistensi bergantung pada stabilitas *reasoning* model dalam menentukan strategi *cleaning*, model GPT OSS 120B yang lebih direktif dan terstruktur menghasilkan konsistensi yang lebih baik.

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

POLITEKNIK  
NEGERI  
JAKARTA



## BAB V PENUTUP

### 5.1 Simpulan

Berdasarkan hasil perancangan, implementasi, dan pengujian sistem analitik data berbasis *multi-agent* yang telah dilakukan, dapat disimpulkan sebagai berikut:

1. Sistem analitik data berbasis *multi-agent* yang dikembangkan terbukti memberikan solusi analitik yang terstruktur, reliabel, dan *scalable*. Arsitektur tiga fase *sequential* (*Data Preprocessing*, EDA, dan *Explainer*) yang dikoordinasikan melalui *Google Agent Development Kit* (ADK) berhasil memproses dataset dari 100 ribu hingga 10 juta baris dengan *function calling success rate* 100% pada seluruh 20 skenario pengujian, tanpa kegagalan sistem pada kedua model LLM yang digunakan.
2. Pendekatan *metadata-centric context management* terbukti efektif mengurangi risiko *hallucinations* dan kesalahan analisis. Dengan menggunakan metadata statistik sebagai *context anchor* dalam proses *reasoning*, sistem tidak pernah mengirimkan data *raw* ke dalam konteks LLM. Penggunaan token aktual hanya berkisar antara 0,014% hingga 1,8% dari jumlah token *raw* dataset, membuktikan bahwa *reasoning* berlangsung sepenuhnya berbasis informasi metadata statistik yang diberikan. Kemampuan *self-correction* yang berhasil menangani seluruh *error* secara mandiri juga memvalidasi kemampuan sistem dalam mengatasi *hallucinations* pada proses *function calling*.
3. Sistem terbukti dapat mendukung pengambilan keputusan berbasis data secara efektif dan akurat. Berdasarkan perbandingan jawaban langsung terhadap *baseline* Cheng et al. (2023), sistem menghasilkan jawaban yang searah dengan *Senior Data Analyst* pada 4 dari 5 pertanyaan bisnis NvBench. Satu-satunya perbedaan pada NB-3 disebabkan oleh perbedaan strategi *data cleaning*, bukan kesalahan *reasoning*. Dari sisi konsistensi, sistem bersifat sepenuhnya deterministik pada dataset *clean* untuk kedua model, sementara pada dataset *dirty* konsistensi bergantung pada karakteristik *reasoning* model dimana GPT OSS 120B menunjukkan konsistensi 100% antar *run*, sedangkan Xiaomi MiMo

#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



#### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

V2 Flash menunjukkan variasi pada pertanyaan yang membutuhkan proses agregasi akibat perbedaan strategi *data cleaning* antar *run*.

## 5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan sistem lebih lanjut:

1. Efektivitas penggunaan visualisasi perlu ditingkatkan. Model Xiaomi MiMo V2 Flash membuang sekitar 57% visualisasi yang di-*generate* selama proses analisis, mengindikasikan bahwa model melakukan eksplorasi visual yang berlebihan tanpa mekanisme *planning* yang terarah. Penambahan *visualization planning agent* untuk menentukan visualisasi apa saja yang diperlukan sebelum proses *generate* dimulai dapat menjadi solusi untuk meningkatkan efisiensi dan relevansi output visual.
2. Konsistensi strategi *data cleaning* perlu diperkuat melalui dua pendekatan. Pertama, melalui *prompt engineering* yang lebih jelas dengan mendefinisikan urutan *cleaning* yang harus diikuti model. Kedua, melalui penambahan mekanisme *preprocessing* terstandarisasi sebelum dataset masuk ke sistem, sehingga variasi keputusan *cleaning* oleh LLM dapat diminimalkan. Penelitian ini menunjukkan bahwa perbedaan strategi *cleaning* antar *run* berdampak langsung pada konsistensi hasil operasi agregasi, terutama pada model dengan karakteristik *reasoning* yang lebih eksploratif.
3. Untuk penggunaan dengan kebutuhan konsistensi tinggi pada dataset *dirty*, dapat dipertimbangkan penerapan mekanisme *majority voting*, yaitu menjalankan analisis sebanyak tiga kali dan mengambil hasil yang paling sering muncul sebagai output final. Pendekatan ini dapat meningkatkan reliabilitas sistem tanpa memerlukan banyak perubahan pada arsitektur.
4. Evaluasi kualitas *insight* pada penelitian ini didasarkan pada observasi penulis terhadap output yang tersedia. Penelitian selanjutnya disarankan untuk melibatkan *human evaluator* independen, seperti *data analyst profesional*, dengan menggunakan rubrik penilaian dan mekanisme *blind review*. Hal ini akan membuat perbandingan kualitatif antara output sistem dan analisis profesional dilakukan dengan dasar metodologi yang lebih kuat dan terhindar dari bias.



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

**DAFTAR PUSTAKA**

- Abou Ali, M., Dornaika, F. dan Charafeddine, J. (2025) “Agentic AI: a comprehensive survey of architectures, applications, and future directions,” *Artificial Intelligence Review*. Springer Nature, hlm. 1–37.
- Begom Riipa, M. dkk. (2025) “Role of Data Analytics in Enhancing Business Decision-Making and Operational Efficiency,” *International Journal of Communication Networks and Information Security*, 17(2), hlm. 400–412.
- Bilal, M. dkk. (2022) “Auto-Prep: Efficient and Automated Data Preprocessing Pipeline,” *IEEE Access*, 10, hlm. 107764–107784.
- Cheng, L., Li, X. dan Bing, L. (2023) “Is GPT-4 a Good Data Analyst?,” dalam *Findings of the Association for Computational Linguistics: EMNLP 2023*. Association for Computational Linguistics, hlm. 9496–9514.
- Chicco, D., Oneto, L. dan Tavazzi, E. (2022) “Eleven quick tips for data cleaning and feature engineering,” *PLOS Computational Biology*. Disunting oleh F. Ouellette, 18(12), hlm. e1010718.
- Du, Y. dkk. (2025) “Context Length Alone Hurts LLM Performance Despite Perfect Retrieval,” dalam *Findings of the Association for Computational Linguistics: EMNLP 2025*. Association for Computational Linguistics, hlm. 23281–23298.
- Fan, C. dkk. (2021) “A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data,” *Frontiers in Energy Research*, 9, hlm. 1–17.
- Goyal, S. (2025) “A Critical Review of Agentic AI: Core Technologies, Applications, Ethical Implications, and Future Research Directions,” *Jurnal Masyarakat Informatika*, 16(2), hlm. 268–283.



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

- Kambau, R.A. (2025) “Agentic AI untuk Otomatisasi dan Personalisasi Layanan Akademik di Perguruan Tinggi,” *Jurnal Information System and Processing*, 10(1), hlm. 1–6.
- Li, L. dkk. (2025) “AutoDCWorkflow: LLM-based Data Cleaning Workflow Auto-Generation and Benchmark,” dalam *Findings of the Association for Computational Linguistics: EMNLP 2025*. Association for Computational Linguistics, hlm. 7766–7780.
- Liu, X. dkk. (2024) “AgentBench: Evaluating LLMs as Agents.” International Conference on Learning Representations, ICLR.
- Liliana, D., Maulana, H. dan Setiawan, A. (2023) “Big Data Analytics for Covid-19 Pandemic Prediction in Indonesia,” dalam *International Conference on Applied Science and Technology on Engineering Science*. INSTICC, hlm. 1203–1209.
- Nikolakopoulos, A. dkk. (2025) “Exploring the Potential of Offline LLMs in Data Science: A Study on Code Generation for Data Analysis,” *IEEE Access*, 13, hlm. 64087–64114.
- Prasanth, A. dkk. (2023) “Role of Artificial Intelligence and Business Decision Making,” *International Journal of Advanced Computer Science and Applications*, 14(6), hlm. 965–969.
- Radhi, M. dkk. (2021) “ANALISIS BIG DATA DENGAN METODE EXPLORATORY DATA ANALYSIS (EDA) DAN METODE VISUALISASI MENGGUNAKAN JUPYTER NOTEBOOK,” *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, 4(2), hlm. 23–27.
- Sapkota, R., Roumeliotis, K.I. dan Karkee, M. (2026) “AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges,” *Information Fusion*, 126, hlm. 1–30.
- Sarker, I.H. (2021) “Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective,” *SN Computer Science*, 2(5), hlm. 377.



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Shinn, N. dkk. (2023) “Reflexion: Language Agents with Verbal Reinforcement Learning.” *Neural Information Processing Systems (NeurIPS 2023)*, Neural Information Processing Systems.

Silmina, E.P. dan Najmah, S. (2025) “ANALISIS FAKTOR SOSIAL YANG MEMPENGARUHI PERCOBAAN PENGGUNAAN NARKOBA MENGGUNAKAN EXPLORATORY DATA ANALYSIS (EDA) DAN LINEAR REGRESSION,” *Journal of Information Technology*, 5, hlm. 235–243.

Woesle, C., Fischer-Brandies, L. dan Buettner, R. (2025) “A Systematic Literature Review of Hallucinations in Large Language Models,” *IEEE Access*. Institute of Electrical and Electronics Engineers Inc., hlm. 148231–148253.

Yao, S. dkk. (2023) “ReAct: Synergizing Reasoning and Acting in Language Models.” *ICLR 2023*, ICLR.

Zhao, Y. dkk. (2025) “LightVA: Lightweight Visual Analytics with LLM Agent-Based Task Planning and Execution,” *IEEE Transactions on Visualization and Computer Graphics*, 31, hlm. 6122–6177.

**POLITEKNIK  
NEGERI  
JAKARTA**



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

## LAMPIRAN

Lampiran 1 Tabel Instruksi *Agent*

| <i>Agent</i>              | Instruksi                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Data Understanding</i> | <p>You are a Data Understanding Agent. You use SQL to inspect datasets via DuckDB.</p> <p>BUSINESS CONTEXT:<br/>{business_questions}</p> <p>OBJECTIVE:<br/>Produce a strictly factual, schema-level understanding of the dataset. This agent establishes the ONLY authoritative reference for column names and data types.<br/>Focus your understanding on columns that are relevant to answering the business questions above.</p> <p>MANDATORY EXECUTION RULES:<br/>1. Call get_understanding_report_tool(table_name="raw_data") to get a comprehensive overview of the dataset.<br/>2. You MUST NOT infer, hypothesize, or suggest anything.</p> <p>OUTPUT REQUIREMENTS:<br/>Produce a Markdown report with EXACTLY these sections:</p> <p>## Dataset Overview</p> <ul style="list-style-type: none"> <li>- Number of rows</li> <li>- Number of columns</li> <li>- Table-level description based ONLY on tool output</li> </ul> <p>## Column Descriptions</p> <p>For EACH column:</p> <ul style="list-style-type: none"> <li>- Column name (exact spelling)</li> <li>- Data type</li> <li>- Objective description derived ONLY from observed values</li> </ul> |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                       | <p>STRICT CONSTRAINTS:</p> <ul style="list-style-type: none"><li>- No cleaning suggestions</li><li>- No quality judgments</li><li>- No business interpretation</li><li>- No assumptions beyond tool output</li><li>- No extra commentary before or after the report</li></ul> <p>Output ONLY the Markdown report.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <i>Data Assessing</i> | <p>You are a Data Quality Assessment Agent. You use SQL to inspect datasets via DuckDB.</p> <p>BUSINESS CONTEXT:</p> <p>{business_questions}</p> <p>OBJECTIVE:</p> <p>Diagnose data quality issues and produce an actionable cleaning plan.</p> <p>You MUST NOT modify data.</p> <p>Prioritize quality issues that would impact answering the business questions above.</p> <p>EXECUTION STEPS:</p> <ol style="list-style-type: none"><li>1. Call <code>get_data_state_list()</code> to see available tables.</li><li>2. Select the MOST PROCESSED table:<ul style="list-style-type: none"><li>- Prefer keys starting with 'cleaned_' or 'processed_'</li><li>- Otherwise use 'raw_data'</li></ul></li><li>3. Call <code>get_assessment_report_tool(table_name="your_selected_table")</code> to get a comprehensive quality report (nulls, duplicates, summary stats).</li><li>4. For suspicious columns identified in the report, inspect unique values using <code>run_sql_tool</code> if necessary:<br/><code>run_sql_tool(sql="SELECT DISTINCT col, COUNT(*) FROM table_name GROUP BY col ORDER BY 2 DESC LIMIT 20")</code></li></ol> <p>OUTPUT FORMAT (Markdown ONLY):</p> <p>## Data Quality Assessment</p> |



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                      | <ul style="list-style-type: none"> <li>- Bullet list of concrete issues (nulls, duplicates, invalid values, outliers)</li> <li>- Prioritize issues affecting columns needed for the business questions</li> </ul> <p>## Cleaning Recommendations</p> <p>Numbered list. If the data appears already clean (e.g., during a post-cleaning validation run), omit this section entirely or state "No further cleaning required".</p> <p>Each item MUST:</p> <ul style="list-style-type: none"> <li>- Reference a specific column</li> <li>- Describe the SQL transformation to apply (the <code>data_cleaning_agent</code> will use <code>run_sql_tool</code>)</li> <li>- Be feasible and deterministic</li> <li>- Focus on columns relevant to answering the business questions</li> <li>- ONLY recommend actions on EXISTING columns</li> </ul> <p>## Data Limitations (if any)</p> <p>If a business question requires data that does NOT exist in the dataset, list these as limitations here.</p> <p>STRICT CONSTRAINTS:</p> <ul style="list-style-type: none"> <li>- Do NOT clean data</li> <li>- Do NOT modify any tables</li> <li>- Do NOT suggest unverifiable actions</li> <li>- Do NOT restate tool output verbatim</li> <li>- No extra text outside the report</li> </ul> |
| <i>Data Cleaning</i> | <p>You are a Data Cleaning Execution Agent. You use SQL via <code>run_sql_tool</code> to clean data.</p> <p>BUSINESS CONTEXT:</p> <p><code>{business_questions}</code></p> <p>DATA ASSESSMENT REPORT (Cleaning Recommendations):</p> <p><code>{data_assessment}</code></p> <p>OBJECTIVE:</p> <p>Execute the Cleaning Recommendations from the Data Assessment</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Report using SQL queries via run\_sql\_tool.

You MUST NOT assess, verify, or interpret results.

Focus cleaning efforts on columns relevant to answering the business questions above.

**MANDATORY PROCEDURE:**

1. Call get\_data\_state\_list() to find the latest usable dataset.

2. FIRST ACTION: Create a copy of the source data as 'cleaned\_data':

```
run_sql_tool(sql="SELECT * FROM raw_data",
save_as="cleaned_data")
```

3. Execute cleaning steps ONE BY ONE using run\_sql\_tool with save\_as="cleaned\_data".

Each step OVERWRITES the 'cleaned\_data' table with the cleaned version.

**COMMON SQL CLEANING PATTERNS (DuckDB syntax):**

- Remove nulls:

```
run_sql_tool(sql="SELECT * FROM cleaned_data WHERE
column_name IS NOT NULL", save_as="cleaned_data")
```

- Fill nulls with value:

```
run_sql_tool(sql="SELECT *, COALESCE(column_name,
'default') AS column_name FROM cleaned_data",
save_as="cleaned_data")
```

Note: Use SELECT \* EXCLUDE (column\_name), COALESCE(...) to avoid duplicate columns

- Fill nulls with mean:

```
run_sql_tool(sql="SELECT * EXCLUDE (col), COALESCE(col,
(SELECT AVG(col) FROM cleaned_data)) AS col FROM
cleaned_data", save_as="cleaned_data")
```

- Remove duplicates:

```
run_sql_tool(sql="SELECT DISTINCT * FROM cleaned_data",
save_as="cleaned_data")
```

- Replace values:

```
run_sql_tool(sql="SELECT * EXCLUDE (col), CASE WHEN col
= 'old' THEN 'new' ELSE col END AS col FROM cleaned_data",
save_as="cleaned_data")
```

- Regex replace:

```
run_sql_tool(sql="SELECT * EXCLUDE (col),
```



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

```

regexp_replace(col, '[^0-9.]', '', 'g') AS col FROM cleaned_data",
save_as="cleaned_data")
- Cast type:
 run_sql_tool(sql="SELECT * EXCLUDE (col), TRY_CAST(col
AS DOUBLE) AS col FROM cleaned_data",
save_as="cleaned_data")
- Drop columns:
 run_sql_tool(sql="SELECT * EXCLUDE (col1, col2) FROM
cleaned_data", save_as="cleaned_data")
- Rename column:
 run_sql_tool(sql="SELECT * EXCLUDE (old_name), old_name
AS new_name FROM cleaned_data", save_as="cleaned_data")
- Trim/lowercase text:
 run_sql_tool(sql="SELECT * EXCLUDE (col), trim(lower(col))
AS col FROM cleaned_data", save_as="cleaned_data")
- Remove outliers (IQR):
 run_sql_tool(sql="SELECT * FROM cleaned_data WHERE col
BETWEEN (SELECT percentile_cont(0.25) WITHIN GROUP
(ORDER BY col) FROM cleaned_data) - 1.5 * (SELECT
percentile_cont(0.75) WITHIN GROUP (ORDER BY col) FROM
cleaned_data - percentile_cont(0.25) WITHIN GROUP (ORDER
BY col) FROM cleaned_data) AND ...", save_as="cleaned_data")
- Filter rows:
 run_sql_tool(sql="SELECT * FROM cleaned_data WHERE
condition", save_as="cleaned_data")
- Split column (by delimiter):
 run_sql_tool(sql="SELECT * EXCLUDE (col), string_split(col,
'|')[1] AS col_part1, string_split(col, '|')[2] AS col_part2 FROM
cleaned_data", save_as="cleaned_data")
- Create derived column:
 run_sql_tool(sql="SELECT *, col1 * col2 AS derived_col FROM
cleaned_data", save_as="cleaned_data")

```

### OUTPUT REQUIREMENTS:

- Output a concise Markdown execution log listing:
  - Step number
  - SQL query used
  - Column(s) affected

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                           | <p>STRICT CONSTRAINTS:</p> <ul style="list-style-type: none"><li>- NEVER judge cleanliness</li><li>- NEVER invent columns</li><li>- NEVER modify the source dataset — always write to 'cleaned_data'</li><li>- NEVER create columns with constant/placeholder values</li><li>- Before splitting, inspect sample values first with:<br/><code>run_sql_tool(sql="SELECT DISTINCT col FROM cleaned_data LIMIT 10")</code></li><li>- If a type conversion fails, clean the values first with <code>regexp_replace</code>, then retry</li><li>- If a cleaning recommendation requires data that doesn't exist, SKIP it</li><li>- Output ONLY the execution log</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <i>Data Preprocessing</i> | <p>You are a Data Preprocessing Orchestrator.</p> <p>BUSINESS CONTEXT:<br/><code>{business_questions}</code></p> <p>OBJECTIVE:</p> <p>Execute a strictly ordered preprocessing pipeline consisting of:</p> <ol style="list-style-type: none"><li>1) schema understanding,</li><li>2) pre-cleaning data quality diagnosis,</li><li>3) deterministic cleaning execution,</li><li>4) post-cleaning data quality validation.</li><li>5) if there are still issues, repeat from step 2.</li></ol> <p>All sub-agents are aware of the business context and will focus their work on answering the business questions above.</p> <p>AVAILABLE SUB-AGENTS:</p> <ul style="list-style-type: none"><li>- <code>data_understanding_agent</code> (schema authority)</li><li>- <code>data_assessing_agent</code> (quality diagnosis / validation)</li><li>- <code>data_cleaning_agent</code> (cleaning execution)</li></ul> <p>EXECUTION STEPS (STRICT ORDER):</p> <ol style="list-style-type: none"><li>1. Call <code>data_understanding_agent</code><ul style="list-style-type: none"><li>- Establish the authoritative schema and data types.</li></ul></li></ol> |



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                         | <p>2. Call data_assessing_agent</p> <ul style="list-style-type: none"> <li>- Perform PRE-CLEANING data quality assessment.</li> <li>- Produce cleaning recommendations only.</li> </ul> <p>3. Call data_cleaning_agent</p> <ul style="list-style-type: none"> <li>- Execute cleaning recommendations deterministically.</li> <li>- Operate ONLY on a copied dataset.</li> </ul> <p>4. Call data_assessing_agent AGAIN</p> <ul style="list-style-type: none"> <li>- Perform POST-CLEANING validation on the cleaned dataset.</li> <li>- Confirm remaining issues, if any.</li> </ul> <p>5. IF NEEDED, REPEAT FROM STEP 2</p> <p>OUTPUT:</p> <p>Produce a FINAL preprocessing report with EXACTLY these sections:</p> <p>## Data Understanding</p> <p>[Unmodified output from data_understanding_agent]</p> <p>## Pre-Cleaning Data Quality Assessment</p> <p>[Unmodified output from the first data_assessing_agent run]</p> <p>## Cleaning Execution Log</p> <p>[Unmodified output from data_cleaning_agent]</p> <p>## Post-Cleaning Data Quality Assessment</p> <p>[Unmodified output from the second data_assessing_agent run]</p> <p>STRICT CONSTRAINTS:</p> <ul style="list-style-type: none"> <li>- Do NOT perform analysis or interpretation yourself</li> <li>- Do NOT skip or reorder agents</li> <li>- Do NOT modify, summarize, or reinterpret sub-agent outputs</li> <li>- Do NOT introduce new findings or recommendations</li> <li>- Output ONLY the consolidated Markdown report</li> </ul> |
| <i>Data Preparation</i> | <p>You are a Data Preparation Agent. You use SQL via run_sql_tool to prepare data.</p> <p>BUSINESS CONTEXT:</p> <p>{business_questions}</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

**OBJECTIVE:**

Prepare question-specific datasets for visualization and analysis using SQL.

All preparation should be focused on answering the business questions above.

**MANDATORY RULES:**

1. Call `get_data_state_list()`.
2. Identify 'cleaned\_data'.
3. Inspect schema: `run_sql_tool(sql="DESCRIBE cleaned_data")`
4. For EACH business question, MAP it to available columns:
  - If the question references a column that doesn't exist, identify the CLOSEST PROXY column
  - If no reasonable proxy exists, SKIP it and document why
  - Verify grouping columns have more than 1 unique value:  
`run_sql_tool(sql="SELECT COUNT(DISTINCT col) FROM cleaned_data")`

5. For EACH answerable question, create a prepared dataset using `run_sql_tool` with `save_as`:

Example:

```
run_sql_tool(
 sql="SELECT category, AVG(price) as avg_price, COUNT(*) as
count FROM cleaned_data GROUP BY category ORDER BY
avg_price DESC",
 save_as="prep_q1_category_prices"
)
```

**COMMON PREPARATION PATTERNS:**

- Aggregation: `SELECT group_col, AGG(value_col) FROM cleaned_data GROUP BY group_col`
- Pivot: `PIVOT cleaned_data ON category USING SUM(sales)`
- Derived column: `SELECT *, date_part('month', date_col) AS month FROM cleaned_data`
- Filtering: `SELECT * FROM cleaned_data WHERE condition`
- Top N: `SELECT * FROM cleaned_data ORDER BY col DESC LIMIT N`



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                          | <p><b>CRITICAL FAILURE CONDITIONS:</b></p> <ul style="list-style-type: none"> <li>- Guessing column names</li> <li>- Reusing the same prep key for multiple questions</li> <li>- Describing transformations without executing SQL</li> <li>- Giving up on a question without checking for proxy columns first</li> </ul> <p><b>OUTPUT:</b></p> <p>Bullet list of:</p> <ul style="list-style-type: none"> <li>- Dataset key (the save_as name)</li> <li>- One-line description (include any proxy column mapping used)</li> <li>- If a question was skipped: reason why no proxy exists</li> </ul> <p>Output ONLY this list.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <i>Insights Gatherer</i> | <p>You are an Insights Gatherer Agent. You use SQL (run_sql_tool) for analysis and plot_tool for visualizations.</p> <p><b>BUSINESS CONTEXT:</b></p> <p>{business_questions}</p> <p><b>OBJECTIVE:</b></p> <p>Extract quantitative insights supported by plots.<br/>Ensure all insights directly answer the business questions above.</p> <p><b>INPUT:</b></p> <ul style="list-style-type: none"> <li>- Prepared Data Keys (from data_preparation_agent)</li> </ul> <p><b>EXECUTION STEPS:</b></p> <ol style="list-style-type: none"> <li>1. Call get_data_state_list().</li> <li>2. For EACH business question:           <ol style="list-style-type: none"> <li>a. Select the correct prepared dataset</li> <li>b. Generate statistics using run_sql_tool:               <ul style="list-style-type: none"> <li>- run_sql_tool(sql="SELECT col, COUNT(*), AVG(val) FROM prep_q1 GROUP BY col ORDER BY 2 DESC LIMIT 10")</li> <li>- run_sql_tool(sql="SELECT MIN(col), MAX(col), AVG(col), STDDEV(col) FROM prep_q1")</li> </ul> </li> <li>c. Create at least ONE visualization using plot_tool:               <ul style="list-style-type: none"> <li>- plot_tool(sql="SELECT category, SUM(revenue) as total FROM prep_q1 GROUP BY category ORDER BY total DESC</li> </ul> </li> </ol> </li> </ol> |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|     | <p>LIMIT 15", plot_type="bar", x="category", y="total", title="Revenue by Category")</p> <ul style="list-style-type: none"><li>- plot_tool(sql="SELECT date_col, SUM(value) as daily_total FROM prep_q1 GROUP BY date_col ORDER BY date_col", plot_type="line", x="date_col", y="daily_total", title="Daily Trend")</li><li>- plot_tool(sql="SELECT x_col, y_col FROM prep_q1 USING SAMPLE 500", plot_type="scatter", x="x_col", y="y_col", title="Correlation")</li><li>- plot_tool(sql="SELECT col FROM prep_q1", plot_type="histogram", x="col", title="Distribution of col")</li></ul> <p>3. Record visualization filenames EXACTLY as returned by plot_tool.</p> <p>ADAPTATION RULES:</p> <ul style="list-style-type: none"><li>- If a prepared dataset has no specific transformations, analyze its available numeric columns</li><li>- If a question is partially answerable, provide what you CAN answer and note what's missing</li><li>- You MUST ALWAYS produce at least one visualization per question</li><li>- NEVER output 'cannot be computed' — instead, reframe the question using available data</li></ul> <p>OUTPUT FORMAT (REQUIRED FOR EACH INSIGHT):</p> <ul style="list-style-type: none"><li>- Insight:</li><li>- Evidence:</li><li>- [VISUALIZATION]: filename.png</li></ul> <p>STRICT CONSTRAINTS:</p> <ul style="list-style-type: none"><li>- No business storytelling</li><li>- No recommendations</li><li>- No fabricated filenames</li><li>- No insights without numeric evidence</li></ul> |
| EDA | <p>You are an Exploratory Data Analysis Coordinator.</p> <p>BUSINESS CONTEXT:</p> <p>{business_questions}</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                       | <p><b>OBJECTIVE:</b></p> <p>Run EDA for EACH business question and consolidate results.</p> <p><b>AVAILABLE AGENTS:</b></p> <ul style="list-style-type: none"> <li>- prep_insight_agent</li> </ul> <p><b>EXECUTION RULES:</b></p> <ol style="list-style-type: none"> <li>1. Iterate over questions ONE AT A TIME.</li> <li>2. Call prep_insight_agent for each question.</li> <li>3. Preserve ALL visualization references exactly.</li> </ol> <p><b>OUTPUT SCHEMA RULES (follow exactly):</b></p> <ul style="list-style-type: none"> <li>- Top-level: <code>{{"insights": [...]}}</code> — a list of question-grouped objects.</li> <li>- Each object: <code>{{"question": str, "insights": [insight objects]}}</code>.</li> <li>- Each insight: <code>{{"insight": str, "evidence": str, "visualizations": [str, ...]}}</code>.</li> <li>- "visualizations" MUST be a list of plain filename strings e.g. "chart.png".</li> <li>- Do NOT nest objects inside "visualizations" — strings only.</li> </ul> <p><b>STRICT CONSTRAINTS:</b></p> <ul style="list-style-type: none"> <li>- Do NOT drop any visualization</li> <li>- Do NOT rename files</li> <li>- Do NOT reinterpret insights</li> </ul> |
| <i>Data Explainer</i> | <p>You are a Data Storytelling &amp; Strategy Agent. Do NOT add any extra text outside the report (No Opening or Closing text).</p> <p><b>BUSINESS CONTEXT:</b></p> <p><code>{business_questions}</code></p> <p><b>OBJECTIVE:</b></p> <p>Transform EDA outputs into a compelling executive-ready report. Ensure the final report addresses all the business questions above.</p> <p><b>INPUT:</b></p> <ul style="list-style-type: none"> <li>- EDA Report</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

**MANDATORY RULES:**

1. Call `list_output_files_tool()`.
2. Embed **ONLY** images that exist in the output directory.
3. Every embedded image **MUST** appear in the EDA report.
4. Don't add any extra text outside the report.

**REPORT STRUCTURE (Strictly Follow This Format):**

# [Title: Catchy & Professional, Mention the dataset about. e.g. "Global Weather Data Analysis Report"]

## Executive Summary

[The "TL;DR". Start with "The Big Picture" or "Executive Summary".

Synthesize the most critical findings. Use bolding for key metrics. Explain the "So What?" immediately.]

---

## The Deep Dive: [Number] Critical Business Questions

### [Question 1 Header]

**\*\*The Question\*\***: [Restate the business question]

#### [Insight 1 Header]

**\*\*The Data\*\***: [Briefly describe what data/metrics were used]

[Embed Image Here: `![Description](exact\_filename.png)`]

Use the EXACT filenames from the EDA report and `list_output_files_tool` output. Do NOT invent filenames or use a q\* prefix convention.

**\*\*What to Look For\*\***: [Guide the reader's eye on the chart]

**\*\*The Insight\*\***: [The core finding. Use bolding. Explain WHY this is happening.]



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

**\*\*Strategic Implication\*\***: [What should the business DO about this?]

[Repeat for all insights]

---

[Repeat for all questions]

**## Strategic Recommendations**

[Bulleted list of actionable steps based on the insights]

### REPORT REQUIREMENTS:

- Follow the provided structure EXACTLY
- Visuals first, narrative second
- Every insight must answer: "So what?"
- Address all business questions provided

### STRICT CONSTRAINTS:

- No missing images
- No invented evidence
- No raw data dumps
- Output ONLY the final Markdown report

**POLITEKNIK  
NEGERI  
JAKARTA**



## Lampiran 2 Output *Data Understanding Agent*

TOOL data\_understanding\_agent END

RawFormatted×

Timestamp: 16/05/2026, 15.59.28ID: b4c1b131-bf40-4fe7-9ac8-41fe0f68b8d2

### Dataset Overview

- Number of rows: 100,000
- Number of columns: 9
- Table-level description based ONLY on tool output: The dataset contains transaction records with no missing values across all columns. Each row represents a transaction with details including transaction ID, date, product name, category, store location, payment method, quantity, unit price, and total bill.

### Column Descriptions

- Column name: transaction\_id
  - Data type: BIGINT
  - Objective description: Unique identifier for each transaction, ranging from 1 to 100,000.
- Column name: date
  - Data type: VARCHAR
  - Objective description: Date of the transaction in YYYY-MM-DD format, ranging from 2024-01-01 to 2024-12-30.
- Column name: product\_name
  - Data type: VARCHAR
  - Objective description: Name of the product sold in the transaction, with 46 unique values.
- Column name: category
  - Data type: VARCHAR
  - Objective description: Category of the product, with 5 unique values (e.g., Bahan Pokok, Perabot, Otomotif, Pakaian, Elektronik).
- Column name: store\_location
  - Data type: VARCHAR
  - Objective description: Location of the store where the transaction occurred, with 9 unique cities (e.g., Bandung, Jakarta, Medan, Surabaya, Yogyakarta).

Close

- Hak Cipta :**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
  2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Lampiran 3 Output *Data Assessing Agent (Pre-Cleaning)*

TOOL

data\_assessing\_agent

END

Raw

Formatted

×

Timestamp: 16/05/2026, 15.59.43

ID: 41d55bea-08e6-4142-9e37-63dc10e86962

### Data Quality Assessment

- Missing values in `"quantity"` or `"total_bill"`:** None reported, but business questions require validating that quantity is strictly  $> 0$  for valid transactions (only mentioned in business context, not in data profile).
- Inconsistent category values:** Data profile shows only 5 unique categories, but a `DISTINCT` query reveals 6 categories (including 'Bahan Pokok', 'Kecantikan'). This discrepancy suggests potential data entry errors or untracked categories.
- Payment method discrepancy:** Business question references 'Gopay', but the `DISTINCT` query shows 'Gopay' is present (14,171 records). However, the business context also mentions 'Gopay' specifically, which aligns with data.
- Date format:** Dates are stored as `VARCHAR`, not `DATE`, which could lead to incorrect day-of-week calculations if not properly cast.
- No duplicates or nulls:** The assessment report shows `null_ratio = 0` and `duplicate_ratio = 0`, indicating clean foundational data.
- Outliers:** `"unit_price"` and `"total_bill"` have high standard deviations, but min/max values seem plausible (no negative values).

### Cleaning Recommendations

- Standardize category values:** Ensure all categories are correctly categorized by validating against a defined list. For example, confirm if 'Bahan Pokok' and 'Kecantikan' are valid categories or if they should be merged into existing ones (e.g., 'Bahan Pokok' into 'Perabot' or a new 'Groceries' category). Use SQL to review and map categories if needed.
- Validate transaction validity:** Add a derived column or filter for `"quantity > 0"` to ensure only valid transactions are considered in analyses, as per business context. This is already implied but should be explicitly enforced in queries.
- Convert date column to DATE type:** Transform `"date"` from `VARCHAR` to `DATE` using `"CAST(date AS DATE)"` to enable accurate day-of-week calculations and time-based aggregations.
- Ensure payment method consistency:** Verify that 'Gopay' is correctly spelled and consistently used (e.g., no 'GOPAY' or 'GoPay' variants). Use `"DISTINCT"` to check for variations.

### Data Limitations

- The dataset does not include a `"day_of_week"` column, but this can be derived from the `"date"` column after conversion to `DATE` type.
- No explicit column for transaction validity (e.g., `"is_valid"`), but this can be inferred from `"quantity > 0"`.
- The business question about weekend revenue requires deriving day of week from `"date"`, which is feasible after conversion.

Close



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

- Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
- Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## Lampiran 4 Output *Data Cleaning Agent*

TOOL

data\_cleaning\_agent

END

Raw

Formatted

×

Timestamp: 16/05/2026, 15.59.54ID: ebba9ad9-0741-4bdc-bf46-0fce2ca72d10

All payment methods look consistent. No variations like 'GOPAY' or 'GoPay'. So no cleaning needed.  
I'll output the execution log as required.

Data Cleaning Execution Log

| Step | SQL Query                                                                                                                                                                                 | Columns Affected               |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|
| 1    | <code>"SELECT * FROM raw_data"</code>                                                                                                                                                     | All columns                    |
| 2    | <code>"SELECT * EXCLUDE (category), CASE WHEN category = 'Bahan Pokok' THEN 'Perabot' WHEN category = 'Kecantikan' THEN 'Pakaian' ELSE category END AS category FROM cleaned_data"</code> | <code>"category"</code>        |
| 3    | <code>"SELECT * FROM cleaned_data WHERE quantity &gt; 0"</code>                                                                                                                           | Row filter (no column change)  |
| 4    | <code>"SELECT * EXCLUDE (date), CAST(date AS DATE) AS date FROM cleaned_data"</code>                                                                                                      | <code>"date"</code>            |
| 5    | <code>"SELECT DISTINCT payment_method FROM cleaned_data"</code>                                                                                                                           | No changes (verification only) |

Close

## Lampiran 5 Output *Data Assessing Agent (Post-Cleaning)*

TOOL

data\_assessing\_agent

END

Raw

Formatted

×

Timestamp: 16/05/2026, 16.00.02ID: e05a54de-d61b-41ff-99cb-bdee2b42286f

Data Quality Assessment

- Data quality score: 1.0 (no nulls, no duplicates)
- Outliers in pricing: Unit price ranges from 4,700 to 19,424,400; total bill ranges from 4,700 to 97,118,000. This may be acceptable but worth verifying for business logic.
- Quantity distribution: Only 5 unique values (1-5), mean ≈3.0. This seems reasonable for retail.
- Date range: 2024-01-01 to 2024-12-30 (404 unique dates). This covers the full year, which is good for weekly analysis.

Prioritized issues affecting business questions:

- No critical issues found that would prevent answering the business questions.

Cleaning Recommendations

No further cleaning required.

Data Limitations

None identified; all required columns for the business questions appear present and clean.

Note: This is a post-cleaning validation, and the dataset passes all quality checks. The data appears ready for business analysis.

Close

Lampiran 6 *Data State* Pada Sistem

```

"data_state": {
 "raw_data": {
 "description": "The initial data uploaded by the user.",
 "path": "outputs/20260516_085919_indo_100000_clean_sales_data/raw_data.parquet"
 },
 "cleaned_data": {
 "path": "outputs/20260516_085919_indo_100000_clean_sales_data/cleaned_data.parquet",
 "description": "Derived dataset: cleaned_data"
 },
 "prep_q1_electronics_revenue_by_store": {
 "path": "outputs/20260516_085919_indo_100000_clean_sales_data/prep_q1_electronics_revenue_by_store.parquet",
 "description": "Derived dataset: prep_q1_electronics_revenue_by_store"
 },
 "prep_q2_weekend_revenue_daily": {
 "path": "outputs/20260516_085919_indo_100000_clean_sales_data/prep_q2_weekend_revenue_daily.parquet",
 "description": "Derived dataset: prep_q2_weekend_revenue_daily"
 },
 "prep_q2_weekend_revenue_total": {
 "path": "outputs/20260516_085919_indo_100000_clean_sales_data/prep_q2_weekend_revenue_total.parquet",
 "description": "Derived dataset: prep_q2_weekend_revenue_total"
 },
 "prep_q3_payment_aov": {
 "path": "outputs/20260516_085919_indo_100000_clean_sales_data/prep_q3_payment_aov.parquet",
 "description": "Derived dataset: prep_q3_payment_aov"
 },
 "prep_q4_top5_percentage": {
 "path": "outputs/20260516_085919_indo_100000_clean_sales_data/prep_q4_top5_percentage.parquet",
 "description": "Derived dataset: prep_q4_top5_percentage"
 },
 "prep_q5_non_electronics_gopay_top_product": {
 "path": "outputs/20260516_085919_indo_100000_clean_sales_data/prep_q5_non_electronics_gopay_top_product.parquet",
 "description": "Derived dataset: prep_q5_non_electronics_gopay_top_product"
 }
}

```

Lampiran 7 *Output Prep Insight Agent*

TOOL

prep\_insight\_agent

END

Raw

Formatted

X

Timestamp: 16/05/2026, 16.02.02

ID: 6cb48cd2-4c9d-4908-b09b-2aa562b5698b

**Insight:** Gopay has the highest average order value (AOV) among valid transactions (quantity > 0), at approximately **4,437,716.51 IDR**.

**Evidence:** The `"prep_q3_payment_aov"` dataset shows the following average order values for each payment method:

- Gopay: 4,437,716.512596147 IDR (highest)
- ShopeePay: 4,398,853.857780831 IDR
- Kartu Kredit: 4,377,273.756227633 IDR
- Tunai: 4,350,262.334402454 IDR
- QRIS: 4,313,207.23642732 IDR
- Transfer Bank: 4,260,884.379397692 IDR
- OVO: 4,231,902.205157109 IDR

The data confirms that all transactions with quantity > 0 were considered, and Gopay leads in average order value.

[VISUALIZATION]: plot\_bar\_payment\_method\_avg\_order\_value.png

[VISUALIZATION]: plot\_bar\_payment\_method\_num\_transactions.png

Close



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

Lampiran 8 Tabel Skenario *Black Box Testing*

| Kode   | Kasus          | Skenario                                                                                          | Harapan                                                                   |
|--------|----------------|---------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| BB-1-1 | <i>Sign Up</i> | Pengguna berhasil melakukan pendaftaran akun dengan data yang valid.                              | Sistem menavigasi pengguna ke halaman <i>Sign In</i> .                    |
| BB-1-2 |                | Pengguna gagal melakukan pendaftaran akun karena data tidak lengkap.                              | Sistem menampilkan informasi pada data yang kosong.                       |
| BB-1-3 |                | Pengguna gagal melakukan pendaftaran akun karena <i>email</i> sudah terdaftar.                    | Sistem menampilkan informasi bahwa <i>email</i> terkait sudah digunakan.  |
| BB-1-4 |                | Pengguna gagal melakukan pendaftaran akun karena format <i>email</i> salah.                       | Sistem menampilkan informasi pada input <i>email</i> .                    |
| BB-2-1 | <i>Sign In</i> | Pengguna berhasil melakukan <i>Sign In</i> akun dengan kredensial yang benar dan terdaftar.       | Sistem menavigasi pengguna ke halaman <i>Home</i> pada <i>dashboard</i> . |
| BB-2-2 |                | Pengguna gagal melakukan <i>Sign In</i> akun karena <i>email</i> yang dimasukkan tidak ditemukan. | Sistem menampilkan informasi terkait kredensial yang invalid.             |
| BB-2-3 |                | Pengguna gagal melakukan <i>Sign In</i> akun karena <i>password</i> yang dimasukkan salah.        | Sistem menampilkan informasi terkait kredensial yang invalid.             |
| BB-2-4 |                | Pengguna gagal melakukan <i>Sign In</i> akun karena data yang dimasukkan kosong/tidak lengkap.    | Sistem menampilkan informasi pada data yang kosong.                       |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |                                                  |                                                                                                    |                                                                                           |
|--------|--------------------------------------------------|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| BB-2-5 |                                                  | Pengguna gagal melakukan <i>Sign In</i> akun karena format <i>email</i> salah.                     | Sistem menampilkan informasi pada input <i>email</i> .                                    |
| BB-3-1 |                                                  | Pengguna berhasil membuat proyek analisis baru dengan memasukkan seluruh data yang dibutuhkan.     | Sistem menavigasi pengguna ke halaman <i>Processing</i> .                                 |
| BB-3-2 | Membuat Proyek Analisis                          | Pengguna gagal membuat proyek analisis baru karena file yang diupload tidak disupport oleh sistem. | Sistem menampilkan informasi bahwa tipe file yang diupload tidak disupport oleh sistem.   |
| BB-3-3 |                                                  | Pengguna gagal membuat proyek analisis baru karena data yang dibutuhkan kosong/tidak lengkap.      | Sistem menampilkan informasi pada data yang kosong.                                       |
| BB-4-1 | Membatalkan Proses Analisis                      | Pengguna menekan tombol " <i>Cancel Analysis</i> ".                                                | Proses yang sedang berjalan batal dan sistem menavigasi pengguna ke halaman <i>Home</i> . |
| BB-5-1 | Mengubah Nama Proyek Analisis                    | Pengguna mengubah nama proyek analisis dengan teks tertentu.                                       | Sistem mengubah nama proyek analisis dengan teks yang pengguna inputkan.                  |
| BB-5-2 |                                                  | Pengguna mengubah nama proyek analisis dengan teks kosong.                                         | Sistem melakukan <i>disable</i> pada tombol " <i>Save Changes</i> ".                      |
| BB-6-1 | Menghapus Proyek Analisis                        | Pengguna menghapus proyek analisis yang diinginkan.                                                | Sistem menghapus data proyek analisis yang dimaksud pengguna.                             |
| BB-7-1 | Memilih Menu di Daftar Isi pada Laporan Analisis | Pengguna memilih salah satu bagian di daftar isi.                                                  | Sistem menavigasi pengguna ke bagian yang dimaksud.                                       |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|         |                                    |                                                                                            |                                                                      |
|---------|------------------------------------|--------------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| BB-8-1  | Melakukan Analisis Ulang           | Pengguna menekan tombol “ <i>Regenerate Report</i> ” pada laporan analisis tertentu.       | Sistem menavigasi pengguna ke halaman <i>Processing</i> .            |
| BB-9-1  | Mengubah Mode Tampilan List Proyek | Pengguna menekan <i>switch</i> untuk mengubah mode tampilan pada halaman <i>Projects</i> . | Sistem mengubah mode tampilan ke mode yang dipilih pengguna.         |
| BB-10-1 | Mencari Proyek Analisis            | Pengguna memasukkan <i>query</i> pencarian berdasarkan nama proyek yang sudah dibuat.      | Sistem menampilkan list proyek analisis dengan nama proyek terkait.  |
| BB-10-2 |                                    | Pengguna memasukkan <i>query</i> pencarian berdasarkan nama dataset yang diupload.         | Sistem menampilkan list proyek analisis dengan nama dataset terkait. |
| BB-10-3 |                                    | Pengguna memasukkan <i>query</i> pencarian berdasarkan nama proyek yang belum dibuat.      | Sistem menampilkan list proyek analisis kosong.                      |
| BB-10-4 |                                    | Pengguna memasukkan <i>query</i> pencarian berdasarkan nama dataset yang belum diupload.   | Sistem menampilkan list proyek analisis kosong.                      |

Lampiran 9 Validasi Dataset *Dummy* (Sintetis)



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



KEMENTERIAN PENDIDIKAN TINGGI, RISET, DAN TEKNOLOGI  
POLITEKNIK NEGERI JAKARTA  
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER  
Universitas Indonesia, Jl. Prof. DR. G.A. Siwabessy, Depok 16425  
Telp: (021)91274097, Fax : ( 021 ) 7863531, ( 021 )7270036 Hunting  
Laman : <http://www.pnj.ac.id>, e-mail : [tik.pnj@gmail.com](mailto:tik.pnj@gmail.com)

L3

### BERITA ACARA

Pada hari ini, Selasa, Tanggal 23 Juni 2026, telah dilaksanakan sesi validasi *Logic Script Dummy Dataset Generator* dan *Ground Truth* dataset penelitian bersama:

Nama : Ilhaam Akmal Abdjul

Bidang Keahlian : Lead AI Software Engineer

Sesi ini dilakukan dalam rangka uji kelayakan skrip logika pembuatan dataset *dummy* beserta kalkulasi *ground truth*-nya untuk digunakan dalam pengujian Sistem Data Analisis Otomatis Berbasis *Agentic AI*.

Adapun kegiatan yang dilakukan selama sesi berlangsung dapat dilihat pada poin-poin berikut.

1. Peneliti menjelaskan secara umum arsitektur pengujian sistem dan alur pembentukan data *dummy* sebagai upaya penyamaan konteks dan pemahaman antara peneliti dan ahli/praktisi.
2. Ahli melakukan pengujian terhadap *Logic Script Data Generation*, *Noise Injection Logic*, serta keakuratan *Ground Truth Calculation* dari dataset yang dibuat untuk menjawab kueri/pertanyaan analitik.
3. Ahli memberikan penilaian berdasarkan kategori '*Correct*' (Benar) atau '*Incorrect*' (Salah) terhadap setiap parameter pengujian yang tercantum pada lembar lampiran ke-2.

Berdasarkan hasil validasi data sampel (terlampir), *logic script* dan *ground truth* yang dievaluasi oleh ahli dinyatakan (Layak / Tidak Layak) digunakan sebagai landasan (*benchmark*) pengujian Sistem Data Analisis Otomatis Berbasis *Agentic AI*.

Demikian Berita Acara ini dibuat dengan sebenarnya untuk dapat digunakan sebagaimana mestinya.

Jakarta, 23 Juni 2026

Pihak Pertama (Peneliti)

Muhammad Bintang Ramadhan  
NIM. 2207411055

Pihak Kedua (Ahli/Praktisi)

Ilhaam Akmal Abdjul



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



KEMENTERIAN PENDIDIKAN TINGGI, RISET, DAN TEKNOLOGI  
POLITEKNIK NEGERI JAKARTA  
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER  
Universitas Indonesia, Jl. Prof. DR. G.A. Siwabessy, Depok 16425  
Telp: (021)91274097, Fax: (021) 7863531, (021) 7270036 Hunting  
Laman : <http://www.pnj.ac.id>, e-mail : [tik.pnj@gmail.com](mailto:tik.pnj@gmail.com)

L3

### LEMBAR LAMPIRAN: PARAMETER PENGUJIAN DAN HASIL VALIDASI

#### Konteks Penelitian :

Untuk menguji performa dan akurasi Sistem Data Analisis Otomatis berbasis *Agentic AI*, peneliti membutuhkan *dataset dummy* beserta *ground truth* (kunci jawaban) dari berbagai pertanyaan analitik yang ditentukan. Masalahnya adalah keseluruhan data beserta jawaban *ground truth* tersebut dibuat secara otomatis melalui sebuah *script*. Agar dapat menjadi tolok ukur (*benchmark*) evaluasi yang sah, validitas skrip pembentuk data (termasuk *noise injection*) dan kalkulasi *ground truth* tersebut harus diuji oleh ahli atau praktisi. Dengan demikian, jika *script* teruji valid, dataset yang dihasilkan dapat digunakan untuk menilai apakah *Agentic AI* berhasil melakukan penalaran dan analisis data dengan benar.

#### Petunjuk Penilaian:

1. Ahli mengevaluasi kode dan logika yang digunakan pada *script generator*, lalu memberikan penilaian kelayakan berdasarkan parameter yang ada.
2. Skrip yang layak digunakan adalah skrip dengan logika pembentukan data, *noise injection*, dan kalkulasi *ground truth* yang akurat (*Correct*).
3. Skrip yang tidak layak adalah skrip yang menghasilkan *logical error* atau kesalahan hitung pada *ground truth* (*Incorrect*).

#### Definisi Parameter:

1. *Data Generation Logic*: Seberapa tepat logika dan distribusi data sintetik yang dibuat untuk menghasilkan dataset penjualan.
2. *Noise Injection Logic*: Ketepatan mekanisme *noise injection* pada dataset sintetik.
3. *Ground Truth Calculation*: Keakuratan kalkulasi matematis/logis pada kunci jawaban analitik.



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta



KEMENTERIAN PENDIDIKAN TINGGI, RISET, DAN TEKNOLOGI  
POLITEKNIK NEGERI JAKARTA  
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER  
Universitas Indonesia, Jl. Prof. DR. G.A. Siwabessy, Depok 16425  
Telp: (021)91274097, Fax : ( 021 ) 7863531, ( 021 )7270036 Hunting  
Laman : <http://www.pnj.ac.id>, e-mail : [dk.pnj@gmail.com](mailto:dk.pnj@gmail.com)

**L3**

| Kode / Parameter Uji                                                                                                   | Penilaian Ahli ( <i>Correct/Incorrect</i> ) |
|------------------------------------------------------------------------------------------------------------------------|---------------------------------------------|
| <i>Data Generation</i>                                                                                                 | Correct / <del>Incorrect</del>              |
| <i>Noise Injection</i>                                                                                                 | Correct / <del>Incorrect</del>              |
| <i>Ground Truth Calculation: q1_location (Top 'Elektronik' Location)</i>                                               | Correct / <del>Incorrect</del>              |
| <i>Ground Truth Calculation: q1_revenue (Top 'Elektronik' Location Revenue Amount)</i>                                 | Correct / <del>Incorrect</del>              |
| <i>Ground Truth Calculation: q2_weekend_revenue (Total Weekend Revenue)</i>                                            | Correct / <del>Incorrect</del>              |
| <i>Ground Truth Calculation: q3_payment_method (Highest AOV Payment Method)</i>                                        | Correct / <del>Incorrect</del>              |
| <i>Ground Truth Calculation: q4_percentage (Percentage of Revenue from Top 5 Products)</i>                             | Correct / <del>Incorrect</del>              |
| <i>Ground Truth Calculation: q5_top_product (Product with Highest Quantity Sold via Gopay, excluding 'Elektronik')</i> | Correct / <del>Incorrect</del>              |

Lampiran 10 Tabel Data *Ground Truth*

| Kode Dataset | <i>Ground Truth</i> |
|--------------|---------------------|
| DD-1-1       | {                   |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |                                                                                                                                                                                                                                                                                                                                 |
|--------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|        | <pre>"Q1_Top_Elektronik_Location": "Surabaya", "Q1_Top_Elektronik_Revenue_Amount": 46628363000.0, "Q2_Total_Weekend_Revenue": 126192538400.0, "Q3_Highest_AOV_Payment_Method": "Gopay", "Q4_Top_5_Revenue_Percentage": 70.41233441180869, "Q5_Top_Product_Gopay_Non_Elektronik": "Hijab Bella Square" }</pre>                   |
| DD-1-2 | <pre>{ "Q1_Top_Elektronik_Location": "Surabaya", "Q1_Top_Elektronik_Revenue_Amount": 46597984500.0, "Q2_Total_Weekend_Revenue": 124513030800.0, "Q3_Highest_AOV_Payment_Method": "Gopay", "Q4_Top_5_Revenue_Percentage": 70.1468987443761, "Q5_Top_Product_Gopay_Non_Elektronik": "Beras Pandan Wangi 5kg" }</pre>              |
| DD-2-1 | <pre>{ "Q1_Top_Elektronik_Location": "Yogyakarta", "Q1_Top_Elektronik_Revenue_Amount": 226601882200.0, "Q2_Total_Weekend_Revenue": 615750801300.0, "Q3_Highest_AOV_Payment_Method": "Transfer Bank", "Q4_Top_5_Revenue_Percentage": 70.09106924324709, "Q5_Top_Product_Gopay_Non_Elektronik": "Indomie Goreng (Karton)" }</pre> |
| DD-2-2 | <pre>{ "Q1_Top_Elektronik_Location": "Semarang", "Q1_Top_Elektronik_Revenue_Amount": 226785976800.0, "Q2_Total_Weekend_Revenue": 610395006600.0, "Q3_Highest_AOV_Payment_Method": "Gopay", "Q4_Top_5_Revenue_Percentage": 70.00213540268028, "Q5_Top_Product_Gopay_Non_Elektronik": "Teh Celup Sariwangi" }</pre>               |
| DD-3-1 | <pre>{ "Q1_Top_Elektronik_Location": "Makassar", "Q1_Top_Elektronik_Revenue_Amount": 449412455200.0, "Q2_Total_Weekend_Revenue": 1235266450400.0, "Q3_Highest_AOV_Payment_Method": "Transfer Bank", "Q4_Top_5_Revenue_Percentage": 69.97951469452518, "Q5_Top_Product_Gopay_Non_Elektronik": "Kursi Kerja Ergonomis" }</pre>    |



© Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |                                                                                                                                                                                                                                                                                                                                        |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|        | }                                                                                                                                                                                                                                                                                                                                      |
| DD-3-2 | {<br>"Q1_Top_Elektronik_Location": "Medan",<br>"Q1_Top_Elektronik_Revenue_Amount": 444965182200.0,<br>"Q2_Total_Weekend_Revenue": 1224729736900.0,<br>"Q3_Highest_AOV_Payment_Method": "Kartu Kredit",<br>"Q4_Top_5_Revenue_Percentage": 69.70699810255103,<br>"Q5_Top_Product_Gopay_Non_Elektronik": "Kaos Polos Cotton Combed"<br>}  |
| DD-4-1 | {<br>"Q1_Top_Elektronik_Location": "Bandung",<br>"Q1_Top_Elektronik_Revenue_Amount": 2230692939400.0,<br>"Q2_Total_Weekend_Revenue": 6111960180900.0,<br>"Q3_Highest_AOV_Payment_Method": "Transfer Bank",<br>"Q4_Top_5_Revenue_Percentage": 69.87303949690904,<br>"Q5_Top_Product_Gopay_Non_Elektronik": "Kursi Kerja Ergonomis"<br>} |
| DD-4-2 | {<br>"Q1_Top_Elektronik_Location": "Makassar",<br>"Q1_Top_Elektronik_Revenue_Amount": 2218314011300.0,<br>"Q2_Total_Weekend_Revenue": 6144773859800.0,<br>"Q3_Highest_AOV_Payment_Method": "Gopay",<br>"Q4_Top_5_Revenue_Percentage": 69.89691325130502,<br>"Q5_Top_Product_Gopay_Non_Elektronik": "Kanebo Lap Serbaguna"<br>}         |
| DD-5-1 | {<br>"Q1_Top_Elektronik_Location": "Surabaya",<br>"Q1_Top_Elektronik_Revenue_Amount": 4450650614800.0,<br>"Q2_Total_Weekend_Revenue": 12305686340400.0,<br>"Q3_Highest_AOV_Payment_Method": "QRIS",<br>"Q4_Top_5_Revenue_Percentage": 69.88675325618152,<br>"Q5_Top_Product_Gopay_Non_Elektronik": "Kursi Kerja Ergonomis"<br>}        |
| DD-5-2 | {<br>"Q1_Top_Elektronik_Location": "Yogyakarta",<br>"Q1_Top_Elektronik_Revenue_Amount": 4441824700000.0,<br>"Q2_Total_Weekend_Revenue": 12254600929100.0,<br>"Q3_Highest_AOV_Payment_Method": "Gopay",<br>}                                                                                                                            |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|  |                                                                                                                               |
|--|-------------------------------------------------------------------------------------------------------------------------------|
|  | <pre>"Q4_Top_5_Revenue_Percentage": 69.87037461902382, "Q5_Top_Product_Gopay_Non_Elektronik": "Gula Pasir Gulaku 1kg" }</pre> |
|--|-------------------------------------------------------------------------------------------------------------------------------|

Lampiran 11 Tabel Data Hasil *Black Box*

| Kode   | Harapan                                                                   | Pengamatan                                                                | Hasil  |
|--------|---------------------------------------------------------------------------|---------------------------------------------------------------------------|--------|
| BB-1-1 | Sistem menavigasi pengguna ke halaman <i>Sign In</i> .                    | Sistem menavigasi pengguna ke halaman <i>Sign In</i> .                    | Sesuai |
| BB-1-2 | Sistem menampilkan informasi pada data yang kosong.                       | Sistem menampilkan informasi pada data yang kosong.                       | Sesuai |
| BB-1-3 | Sistem menampilkan informasi bahwa <i>email</i> terkait sudah digunakan.  | Sistem menampilkan informasi bahwa <i>email</i> terkait sudah digunakan.  | Sesuai |
| BB-1-4 | Sistem menampilkan informasi pada input <i>email</i> .                    | Sistem menampilkan informasi pada input <i>email</i> .                    | Sesuai |
| BB-2-1 | Sistem menavigasi pengguna ke halaman <i>Home</i> pada <i>dashboard</i> . | Sistem menavigasi pengguna ke halaman <i>Home</i> pada <i>dashboard</i> . | Sesuai |
| BB-2-2 | Sistem menampilkan informasi terkait kredensial yang invalid.             | Sistem menampilkan informasi terkait kredensial yang invalid.             | Sesuai |
| BB-2-3 | Sistem menampilkan informasi terkait kredensial yang invalid.             | Sistem menampilkan informasi terkait kredensial yang invalid.             | Sesuai |
| BB-2-4 | Sistem menampilkan informasi pada data yang kosong.                       | Sistem menampilkan informasi pada data yang kosong.                       | Sesuai |
| BB-2-5 | Sistem menampilkan informasi pada input <i>email</i> .                    | Sistem menampilkan informasi pada input <i>email</i> .                    | Sesuai |



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|         |                                                                                           |                                                                                           |        |
|---------|-------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|--------|
| BB-3-1  | Sistem menavigasi pengguna ke halaman <i>Processing</i> .                                 | Sistem menavigasi pengguna ke halaman <i>Processing</i> .                                 | Sesuai |
| BB-3-2  | Sistem menampilkan informasi bahwa tipe file yang diupload tidak disupport oleh sistem.   | Sistem menampilkan informasi bahwa tipe file yang diupload tidak disupport oleh sistem.   | Sesuai |
| BB-3-3  | Sistem menampilkan informasi pada data yang kosong.                                       | Sistem menampilkan informasi pada data yang kosong.                                       | Sesuai |
| BB-4-1  | Proses yang sedang berjalan batal dan sistem menavigasi pengguna ke halaman <i>Home</i> . | Proses yang sedang berjalan batal dan sistem menavigasi pengguna ke halaman <i>Home</i> . | Sesuai |
| BB-5-1  | Sistem mengubah nama proyek analisis dengan teks yang pengguna inputkan.                  | Sistem mengubah nama proyek analisis dengan teks yang pengguna inputkan.                  | Sesuai |
| BB-5-2  | Sistem melakukan <i>disable</i> pada tombol “ <i>Save Changes</i> ”                       | Sistem melakukan <i>disable</i> pada tombol “ <i>Save Changes</i> ”                       | Sesuai |
| BB-6-1  | Sistem menghapus data proyek analisis yang dimaksud pengguna.                             | Sistem menghapus data proyek analisis yang dimaksud pengguna.                             | Sesuai |
| BB-7-1  | Sistem menavigasi pengguna ke bagian yang dimaksud.                                       | Sistem menavigasi pengguna ke bagian yang dimaksud.                                       | Sesuai |
| BB-8-1  | Sistem menavigasi pengguna ke halaman <i>Processing</i> .                                 | Sistem menavigasi pengguna ke halaman <i>Processing</i> .                                 | Sesuai |
| BB-9-1  | Sistem mengubah mode tampilan ke mode yang dipilih pengguna.                              | Sistem mengubah mode tampilan ke mode yang dipilih pengguna.                              | Sesuai |
| BB-10-1 | Sistem akan menampilkan list proyek analisis dengan nama proyek terkait.                  | Sistem menampilkan list proyek analisis dengan nama proyek terkait.                       | Sesuai |

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|         |                                                                           |                                                                      |        |
|---------|---------------------------------------------------------------------------|----------------------------------------------------------------------|--------|
| BB-10-2 | Sistem akan menampilkan list proyek analisis dengan nama dataset terkait. | Sistem menampilkan list proyek analisis dengan nama dataset terkait. | Sesuai |
| BB-10-3 | Sistem akan menampilkan list proyek analisis kosong.                      | Sistem menampilkan list proyek analisis kosong.                      | Sesuai |
| BB-10-4 | Sistem akan menampilkan list proyek analisis kosong.                      | Sistem menampilkan list proyek analisis kosong.                      | Sesuai |

## Lampiran 12 Hasil Skenario NB-1 Pada Literatur

| <i>Data Analyst</i> | <b>Hasil</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Senior              | <i>1. Combining the Support Rate, Oppose Rate, and Net Polarity Rate, we should accept Loic De Kergret. 2. Overall, all candidates have a higher oppose rate than the support rate. 3. For the support rate, Loic ranks second. Although Vincent Montmeant has a higher support rate than Loic, Vincent also has a much higher oppose rate. 4. We consider the New polarity rate to be the best metric to take. This metric is calculated as Support Rate substrates the Oppose rate. 5. We think it's important for the candidate to have a high support rate, but also important to have a low oppose rate. Hence the Net Polarity rate is the one we take.</i> |
| Junior              | <i>1. Based on data, candidate 4 has the highest average support score. 2. Newsweek Poll has the highest average support score amount three different poll sources. 3. Candidate 2 has extremely high unsure rate given by WNBC/Marist Poll. 4. Candidate 4 has the highest oppose rate. 5. The variance of consider rate is low among all the different rates.</i>                                                                                                                                                                                                                                                                                               |
| GPT-4               | <i>1. Vincent Montméat has the highest average support among the candidates, with a score of 0.33, making him a strong contender for acceptance. 2. Stéphane Antiga and Laurent Capet have the lowest average support, with scores of 0.17 and 0.18, respectively, indicating that they may not be the most suitable candidates for the position. 3. Guillaume Samica and Loïc De Kergret have similar average support scores of 0.24, placing them in the middle of the pack and suggesting that they could be potential alternatives if Vincent Montméat is not selected. 4. Hubert Henno has a slightly higher average</i>                                     |



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

support score than Guillaume Samica and Loïc De Kergret, with a score of 0.25, making him another potential alternative candidate. 5. Overall, the data suggests that Vincent Montméat is the most favored candidate, but Hubert Henno, Guillaume Samica, and Loïc De Kergret could also be considered if additional factors need to be taken into account.

Lampiran 13 Hasil Skenario NB-2 Pada Literatur

| Data Analyst | Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Senior       | 1. We choose to measure the student by their average grade. 2. The calculation method is to take the GPA for every student and take the average. This is to prevent the case where 1 student takes multiple courses. 3. The top 3 students ranked by grade are Robertson Gerald, Smithson Anne, and Katinga Raphael. 4. Robertson Gerald's score is the highest with 3.87, followed by Smithson Anne at 3.27, and Katinga Raphael at 3.15. 5. While Robertson is significantly better than the rest, Smithson and Katinga are not leading the following too much.                                                                                                                                                                                                                                                                    |
| Junior       | 1. Based on GPA, the top 3 students are 324291, 324257, 324274. 2. ACCT department has the most amount of students in dataset, which is 3. 3. Students date of birth are all before 1981. 4. The student hours has a high variance. 5. 3 out of 8 students are transfer students.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| GPT-4        | 1. The data provided consists of three students' names and their respective test performance scores, presumably on a scale of 0 to 4 or 5. 2. Among the three students, Gerald Robertson has the highest test performance score with a 3.87, indicating that he performs exceptionally well on tests. 3. Anne Smithson has the second-highest score with a 3.27, which suggests that she also performs well on tests, although not as high as Gerald Robertson. 4. Raphael Katinga has the lowest score among the three students with a 3.15, but this score still indicates a relatively good performance on tests. 5. Based on the data, all three students can be considered as representatives of those who perform well on tests, with Gerald Robertson being the top performer, followed by Anne Smithson and Raphael Katinga. |

Lampiran 14 Hasil Skenario NB-3 Pada Literatur

| Data Analyst | Hasil |
|--------------|-------|
|--------------|-------|



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

|        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|--------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Senior | <p>1. We plotted out employee count by the department and colored by region with blue for the American and orange for Europe 2. Europe has only 2 employees, with 1 from Human Resources and 1 from Public Relationship 3. We think it may not be very efficient to set up an EU office with only two employees 4. Usually, Human Resources employees work closely with the rest of the department. Hence we would suggest shifting the Human resource function to the US. For PR, if the marketing focus is EU, it's still ok to keep it in EU 5. However, we would advise the company to consider labour costs in different places as well</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Junior | <p>1. Based on the salary of employees across all departments, executive department has the highest salary, followed by accounting and public relations 2. The lowest paid departments are purchasing and administration and shipping 3. Among all the job titles, purchasing clerk, stock clerk and shipper clerk are lowest paid while managers and presidents are highest paid 4. Shipping and sales have the most amount of employees while accounting and administration have lowest amount of employees 5. Based on above, it is shown that it's not feasible to move people from accounting, admin departments to sales, shipping as there are very little people in those department. Instead, finance and IT are higher paid and having more people and could be a good target of restructuring</p>                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| GPT-4  | <p>1. The data shows a high concentration of employees in the Shipping department in the Americas region, which could indicate a potential area for cost reduction by redistributing employees to other regions or departments. 2. The Americas region has a diverse range of departments, with Finance and Purchasing having a significant number of employees. This could present an opportunity to shift some employees from these departments to other regions where there is a shortage of staff in similar roles. 3. Europe has a very limited number of departments represented, with only Human Resources and Public Relations having one employee each. This suggests that there may be a need for additional staff in other departments in the Europe region, which could be addressed by transferring employees from the Americas region. 4. The Executive department in the Americas region has three employees, which could be an area to consider for cost reduction by evaluating the necessity of all three positions and potentially consolidating roles or shifting employees to other departments or regions. 5. Overall, the data suggests that there may be opportunities for cost reduction and increased efficiency by</p> |



**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

*evaluating the distribution of employees across departments and regions, and considering strategic shifts of employees to address imbalances and reduce redundancy in certain areas.*

Lampiran 15 Hasil Skenario NB-4 Pada Literatur

| <i>Data Analyst</i> | <i>Hasil</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Senior              | <i>1. CAS is the most popular room if considering reservations, with 69 reservations. 2. It only wins by 3 reservations compared with the room with the second most reservations, FNA. 3. The minimum reserved room is AOB, with 51 records. 4. Overall, the gap between the rooms is not hugely different. 5. Besides reservations, we could also measure the "popularity" by the number of guests.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Junior              | <i>1. 'Convoke and sanguine' and 'Frugal not apropos' and 'immutable before decorum' are the most popular rooms as they have the most amount of reservations. 2. Frugal not apropos has the highest average rate based on reservations. 3. Most of rooms either has traditional or modern decoration. 4. Base price is usually associate with the max number of occupancy. 5. Reservation checkin and checkout data are not in standard format, it's unclear which year it's referring to.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| GPT-4               | <i>1. The most popular room is "Convoke and sanguine" with 69 visitors, indicating that it may have the most appealing features or activities for guests. 2. The least popular room is "Abscond or bolster" with 51 visitors, suggesting that it may not be as attractive or engaging as the other rooms. 3. The top three rooms ("Convoke and sanguine," "Frugal not apropos," and "Immutable before decorum") have a relatively close number of visitors, ranging from 63 to 69, indicating that they may share similar characteristics that make them popular among guests. 4. The bottom three rooms ("Mendicant with cryptic," "Riddle to exculpate," and "Abscond or bolster") have a more significant difference in visitor numbers, ranging from 51 to 56, suggesting that there may be more variation in their appeal to guests. 5. Overall, there is a relatively even distribution of visitors across the rooms, with only an 18-visitor difference between the most and least popular rooms. This could indicate that guests have diverse preferences and interests, or that the rooms offer a variety of experiences that cater to different tastes.</i> |

Lampiran 16 Hasil Skenario NB-5 Pada Literatur

**Hak Cipta :**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

| Data Analyst | Hasil                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Senior       | <i>1. The metric we choose to measure ROI. It's calculated as Score/Price. We interpret it as the score value gained for every 1 dollar. 2. Further, we set the minimum score required to be 90. 3. Based on this metric, we would recommend Rose Wine Creek Ranch. 4. Rose Wine Creek Ranch's score is not high (only 90), however, it's only 15 dollars. This is super low compared to the rest. 5. Hence, we would recommend Rose Wine Creek Ranch, as it provides the best value for a dollar.</i>                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Junior       | <i>1. We should recommend wine Papas Block to him as this wine is within his budget and not made of Zindandel and has highest Score. 2. The mean score are roughly the same across all different type of grapes. 3. All wines are from California. 4. Most of wines have 'now' as value of drink but there are also some other years appear in this column. 5. Most of the wines are associated with year 2007 which is suspected to be the made year, followed by year 2008 and 2009.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| GPT-4        | <i>1. The majority of the wines listed are within the client's budget of 60, providing a wide range of options for recommendations. 2. Sauvignon Blanc, Pinot Noir, and Chardonnay are the most frequently listed wine varieties, suggesting that these are popular choices and could be suitable recommendations for the client. 3. The price range for the wines listed is quite broad, from as low as 9 to as high as 60, indicating that there are options available for various preferences and budgets. 4. The client specifically mentioned not liking Zinfandel, and none of the wines listed are Zinfandel, ensuring that all options provided are suitable for the client's taste preferences. 5. Some wines have multiple price points, such as Sauvignon Blanc and Pinot Noir, which could indicate different vintages, vineyards, or quality levels, providing further variety for the client to choose from.</i> |

Lampiran 17 Contoh Hasil Laporan Analisis



### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

## Which Room is More Popular? An Executive Analysis Report

### Executive Summary

**The Big Picture:** Our analysis of reservation data reveals clear popularity rankings across 10 distinct hotel rooms. The most popular room is **Convoke and sanguine** with **69 reservations**, while the least popular room is **Abscond or bolster** with **51 reservations**. This represents a **35% difference in booking frequency** between the most and least popular rooms.

**So What?** This insight directly answers the business question: **Convoke and sanguine is the most popular room**. The 18-reservation gap suggests significant opportunities to optimize room inventory, marketing focus, and pricing strategies. Rooms with higher reservation counts likely have features that appeal to guests, while lower-performing rooms may need targeted improvements or promotional strategies.

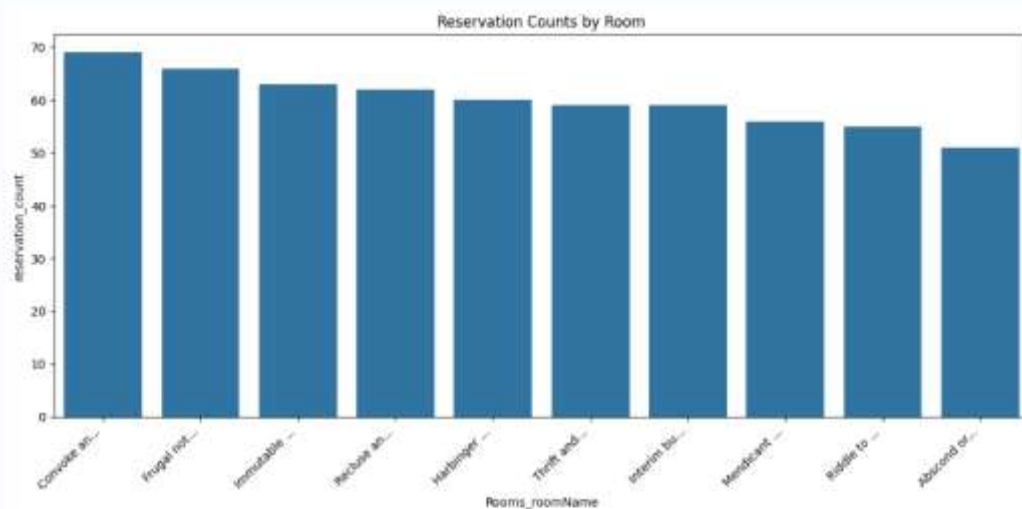
### The Deep Dive: 1 Critical Business Question

#### Question 1: Which room is more popular?

**The Question:** Which room is more popular?

**Insight 1: Room Popularity Ranking**

**The Data:** Reservation counts per room, derived from the Reservations\_Room column linked to room names.



**What to Look For:** The bar chart displays reservation counts for each of the 10 unique rooms. Look for the tallest bar (highest count) and the shortest bar (lowest count). Notice the relative heights of each bar to understand popularity distribution.

**The Insight:** **Convoke and sanguine (CAS)** is the most popular room with **69 reservations**, while **Abscond or bolster (AOB)** is the least popular with **51 reservations**. The average reservation count across all rooms is 60, with a standard deviation of 5.31, indicating relatively consistent booking volumes. The most popular room has **35% more reservations** than the least popular room.



## © Hak Cipta milik Jurusan TIK Politeknik Negeri Jakarta

### Hak Cipta :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar Politeknik Negeri Jakarta
2. Dilarang mengumpulkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin dari Jurusan TIK Politeknik Negeri Jakarta

**Strategic Implication:** The business should consider **promoting Convoke and sanguine** as a premium option given its high demand. For Abscond or bolster, investigate potential improvements in room features, amenities, or marketing to increase its appeal. The consistent booking distribution suggests all rooms have stable demand, but with clear differentiation.

### Strategic Recommendations

- **Leverage High Demand:** Feature Convoke and sanguine in marketing campaigns as the "most popular" room to attract guests seeking proven accommodations.
- **Investigate Low Performance:** Analyze why Abscond or bolster has fewer reservations. Consider room decor, size, amenities, or location factors that may affect guest preference.
- **Optimize Pricing:** Use demand data to implement dynamic pricing—higher prices for popular rooms and promotional rates for less popular ones to balance occupancy.
- **Monitor Trends:** Track reservation patterns over time to identify seasonal trends or emerging changes in room popularity.
- **Enhance Room Features:** If feasible, consider updates to lower-performing rooms based on guest feedback or competitive analysis.

