



© Hak Cipta milik 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 Politeknik Negeri Jakarta



IMPLEMENTASI METODE SUPPORT VECTOR REGRESSION UNTUK PREDIKSI KANDUNGAN GULA PADA MADU LEBAH TANPA SENGAT

LAPORAN SKRIPSI

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

ADAM CHRISAMUEL ALEXANDER

2007412019

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER
POLITEKNIK NEGERI JAKARTA
DEPOK
2025**



© Hak Cipta milik 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



IMPLEMENTASI METODE SUPPORT VECTOR REGRESSION UNTUK PREDIKSI KANDUNGAN GULA PADA MADU LEBAH TANPA SENGAT

LAPORAN SKRIPSI

ADAM CHRISAMUEL ALEXANDER

2007412019

**POLITEKNIK
NEGERI
JAKARTA**

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER
POLITEKNIK NEGERI JAKARTA
DEPOK
2025**



© Hak Cipta milik 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertanda tangan di bawah ini:.

Nama Mahasiswa : Adam Chrisamuel Alexander
NIM : 2007412019
Jurusan / Program Studi : T. Informatika dan Komputer / Teknik Informatika
Judul Skripsi : Implementasi Metode Support Vector Regression
Untuk Prediksi Kandungan Gula Pada Madu Lebah
Tanpa Sengat

Menyatakan bahwa skripsi yang berjudul "Implementasi Metode Support Vector Regression Untuk Prediksi Kandungan Gula Pada Madu Lebah Tanpa Sengat" benar-benar merupakan hasil karya saya sendiri, bebas dari peniruan terhadap karya 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 atau bentuk-bentuk peniruan lain yang dianggap melanggar peraturan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Tangerang 7 Juli 2025,

Yang membuat pernyataan



Adam Chrisamuel Alexander

2007412019



© Hak Cipta milik 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

LEMBAR PENGESAHAN

Skripsi diajukan oleh:

Nama : Adam Chrisamuel Alexander
NIM : 2007412019
Jurusan/Program Studi : T.Informatika dan Komputer/Teknik Informatika
Judul Skripsi : Implementasi Metode Support Vector Regression
Untuk Prediksi Kandungan Gula Pada Madu Lebah
Tanpa Sengat

Telah diuji oleh tim penguji dalam Sidang Skripsi pada hari Jumat, Tanggal 11,
Bulan Juli, Tahun 2025 dan dinyatakan LULUS

Disahkan oleh

Pembimbing I : Euis Oktavianti, S.Si., M.T.I

Penguji I : Rizki Elisa Nalawati, S.T., M.T
Penguji II : Iklima Ermis Ismail, S.Kom., M.Kom
Penguji III : Maria Agustin, S.Kom., M.Kom

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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

KATA PENGANTAR

Puji serta Syukur penulis panjatkan kehadirat Tuhan YME atas anugerah, dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini dengan judul “Implementasi Metode Support Vector Regression Untuk Prediksi Kandungan Gula Pada Madu Lebah Tanpa Sengat”. Penyusunan skripsi ini sebagai salah satu syarat kelulusan Program Studi Teknik Informatika, Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta. Dalam penyusunan skripsi ini banyak pihak yang memberikan bantuan kepada penulis sehingga skripsi ini dapat diselesaikan. Oleh karena itu penulis ingin menyampaikan rasa terima kasih sebesar-besarnya kepada:

1. Tuhan YME yang telah memberikan karunia-Nya berupa kesehatan fisik, kesehatan mental dan akal sehat sehingga dapat menyelesaikan penelitian dan penulisan skripsi ini dengan baik
2. Orang tua dan keluarga yang telah memberikan dukungan berupa doa serta kebutuhan materi dan non materi
3. Ibu Dr. Anita Hidayati, S.Kom., M.Kom., selaku Ketua Jurusan Teknik Informatika dan Komputer.
4. Ibu Euis Oktavianti, S.Si., M.TI., selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran dalam mengarahkan dalam penulisan skripsi.
5. Teman-teman yang selalu mendukung dan memberi semangat pada saat proses pengerjaan skripsi.

Tidak ada kata yang pantas penulis sampaikan selain terima kasih kepada semua pihak yang telah membantu hingga laporan ini terselesaikan dengan baik. Oleh karena itu, penulis mengharapkan segala bentuk kritik, saran, dan masukan yang dapat membantu memperbaiki serta menyempurnakan skripsi ini.

Tangerang, 7 Juli 2025

Adam Chrisamuel Alexander



IMPLEMENTASI METODE SUPPORT VECTOR REGRESSION UNTUK PREDIKSI KANDUNGAN GULA PADA MADU LEBAH TANPA SENGAT

ABSTRAK

Pemalsuan madu lebah tanpa sengat (SBH) merupakan masalah serius yang memerlukan metode deteksi akurat. Penelitian ini bertujuan mengembangkan model prediksi kandungan gula madu menggunakan Support Vector Regression (SVR) yang kinerjanya dioptimalkan dengan seleksi fitur Forward Selection (FS). Model dikembangkan dari dataset spektrum UV-Vis 200 sampel, di mana tantangan data tidak lengkap secara signifikan berhasil diatasi menggunakan imputasi. Kernel Radial Basis Function (RBF) dipilih karena kemampuannya menangani pola data non-linear yang kompleks. Evaluasi dengan K-Fold Cross-Validation menunjukkan bahwa model dengan Forward Selection lebih unggul. Performa optimal dicapai pada 6-Fold Cross-Validation dengan nilai R^2 sebesar 0.8697, RMSE 0.0519, dan MAE 0.0398. Metode Forward Selection juga berhasil mengidentifikasi panjang gelombang spesifik yang paling berpengaruh terhadap kandungan gula. Kemampuan generalisasi model kemudian divalidasi pada unseen data dengan hasil R^2 0.8671. Hasil ini membuktikan bahwa kombinasi SVR dan Forward Selection efektif untuk membangun sistem prediksi kemurnian madu yang akurat.

Kata Kunci: *Support Vector Regression, Forward Selection, Madu Lebah Tanpa Sengat, Ridge Imputation, Prediksi.*

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 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

DAFTAR ISI

LEMBAR PENGESAHAN	ii
KATA PENGANTAR	iii
ABSTRAK	iv
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
BAB I	1
1. 1 Latar Belakang	1
1. 2 Perumusan Masalah	2
1. 3 Batasan Masalah	2
1. 4 Tujuan dan Manfaat	2
1.4.1 Tujuan	2
1.4.2 Manfaat	3
1. 5 Sistemika Penulisan	3
BAB II	4
2. 1 Penelitian Terdahulu	4
2. 2 Madu	8
2. 3 Lebah Tanpa Sengat	8
2. 4 Machine Learning	9
2.4.1 Support Vector Regression	9
2.4.2 Forward Selection	12
2. 5 Cross Industry Standard Process Data Mining (CRISP-DM)	12
Business Understanding	13
Data Understanding	13



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 Politeknik Negeri Jakarta

Data preparation.....	13
Model	14
Evaluation	14
Deployment.....	15
2.6 <i>System Development Life Cycle (SDLC)</i>	15
2.6.1 Waterfall	16
2.7 Python	17
2.7.1 Streamlit.....	18
2.7.2 FastAPI	18
2.8 Aplikasi Web	19
BAB III	20
3.1 Rancangan Penelitian.....	20
3.2 Tahapan Penelitian.....	20
3.2.1 Identifikasi Masalah.....	20
3.2.2 Studi Literatur	20
3.2.2 Persiapan Data	21
3.2.3 Pengembangan Model.....	21
3.2.4 Evaluasi Model	21
3.2.5 Implementasi.....	21
3.2.6 Analisis Data	21
3.2.7 Pelaporan.....	22
3.3 Objek Penelitian.....	22
BAB IV	23
4.1 Analisis Kebutuhan.....	23
4.1.1 Kebutuhan Pengembangan Model	23
4.1.2 Kebutuhan Pengembangan Web	24



© Hak Cipta milik 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 Politeknik Negeri Jakarta

4.1.3	Kebutuhan Perangkat Keras dan Perangkat Lunak.....	25
4.2	Perancangan Sistem	26
4.2.1	Perancangan Model.....	26
4.2.2	Perancangan Web.....	29
4.3	Implementasi Sistem	32
4.3.1	Implementasi Model Regresi	33
4.3.2	Implementasi Web	43
4.4	Pengujian.....	44
4.4.1	Deskripsi Pengujian	44
4.4.2	Prosedur Pengujian	46
4.4.3	Data Hasil Pengujian.....	47
4.4.4	Analisis Data Pengujian	49
BAB V		51
5.1	KESIMPULAN.....	51
5.2	SARAN	51
DAFTAR PUSTAKA		52

**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 Politeknik Negeri Jakarta

DAFTAR GAMBAR

Gambar 2. 1 Lebah Tanpa Sengat.....	8
Gambar 2. 2 Cross Industry Standard Process Data Mining (CRISP-DM) .	12
Gambar 2. 3 SDLC	15
Gambar 2. 4 Metode Waterfall.....	16
Gambar 2. 5 Python.....	17
Gambar 2. 6 Streamlit	18
Gambar 2. 7 FastAPI.....	18
Gambar 4. 1 Dataset yang digunakan.....	27
Gambar 4. 2 Use Case Diagram.....	29
Gambar 4. 3 Activity Diagram Mengunggah dan Prediksi file	30
Gambar 4. 4 Activity Diagram Visualisasi Data	31
Gambar 4. 5 Activity Diagram Mengunduh Data	32
Gambar 4. 6 Informasi Dataset	34
Gambar 4. 7 Informasi missing values.....	34
Gambar 4. 8 Informasi Data	35
Gambar 4. 9 Dataset setelah Ridge Imputation	36
Gambar 4. 10 K-fold	37
Gambar 4. 11 Inisiasi Target dan Fitur	38
Gambar 4. 12 Pipeline	39
Gambar 4. 13 K-fold cross validation	40
Gambar 4. 14 Halaman Web	43
Gambar 4. 15 Web Setelah mengunggah Data	43
Gambar 4. 16 Halaman Visualisasi	44



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 Politeknik Negeri Jakarta

DAFTAR TABEL

Tabel 1. Penelitian Terdahulu	4
Tabel 2. Kebutuhan Dataset.....	23
Tabel 3. Kebutuhan Model.....	24
Tabel 4. Kebutuhan Fungsional.....	25
Tabel 5. Kebutuhan Non-Fungsional	25
Tabel 6. Kebutuhan Perangkat Keras	26
Tabel 7. Tabel Perangkat Lunak.....	26
Tabel 8. Perbandingan Kernel SVR.....	38
Tabel 9. SVR dengan Forward Selection	41
Tabel 10. SVR tanpa Forward Selection.....	41
Tabel 11. Perbandingan Model.....	42
Tabel 12. Skenario Blackbox Testing	45
Tabel 13. Hasil Blackbox Testing	48
Tabel 14. Perbandingan Model.....	48
Tabel 15 Perbandingan Model dataset berbeda	49
Tabel 16. Evaluasi pada data unseen	49



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 Politeknik Negeri Jakarta

BAB I

PENDAHULUAN

1. 1 Latar Belakang

Madu menempati peringkat ketiga sebagai makanan yang paling sering dipalsukan di dunia, dengan madu lebah tanpa sengat (SBH) menjadi target utama karena keterbatasan produksi dan meningkatnya permintaan pasar. Penipuan ini sering terjadi melalui penambahan bahan kimia industri (seperti sirup jagung dan tebu) atau label palsu, yang tidak hanya mengurangi nilai gizi madu tetapi juga dapat menyebabkan tekanan darah tinggi, diabetes, dan penumpukan lemak perut. (Suhandy et al., 2024). Karakteristik alami madu lebah tanpa sengat, seperti kadar air tinggi dan keasaman rendah, meningkatkan kerentanan terhadap pemalsuan (Yong et al., 2022). Salah satu metode yang dapat digunakan untuk menganalisis karakteristik kandungan madu adalah spektroskopi fluoresensi. Metode ini memungkinkan evaluasi komposisi dan keaslian sampel madu yang lebih akurat melalui pengukuran spektrum cahaya yang dihasilkan oleh molekul akibat eksitasi elektronik pada molekul (Hassoun et al., 2020).

Implementasi prediksi kandungan gula pada madu lebah tanpa sengat, salah satunya memakai metode berbasis komputer, khususnya *machine learning*. Pembelajaran mesin atau *machine learning* merupakan pengembangan sistem atau mesin cerdas yang dapat belajar secara mandiri tanpa diprogram berulang kali oleh manusia (Dash et al., 2020). Beberapa algoritma yang dipakai dalam *machine learning* untuk menyelesaikan masalah regresi antara lain *Support Vector Regression*, *Random Forest Regression*, serta *Multiple Linear Regression* (Wang et al., 2022). Masing-masing algoritma memiliki keunggulan tersendiri tergantung pada karakteristik data dan tujuan analisis yang diinginkan.

Penyelesaian persoalan regresi pada penelitian ini akan dilakukan dengan menggunakan *Support Vector Regression* (SVR), di mana nilai-nilai numerik diprediksi berdasarkan fitur input yang disediakan melalui *forward selection*. SVR sering digunakan dalam pemodelan regresi karena kemampuannya dalam



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 Politeknik Negeri Jakarta

menangani hubungan linier maupun non-linier melalui fungsi kernel yang fleksibel (Parbat and Chakraborty, 2020). SVR juga menunjukkan performa yang sangat baik dalam prediksi numerik, terutama ketika diterapkan pada data multivariabel yang kompleks. Selain akurat, metode ini juga mampu mempertahankan efisiensi komputasi, sehingga cocok untuk regresi berbasis data berskala menengah hingga besar (Razavi and Kenari, 2023).

1. 2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, maka dapat dirumuskan masalahnya adalah “Bagaimana mengimplementasikan metode algoritma *support vector regression* untuk prediksi kandungan gula pada madu lebah tanpa sengat?”

1. 3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah

- a) Metode yang digunakan dalam pembuatan model adalah *support vector regression*
- b) Data yang digunakan dari *sample* madu lebah tanpa sengat
- c) *Sample* madu yang diperoleh dari 2 lokasi; Lampung dan Rangkasbitung
- d) Data terdiri dari 200 sampel, dimana hanya 4 data yang memiliki nilai lengkap secara keseluruhan.
- e) Aplikasi ada

1. 4 Tujuan dan Manfaat

1.4.1 Tujuan

Tujuan dari penelitian ini adalah untuk mengimplementasikan metode *support vector regression* dalam proses prediksi kandungan gula pada madu lebah tanpa sengat.



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 Politeknik Negeri Jakarta

1.4.2 Manfaat

Manfaat yang diharapkan dari penelitian ini adalah untuk membantu peternak lebah tanpa sengat dalam mengetahui kandungan gula pada madu secara praktis tanpa perlu melalui uji laboratorium.

1.5 Sistemika Penulisan

Sistematika Penulisan dalam penelitian ini terdiri atas 5 bab, yaitu:

BAB I PENDAHULUAN

Bab I berisikan latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat, serta sistematika penulisan yang akan dikaitkan pada implementasi metode algoritma *support vector regression* untuk prediksi kandungan gula pada madu lebah tanpa sengat.

BAB II TINJAUAN PUSTAKA

Bab II merupakan bab yang berisi tentang tinjauan pustaka serta landasan teori atau kajian ilmu secara garis besar mengenai topik pembahasan pada penelitian ini, hingga beberapa penelitian yang telah dilakukan sebelumnya.

BAB III METODE PENELITIAN

Bab III berisi uraian tentang metode yang akan digunakan, meliputi rancangan penelitian, tahapan penelitian, objek penelitian, aplikasi pembangunan yang digunakan, teknik pengumpulan dan analisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab IV dari pembahasan menjelaskan tentang pemaparan dan analisis pengujian seperti deskripsi prosedur pengujian yang terdiri dari pengujian untuk menguji aplikasi prediksi kandungan gula pada madu lebah tanpa sengat dan hasil analisis data atau evaluasi dari pengujian.

BAB V PENUTUP

Bab V dari penutup menjelaskan mengenai kesimpulan akhir dan saran dari penelitian serta untuk proses pengujian 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

BAB II

TINJAUAN PUSTAKA

2. 1 Penelitian Terdahulu

Tabel 1. Penelitian Terdahulu

Peneliti, Tahun, dan Judul	Metode	Kesimpulan
Ultraviolet–visible spectroscopy combined with machine learning as a rapid detection method to the predict adulteration of honey (Razavi and Kenari, 2023)	<i>PLSR dan SVR</i>	Performa model PLSR dan SVR dalam memprediksi kandungan sukrosa pada sampel madu menggunakan data spektrum UV-Vis. Hasilnya menunjukkan bahwa model SVR memiliki akurasi prediksi lebih tinggi dengan nilai RMSE sebesar 0,97 dan R^2 sebesar 0,98, menjadikannya lebih unggul dibandingkan model PLSR.
Research on Support Vector Regression Short-Time Traffic Flow Prediction Model for Secondary Roads Based on Associated Road Analysis (Duan et al., 2025)	<i>SVR, LSTM, Random Forest, GBDT</i>	Penelitian ini membandingkan kinerja empat algoritma prediksi lalu lintas jangka pendek, yaitu SVR, LSTM, Random Forest, dan GBDT. Hasilnya menunjukkan bahwa model SVR menghasilkan performa terbaik dengan nilai RMSE = 11.422 dan MAE = 7.017, mengungguli model lainnya, sehingga dianggap paling akurat dalam memprediksi volume lalu lintas di jalan sekunder.



© Hak Cipta milik 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 Politeknik Negeri Jakarta

Evaluation of the SVR and Random Forest Models Accuracy for Streamflow Prediction under a Data-Scarce Basin in Morocco (Bouchra Bargam et al., 2024)	SVR, RF, Multiple Linear Regression (MLR)	SVR menunjukkan performa prediksi terbaik dibanding RF dan MLR dalam memprediksi debit sungai harian di daerah dengan keterbatasan data, dengan nilai RMSE 1.18 m ³ /s dan NSE 0.59. Hasil ini membuktikan efektivitas SVR untuk regresi dalam konteks hidrologi.
--	---	--

Support Vector Regression (SVR) telah digunakan secara luas dalam berbagai penelitian sebagai metode yang efektif untuk menyelesaikan permasalahan regresi, terutama ketika data yang digunakan memiliki pola yang tidak linier dan kompleks. Razavi dan Kenari (2023) menunjukkan bahwa SVR mampu memprediksi kadar sukrosa pada madu dengan tingkat akurasi yang tinggi menggunakan data spektrum UV-Vis, ditunjukkan oleh nilai RMSE sebesar 0,97 dan R² sebesar 0,98. Penelitian tersebut memperlihatkan bahwa SVR dapat mengolah data spektral secara efisien untuk menghasilkan prediksi numerik yang akurat.

Penelitian, Duan et al. (2025) mengembangkan model prediksi arus lalu lintas jangka pendek dengan pendekatan SVR yang dikombinasikan dengan analisis korelasi antar jalan. Hasil evaluasi menunjukkan bahwa model SVR memberikan performa terbaik dibandingkan dengan algoritma lainnya, dengan nilai RMSE dan MAE yang lebih rendah, menegaskan kemampuan SVR dalam menghadapi data *time-series* yang dinamis dan tidak linier.

Studi oleh Bargam et al. (2024) juga memperkuat efektivitas SVR dalam konteks regresi, di mana SVR digunakan untuk memprediksi debit sungai harian di daerah dengan keterbatasan data. SVR mampu memberikan performa prediksi yang akurat meskipun data latih yang digunakan bersifat terbatas dan tidak ideal.

Berdasarkan temuan dari berbagai studi tersebut, SVR terbukti menjadi salah satu metode regresi yang kuat dan fleksibel. Kelebihannya terletak pada kemampuannya membentuk model prediksi yang tahan terhadap outlier, bekerja optimal pada data berdimensi tinggi, serta mampu menyesuaikan terhadap hubungan nonlinier antar



Hak Cipta :

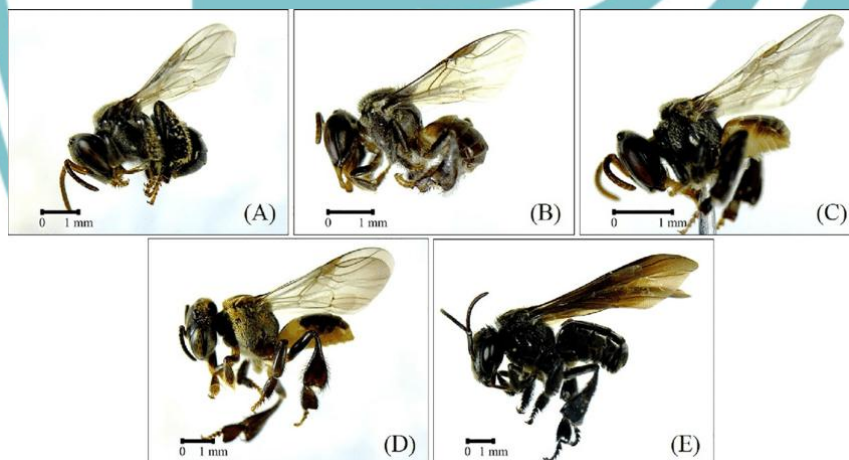
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

fitur. Oleh karena itu, pemanfaatan SVR dalam penelitian ini diharapkan mampu menghasilkan model prediksi yang andal dan akurat, sesuai dengan kebutuhan analisis data yang kompleks.

2.2 Madu

Madu adalah cairan kental berwarna kuning kecokelatan dengan rasa manis khas yang dihasilkan oleh lebah melalui pengolahan nektar bunga. Berdasarkan Standar Nasional Indonesia (SNI) Nomor 3545 Tahun 2013, madu diklasifikasikan sebagai cairan alami manis yang umumnya berasal dari sari bunga tanaman (floral nektar) maupun bagian tanaman lainnya (ekstra floral), dan dihasilkan oleh lebah madu (*Apis* sp.) Madu dapat diperoleh dari berbagai jenis lebah, termasuk lebah ternak seperti *Apis mellifera*, lebah tanpa sengat seperti *Trigona*, maupun lebah hutan seperti *Apis dorsata* (Hidayatullah, Handoko & Maring, 2022).

2.3 Lebah Tanpa Sengat



Gambar 2.1 Lebah Tanpa Sengat

(Sumber: (Sayusti *et al.*, 2020))

Lebah tanpa sengat merupakan jenis lebah sosial yang mendiami wilayah tropis dan subtropis, dengan jumlah spesies yang telah teridentifikasi mencapai lebih dari 550, menjadikannya sebagai kelompok lebah paling besar dan beragam secara ilmiah (Shanahan & Spivak, 2021). Umumnya, lebah ini hidup di pepohonan besar dan lebat yang memiliki suhu serta kelembapan yang stabil, dan kaya akan sumber pakan alami (N. Prasetyo, Senoaji & Hidayat, 2022). Meskipun demikian, lebah



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 Politeknik Negeri Jakarta

tanpa sengat memiliki keterbatasan dalam hal produksi madu, dengan kapasitas sarang yang rendah dan biaya pemeliharaan yang relatif tinggi.

Pengelolaan madu lebah tanpa sengat ini menghadapi sejumlah tantangan, seperti ketergantungan terhadap kondisi cuaca yang memengaruhi vegetasi, ancaman predator, serta akses pasar yang masih terbatas. Secara ilmiah, madu dari lebah tanpa sengat telah terbukti memiliki nilai medis yang tinggi, yang menjadikannya semakin diminati (Hidalgo, Nicolas & Cedon, 2020). Walaupun permintaan terus tumbuh, industri madu tanpa sengat masih menyimpan potensi pertumbuhan yang besar. Namun demikian, isu-isu seperti pemalsuan dan produksi madu buatan menjadi tantangan serius terkait upaya menjaga kemurnian produknya.

2. 4 Machine Learning

Machine learning adalah studi dan pengembangan pendekatan yang mengotomatisasi pengambilan keputusan kompleks, karena hal tersebut memungkinkan komputer untuk menemukan aturan prediksi dari pola dalam data yang berlabel tanpa instruksi eksplisit (Raschka, Patterson and Nolet, 2020).

2.4.1 Support Vector Regression

Support Vector Regression (SVR) digunakan untuk memecahkan masalah regresi dengan memprediksi nilai numerik berdasarkan fitur-fitur input yang diberikan. SVR bekerja dengan mencari fungsi regresi yang menyimpang tidak lebih dari nilai ambang ϵ dari nilai sebenarnya pada data pelatihan, sambil menjaga model tetap sesederhana mungkin. Untuk data non-linear, SVR menggunakan fungsi kernel untuk memproyeksikan data ke ruang berdimensi lebih tinggi, sehingga pemisahan atau fitting dapat dilakukan secara linear di ruang tersebut.

SVR adalah pilihan populer dalam prediksi dan penyesuaian kurva, baik untuk regresi linear maupun non-linear. Model ini didasarkan pada elemen-elemen dari Support Vector Machine (SVM), di mana support vectors merupakan titik-titik yang paling dekat dengan hyperplane yang dihasilkan dalam ruang fitur berdimensi- n . Titik-titik ini secara jelas memisahkan data sekitar hyperplane tersebut (Parbat and Chakraborty, 2020). SVR efektif dalam menangani data 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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

non-linear dan kompleks, memberikan hasil yang lebih akurat dan dapat diandalkan, terutama ketika fitur-fitur yang relevan telah dipilih melalui teknik seperti forward selection (Zhang & O'Donnell, 2020).

SVR mempunyai fungsi dasar seperti berikut:

$$f(x) = \omega^T x + b \quad (1)$$

Dimana:

- ω adalah vektor bobot (weight vector),
- x adalah vektor fitur (fitur input),
- b adalah bias atau intercept.
- ω^T transpose dari ω , untuk memastikan dimensi sesuai dalam perkalian vektor.

Dalam praktiknya, masalah SVR diselesaikan dalam bentuk dual menggunakan metode maksimisasi terhadap fungsi objektif Lagrangian, agar dapat diaplikasikan dengan kernel. Rumus optimisasi dual SVR adalah:

$$\text{Max} - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k (a_i - a_i^*)(a_j - a_j^*) K(X_i, X_j) + \sum_{i=1}^n (a_i - a_i^*) y_i - \varepsilon \sum_{i=1}^n (a_i + a_i^*) \quad (2)$$

Dengan syarat (kendala):

$$\sum_{i=1}^n (a_i - a_i^*) = 0, \quad 0 \leq a_i, a_i^* \leq C \quad (3)$$

Keterangan:

- $a_i - a_i^*$: koefisien Lagrange,
- C : parameter regulasi, mengontrol trade-off antara margin dan pelanggaran error,
- ε : margin toleransi terhadap error kecil,



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 Politeknik Negeri Jakarta

- $K(X_i, X_j)$: fungsi kernel yang menghitung kesamaan antar data.

Fungsi kernel adalah bagian penting dari metode SVR. Tujuan utama dari fungsi kernel adalah untuk mengubah data masukan ke dimensi yang lebih tinggi tanpa perlu menghitung koordinat baru. Ini dilakukan dengan cara yang memungkinkan hubungan non-linear yang ada pada ruang awal untuk diubah menjadi hubungan linear di ruang dimensi baru, yang memungkinkan SVR untuk beroperasi. Dengan cara ini, data yang semula tidak dapat dipisahkan secara linear akan dipeta (Alida and Mustikasari, 2020). Beberapa fungsi kernel yang terdapat pada SVR diantaranya kernel linear, kernel polynomial, dan kernel Radial Basis Function (RBF).

Beberapa jenis kernel umum:

- i. Linear: Saat data dapat dipisahkan atau diregresi secara linear, kernel Linear digunakan. Tidak ada transformasi ke ruang dimensi yang lebih besar, sehingga proses komputasi menjadi lebih mudah dan lebih cepat. Cocok untuk dataset yang memiliki hubungan linier antar fitur.

$$K(X_i, X_j) = X_i \cdot X_j \quad (4)$$

- ii. Polynomial: Kernel ini memproyeksikan data ke dimensi yang lebih tinggi menggunakan kombinasi polinomial. Parameter d adalah derajat polinomial, dan r adalah konstanta penyesuaian. Cocok untuk Pola hubungan yang ada lebih kompleks, tetapi masih dapat dijelaskan melalui kombinasi polinomial antar fitur.

$$K(X_i, X_j) = (x_i^T x_j + c)^d \quad (5)$$

- iii. RBF (Radial Basis Function): Fungsi kernel ini bertugas melakukan transformasi non-linear pada input, memindahkannya ke ruang berdimensi tinggi. Kemampuannya memungkinkan penanganan hubungan data yang kompleks, menjadikannya sesuai untuk data dengan pola non-linear yang tidak dapat diselesaikan oleh model linier.

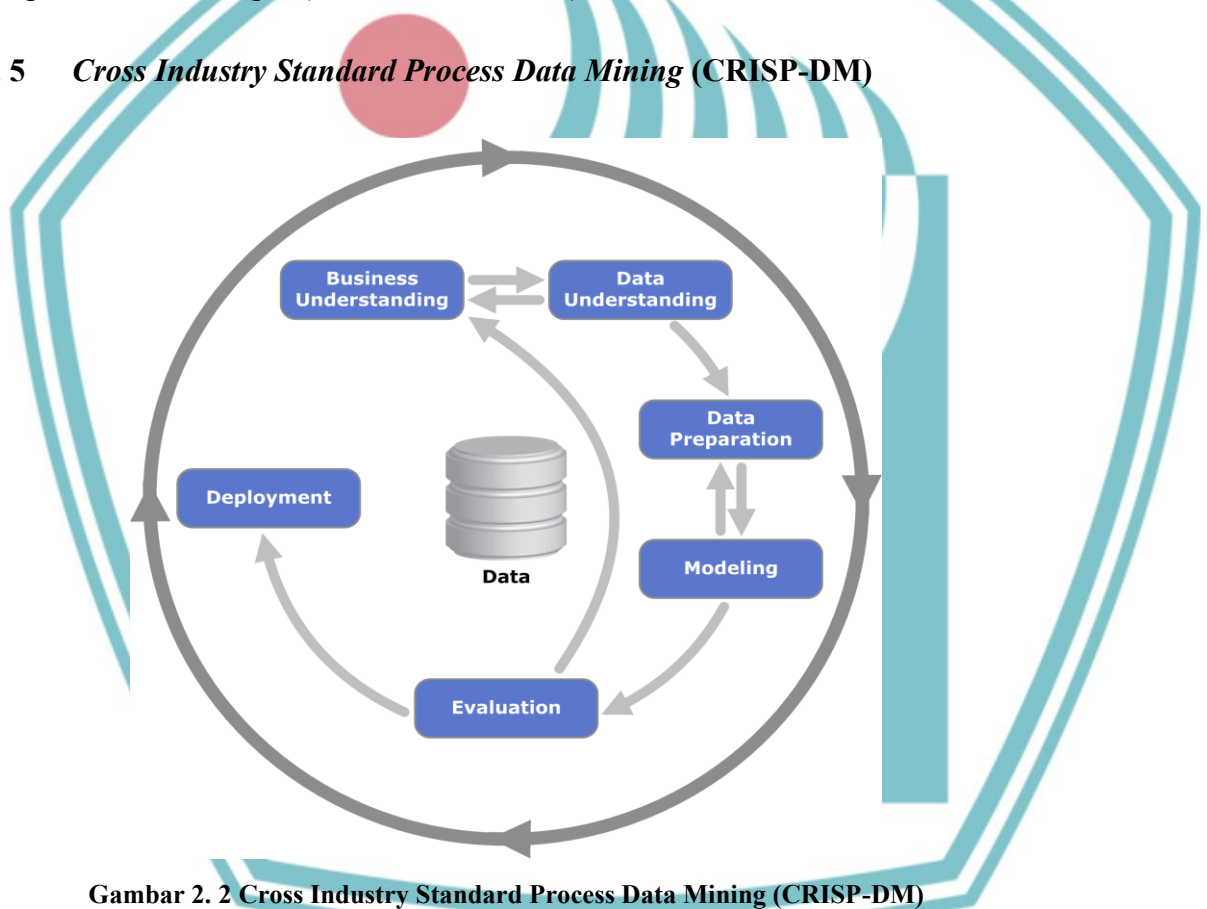
$$K(X_i, X_j) = \exp(-\gamma(X_i - X_j)^2) \quad (6)$$



2.4.2 Forward Selection

Forward selection merupakan salah satu metode *feature selection* yang bersifat *greedy* dan termasuk dalam kategori pendekatan *wrapper*. Teknik ini bekerja dengan cara memulai dari himpunan fitur kosong, kemudian secara bertahap menambahkan satu fitur pada setiap iterasi yang memberikan peningkatan paling signifikan terhadap performa model, misalnya dalam hal menurunkan error atau meningkatkan akurasi prediksi. Proses ini terus berlanjut hingga tidak ada lagi fitur yang dapat menambah kinerja model secara signifikan atau sampai jumlah fitur yang ditentukan tercapai (Kamalov et al, 2021).

2.5 Cross Industry Standard Process Data Mining (CRISP-DM)



Gambar 2. 2 Cross Industry Standard Process Data Mining (CRISP-DM)

(Sumber: (Oktavianti, 2023))

CRISP-DM adalah model proses standar yang banyak digunakan dalam proyek data mining dan data science. Model ini menyediakan kerangka kerja terstruktur yang membagi seluruh proses analisis data menjadi beberapa tahap utama, mulai dari memahami bisnis hingga menerapkan hasil (Schröer et al., 2021).

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 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 Politeknik Negeri Jakarta

CRISP-DM terdiri dari enam fase utama yang saling berurutan dan iteratif, yaitu: Business Understanding (BU), untuk membantu memahami tujuan bisnis dan merumuskan masalah; Data Understanding (DU), yang berkonsentrasi pada pengumpulan dan eksplorasi data awal; Data Preparation (DP), yang mencakup tahap pembersihan dan transformasi data; *Modelling* (M), penerapan algoritma untuk membangun model; Evaluation (E), untuk menilai performa model; dan Deployment (D), yaitu penerapan model agar dapat digunakan secara operasional (Oktavianti, Agustin and Sari, 2023).

Business Understanding

Tahap pertama dalam CRISP-DM yang bertujuan untuk memahami tujuan bisnis yang ingin dicapai melalui penerapan teknik data mining. Tahap ini memberikan penjelasan yang jelas mengenai tujuan bisnis dan alasan mengapa data mining diperlukan untuk mencapainya. Secara umum, tahap ini berisi deskripsi tentang tujuan bisnis serta penjelasan mengapa data mining dianggap relevan untuk konteks tersebut (Wurijanto, Setiawan, & Tjandrarini, 2022).

Data Understanding

Tahap kedua dalam CRISP-DM yang menitikberatkan pada pemahaman mengenai data yang digunakan dalam proses data mining. Pada tahap ini, penulis penelitian menguraikan sumber data yang digunakan dan menjelaskan bagaimana data tersebut dapat mendukung pencapaian tujuan yang telah ditentukan pada tahap Business Understanding (Wurijanto, Setiawan, & Tjandrarini, 2022).

Data preparation

Tahap persiapan data mencakup beberapa aktivitas seperti pemilihan, transformasi, dan pembersihan data. Transformasi bertujuan untuk menyesuaikan data agar lebih sesuai untuk analisis lebih lanjut. Sementara itu, pemilihan data dan pembersihan data mengikuti sebagai tugas yang penting, di mana data yang relevan dipilih dan data yang cacat atau tidak lengkap diperbaiki (Oğuz & Tilda, 2022). Beberapa studi juga menyebutkan secara eksplisit alat atau teknologi yang digunakan dalam tahap ini, seperti RapidMiner, R, MS Access, WEKA, dan Microsoft SQL Server.



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 Politeknik Negeri Jakarta

Feature selection adalah teknik yang bertujuan untuk memilih fitur yang paling relevan dari sekumpulan fitur yang tersedia. Proses ini penting untuk menyederhanakan model, meningkatkan efisiensi komputasi, dan mengurangi risiko overfitting. Proses ini dilakukan dengan menghilangkan fitur yang tidak penting atau bersifat redundan, sehingga model yang dihasilkan cenderung lebih ringan dan memiliki performa prediksi yang lebih baik. (Nurdin et al., 2024). Dalam membuat model prediksi digunakan forward selection sebagai salah satu bagian dari data preparation.

Model

Tahap pemodelan melibatkan penggunaan berbagai algoritma dan teknik data mining untuk membangun model yang dapat digunakan untuk memprediksi atau mengklasifikasikan data. Banyak studi yang membandingkan beberapa model atau algoritma untuk mengevaluasi perbedaan hasil yang dihasilkan oleh masing-masing. Berbagai model yang digunakan dalam teknik data mining meliputi Support Vector Machine (SVM) dan Random Forest untuk klasifikasi, Support Vector Regression (SVR) dan Linear Regression untuk regresi, serta K-Means dan DBSCAN untuk klustering. Masing-masing model ini memiliki keunggulannya sendiri tergantung pada jenis data dan tujuan analisis yang ingin dicapai.

Evaluation

Tahap evaluasi bertujuan untuk mengukur kualitas dari model yang telah dilatih menggunakan data yang telah dipersiapkan. Pada tahap ini, metrik evaluasi digunakan untuk menilai kinerja model. Salah satu metrik yang digunakan adalah *evaluation matrix*, yang membantu untuk menggambarkan sejauh mana model dapat memprediksi data dengan benar. Visualisasi model juga sering digunakan untuk membantu pemahaman hasil evaluasi, seperti visualisasi pohon keputusan yang menjelaskan proses pengambilan keputusan dalam model (Oğuz & Tilda, 2022).

Model yang telah dibangun akan diuji menggunakan berbagai metrik evaluasi, seperti Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2) untuk menilai kinerja model. Hasil pengujian ini kemudian akan dianalisis lebih lanjut pada tahap



Hak Cipta :

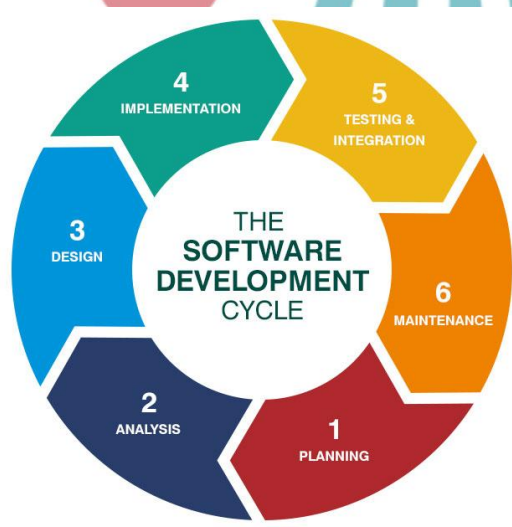
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

penulisan laporan untuk memberikan pemahaman yang lebih mendalam tentang efektivitas model yang dikembangkan.

Deployment

Tahap terakhir, yaitu tahap di mana model yang telah dilatih dan dievaluasi diimplementasikan dalam lingkungan nyata atau sistem yang berfungsi. Pada tahap ini, model yang dihasilkan akan digunakan untuk memberikan hasil yang dapat digunakan untuk mendukung pengambilan keputusan atau melakukan prediksi secara langsung Wurijanto, Setiawan, & Tjandrarini, 2022).

2.6 *System Development Life Cycle (SDLC)*



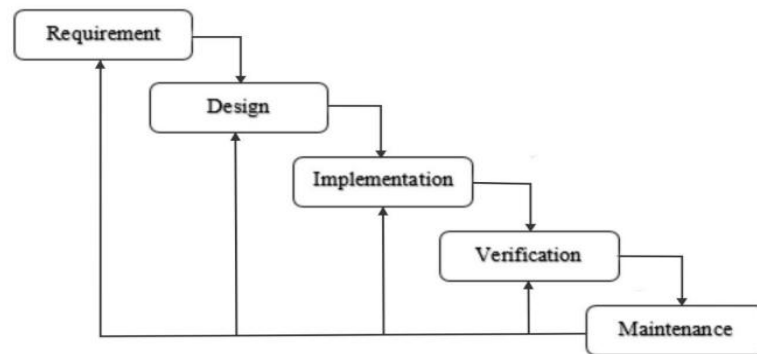
Gambar 2. 3 SDLC

(Sumber: (Chan, 2021))

Pendekatan terstruktur yang digunakan untuk merencanakan, merancang, mengembangkan, menguji, dan menerapkan sistem perangkat lunak. SDLC mencakup serangkaian langkah yang bertujuan untuk memastikan bahwa perangkat lunak yang dikembangkan memenuhi kebutuhan organisasi dan memiliki kualitas yang sesuai dengan harapan. Proses ini sangat penting dalam memastikan bahwa perangkat lunak dapat diproduksi dengan biaya yang efisien, dalam waktu yang tepat, dan dengan fungsionalitas yang sesuai dengan kebutuhan pengguna Ghumatkar & Date (2023).



2.6.1 Waterfall



Gambar 2. 4 Metode *Waterfall*

(Sumber: (Wahid, 2020))

Waterfall adalah metode dalam *System Development Life Cycle* (SDLC) yang bersifat berurutan dan linier. Setiap tahapan pengembangan sistem dilakukan secara berurutan, dan tidak ada kesempatan untuk kembali ke tahap yang sudah selesai. Metode ini sering digambarkan seperti air yang mengalir ke bawah, di mana satu tahap harus diselesaikan sepenuhnya sebelum melanjutkan ke tahap berikutnya. Waterfall sangat terstruktur dan mudah dipahami, tetapi memiliki keterbatasan dalam hal fleksibilitas. (Senarath, 2021). *Waterfall* mempunyai urutan yang terdiri dari beberapa langkah, seperti *requirement*, *design*, *implementation*, *testing*, dan *maintenance* untuk menyempurnakan hasilnya

a. *Requirement*

mengidentifikasi kebutuhan sistem secara menyeluruh. Tujuannya adalah melakukan pengumpulan informasi dari pengguna atau pihak terkait guna memahami fitur, fungsi, serta batasan sistem yang akan dibangun. Hasilnya dituangkan dalam dokumen spesifikasi kebutuhan sebagai acuan pengembangan.

b. *Design*

Setelah kebutuhan ditentukan, langkah selanjutnya adalah menyusun rancangan teknis. Perancangan ini meliputi struktur data, antarmuka pengguna, diagram alur sistem, serta arsitektur logika dan fisik.

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 Politeknik Negeri Jakarta



© Hak Cipta milik 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 Politeknik Negeri Jakarta

c. *Implementation*

Tahap ini adalah proses aktualisasi desain ke dalam bentuk program. Setiap bagian sistem dibangun secara bertahap dan disusun menjadi satu kesatuan fungsional.

d. *Testing*

Setelah perangkat lunak dikembangkan, dilakukan serangkaian pengujian untuk memastikan bahwa aplikasi berjalan sesuai rencana. Pengujian mencakup pengecekan kesalahan logika, integrasi antarmuka, dan kecocokan sistem dengan kebutuhan awal.

e. *Maintenance*

setelah perangkat lunak dibuat, fase pemeliharaan dilakukan untuk menangani berbagai masalah yang muncul. Ini mencakup perbaikan bug, peningkatan performa, atau penyesuaian terhadap kebutuhan baru yang mungkin muncul setelah sistem diterapkan.

2.7 Python

Python adalah bahasa pemrograman yang dinamis dan bermanfaat untuk beragam kebutuhan pengembangan perangkat lunak. Sebagai bahasa pemrograman open source, Python dapat digunakan secara gratis. Python dapat beroperasi di berbagai sistem operasi termasuk Windows, Linux, dan Mac. Selain itu, Python juga memiliki kemampuan untuk melakukan perhitungan matematika dan menyederhanakan proses aritmetika (Retnoningsih and Pramudita, 2020).



Gambar 2. 5 Python



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 Politeknik Negeri Jakarta

(Sumber: <https://www.python.org>)

2.7.1 Streamlit

Streamlit adalah framework open-source berbasis Python yang digunakan untuk membuat aplikasi web interaktif. Tanpa memerlukan keahlian pengembangan web yang mendalam dari pengguna, pembuatan antarmuka visual dapat dilakukan hanya dengan menggunakan skrip Python. Struktur ini sering digunakan dalam pengembangan aplikasi berbasis data, khususnya untuk kebutuhan visualisasi dan analisis data secara langsung, karena struktur sintaksisnya yang sederhana dan mudah dipahami (Khorasani et al., 2022).



Gambar 2. 6 Streamlit

(Sumber: streamlit.io)

2.7.2 FastAPI

FastAPI adalah framework web modern berbasis Python yang berfokus pada kecepatan dan efisiensi. FastAPI membuat pembelajaran dan pengembangan lebih mudah daripada framework seperti Flask dan Django. Meskipun masih relatif muda, popularitasnya meningkat seiring dengan dukungan komunitas. FastAPI sering digunakan dalam pengembangan API dan penerapan model kecerdasan buatan, terutama untuk deployment model pembelajaran mesin (Abdulazeez Adeshina, 2022).



Gambar 2. 7 FastAPI



© Hak Cipta milik Politeknik Negeri Jakarta

(Sumber: <https://fastapi.tiangolo.com>)

2.8 Aplikasi Web

Aplikasi *web* adalah sistem yang berjalan melalui internet dan dapat diakses menggunakan berbagai jenis perangkat melalui browser (Mandang et al., 2020). Keunggulannya terletak pada kemudahan akses dari lokasi mana pun, baik menggunakan komputer maupun gawai.



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 Politeknik Negeri Jakarta

BAB III

METODE PENELITIAN

3.1 Rancangan Penelitian

Secara metodologis, penelitian ini digolongkan sebagai penelitian kuantitatif. Penentuan ini didasarkan pada sifat penelitian yang sepenuhnya bergantung pada data numerik. Seluruh analisis dan penarikan kesimpulan dilakukan secara objektif melalui penerapan metode statistik dan algoritma komputasional. Pendekatan ini memastikan bahwa setiap kesimpulan yang ditarik bersifat objektif, terukur, dan dapat direproduksi.

3.2 Tahapan Penelitian

Tahapan ini memastikan untuk mendapat hasil yang maksimal dan dapat digunakan secara efektif. Tahapan tersebut meliputi identifikasi masalah, studi literatur, persiapan data, pengembangan model, evaluasi model, analisis data, dan pelaporan hasil penelitian.

3.2.1 Identifikasi Masalah

Langkah awal dalam penelitian ini dimulai dengan proses identifikasi masalah dan kebutuhan. Tahapan ini bertujuan untuk merumuskan secara jelas keperluan yang berkaitan dengan penelitian, baik dari segi pengembangan model maupun kebutuhan teknis. Identifikasi dilakukan agar proses penelitian dapat berjalan terarah dan sistematis. Aspek yang dianalisis meliputi kebutuhan fungsional dan non-fungsional, termasuk perangkat keras, perangkat lunak, serta metode pemodelan yang akan digunakan.

3.2.2 Studi Literatur

Mengumpulkan dan mempelajari literatur yang relevan terkait Support Vector Regression (SVR). Studi ini bertujuan untuk memahami teori dasar dan penelitian terdahulu yang mendukung pengembangan model prediktif.



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 Politeknik Negeri Jakarta

3.2.2 Persiapan Data

Dataset diperoleh dari uji lab pengukuran spektrum UV-Vis pada sampel madu lebah tanpa sengat. Tahapan ini mencakup imputasi nilai hilang menggunakan Ridge Regression untuk memastikan dataset bersih dan siap diproses. Setelah itu, dilakukan pra-pemrosesan berupa normalisasi data dan pemisahan antara fitur dan variabel target. Proses seleksi fitur menggunakan algoritma Forward Selection dipilih untuk membantu kinerja model.

3.2.3 Pengembangan Model

Model prediksi dikembangkan menggunakan algoritma Support Vector Regression (SVR). Untuk mengevaluasi kinerja dan stabilitas model, diterapkan teknik validasi silang (K-Fold Cross-Validation) dengan variasi nilai k , yaitu 4, 5, 6, 8, dan 10. Evaluasi model dilakukan dengan mengukur beberapa metrik performa, seperti R^2 , RMSE, MAE, dan MAPE.

3.2.4 Evaluasi Model

Model dilatih dan divalidasi menggunakan K-Fold, model terbaik berdasarkan performa validasi akan diuji lebih. Evaluasi ini bertujuan untuk mengukur kemampuan generalisasi model terhadap data baru, serta membandingkan kinerja antara model dengan dan tanpa penerapan seleksi fitur.

3.2.5 Implementasi

Model yang telah dilatih dan disimpan kemudian diintegrasikan ke dalam aplikasi web berbasis Python. Aplikasi ini dirancang untuk menerima input berupa spektrum UV-Vis dari pengguna dan memberikan output berupa prediksi kandungan gula secara otomatis dan interaktif.

3.2.6 Analisis Data

Analisis data dilakukan terhadap performa model yang telah dibangun. Analisis ini meliputi evaluasi terhadap akurasi prediksi, keefektifan teknik seleksi fitur menggunakan Forward Selection, serta perbandingan performa model dengan dan



© Hak Cipta milik 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

tanpa seleksi fitur. Evaluasi dilakukan berdasarkan metrik-metrik seperti koefisien determinasi (R^2), *root mean squared error* (RMSE), *mean squared error* (MSE), *mean absolute error* (MAE), dan *mean absolute percentage error* (MAPE),

3.2.7 Pelaporan

Hasil dari seluruh proses pengembangan dan evaluasi model akan disusun secara sistematis ke dalam laporan skripsi. Pelaporan ini mencakup dokumentasi tahapan penelitian, hasil pengujian model, serta pembahasan mengenai potensi penerapan model dalam sistem nyata, seperti aplikasi web yang dikembangkan dalam penelitian ini.

3.3 Objek Penelitian

Objek penelitian ini adalah data spektrum UV-Vis madu lebah tanpa sengat yang digunakan untuk membangun model prediksi kandungan gula pada madu melalui algoritma Support Vector Regression (SVR) dengan pendekatan seleksi fitur Forward Selection.

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 Politeknik Negeri Jakarta

BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisis Kebutuhan

Berdasarkan hasil studi literatur atau tinjauan pustaka, telah diidentifikasi sejumlah kebutuhan penting dalam penelitian ini. Kebutuhan tersebut mencakup pengembangan model, pengembangan web, serta pemenuhan perangkat keras dan perangkat lunak yang diperlukan. Seluruh aspek ini harus dipenuhi guna mendukung kelancaran proses penelitian dan memastikan tercapainya hasil yang optimal.

4.1.1 Kebutuhan Pengembangan Model

Aspek kebutuhan pengembangan model diawali dengan penjelasan mengenai dataset yang digunakan dalam penelitian ini, termasuk di dalamnya adalah dataset hasil uji laboratorium yang bertujuan untuk menganalisis kandungan madu lebah tanpa sengat.

4.1.1.1 Kebutuhan Dataset

Pembuatan model yang akurat dan andal memerlukan dataset yang lengkap dan representatif, serta mencakup seluruh variabel yang relevan.. Untuk mengurangi kesalahan, sumber data juga harus terpercaya dan berkualitas tinggi. Sebelum bisa digunakan untuk melatih dan menguji model, dataset perlu diolah terlebih dahulu. Proses ini mencakup pembersihan data dari error dan penyesuaian format agar seragam.

Tabel 2. Kebutuhan Dataset

Kode	Label	Deskripsi
KD-1	Nama Daerah	Daerah Lampung dan Rangkasbitung
KD-2	Data Spektrum	Nilai serapan cahaya pada panjang gelombang 357-725.5 nm
KD-3	Hasil Uji lab	Terdiri dari Hydroxymethylfurfural, Reducing Sugar, Saccharose.
KD-4	Jumlah Data	Jumlah data pada dataset sebanyak 200 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Tabel 2 menunjukkan kriteria detail dataset yang digunakan untuk memprediksi kandungan gula pada madu menggunakan metode *Support Vector Regression* dan teknik *forward Selection*. Dataset yang digunakan di penelitian ini mempunyai 200 data. Dataset yang digunakan ini diambil dari hasil pengujian menggunakan alat spektroskopi pada daerah Lampung dan Rangkasbitung dengan masing-masing daerah memiliki 100 data. Pada penelitian ini akan dipilih data spektrum serapan cahaya dan data kandungan gula sebagai data tes dan data latih dalam pengembangan model. Pengembangan model memanfaatkan data spektrum serapan cahaya sebagai variabel *input*, dengan data kandungan gula sebagai variabel *output*.

4.1.1.2 Kebutuhan Model

Tabel 3. Kebutuhan Model

Kode	Label	Deskripsi
KM-1	Prediksi	Model mampu melakukan prediksi dengan kandungan gula berdasarkan variabel input
KM-2	Performa	Model memiliki akurasi yang cukup baik
KM-3	Visualisasi	Model dilengkapi dengan tampilan visual yang jelas dan informatif.

Tabel 3 menunjukkan kebutuhan model pada penelitian ini yang bertujuan untuk melakukan prediksi kandungan gula pada madu. Dengan terpenuhinya kebutuhan ini, model akan layak digunakan.

4.1.2 Kebutuhan Pengembangan Web

Dua aspek utama, yaitu kebutuhan fungsional dan non-fungsional, menjadi pertimbangan penting dalam pengembangan web pada penelitian ini untuk memastikan sistem berjalan secara optimal. Kebutuhan fungsional mencakup proses yang disediakan system, sedangkan kebutuhan non-fungsional mencakup aspek seperti kecepatan, keandalan, dan kemudahan penggunaan sistem. Aplikasi web ini dirancang sebagai sarana implementasi model regresi yang telah dikembangkan, sehingga pengguna dapat memanfaatkan hasil prediksi secara langsung melalui *web* yang interaktif.



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 Politeknik Negeri Jakarta

4.1.2.1 Kebutuhan Fungsional

Penelitian ini mengembangkan web yang bertujuan untuk menguji model yang telah dibuat sambil tetap memenuhi kebutuhan fungsional dasar seperti kemampuan untuk mengunggah, mengunduh, dan melihat data dengan tampilan yang sederhana.

Tabel 4. Kebutuhan Fungsional

Kode	Nama	Deskripsi
KF-1	Unggah Data	Pengguna dapat mengunggah data
KF-2	Unduh Data	Pengguna dapat mengunduh data
KF-3	Prediksi Data	Pengguna dapat melakukan prediksi pada data yang diunggah
KF-4	Visualisasi Data	Pengguna dapat melihat visualisasi pada data yang diprediksi

Kebutuhan fungsional pada tabel 4 diharapkan mempermudah pengguna melakukan prediksi kandungan gula pada madu serta memvisualisasikan hasil.

4.1.2.2 Kebutuhan Non-Fungsional

Kebutuhan non-fungsional tidak kalah penting dengan kebutuhan fungsional karena kebutuhan ini menentukan kualitas, fitur, dan batasan sistem. Kebutuhan non-fungsional dapat dilihat pada tabel 5 berikut.

Tabel 5. Kebutuhan Non-Fungsional

Kode	Nama	Deskripsi
KNF-1	Kompabilitas	Sistem dapat diakses dengan mudah pada berbagai <i>browser</i>
KNF-2	Aksesibilitas	Memiliki navigasi dan layout yang mudah digunakan
KNF-3	<i>Library</i>	Sistem menggunakan <i>library</i> seperti <i>scikit-learn</i> , <i>pandas</i> , dan <i>numpy</i> untuk kebutuhan pemodelan.
KNF-4	Arsitektur	Sistem dikembangkan menggunakan Python dengan antarmuka berbasis Streamlit dan <i>backend</i> menggunakan FastAPI.

4.1.3 Kebutuhan Perangkat Keras dan Perangkat Lunak

Pengembangan model dan aplikasi web yang optimal membutuhkan perangkat keras dan perangkat lunak dengan spesifikasi yang memadai. Pada tabel 6



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 Politeknik Negeri Jakarta

menampilkan perangkat keras yang digunakan dalam melakukan perancangan sampai dengan pengujian.

Tabel 6. Kebutuhan Perangkat Keras

Kode	Nama	Deskripsi
KPK-1	CPU	Intel CORE i7
KPK-2	RAM	8 GB
KPK-3	Operating Windows	Windows 11

Tabel 7 dijelaskan kebutuhan perangkat lunak yang digunakan untuk mengembangkan model dan melakukan implementasi web.

Tabel 7. Tabel Perangkat Lunak

Kode	Nama	Deskripsi
KPL-1	Visual Studio Code	<i>Code Editor</i>
KPL-2	Google Collab	Pembuatan model

4.2 Perancangan Sistem

Setelah mengidentifikasi kebutuhan, langkah berikutnya adalah merancang sistem yang memenuhi kebutuhan tersebut. Perancangan ini dimaksudkan untuk berfungsi sebagai panduan dalam proses pengembangan dan implementasi sistem. Penelitian ini membagi perancangan sistem menjadi perancangan model dan perancangan web.

4.2.1 Perancangan Model

Dalam perancangan model, diaplikasikan metode CRISP-DM yang dimulai dari tahap pengumpulan dan penyiapan data yang dibutuhkan untuk proses pelatihan. Setelah data siap digunakan, dilakukan pengembangan model yang kemudian dilatih dan diuji untuk memastikan kemampuannya dalam menghasilkan output yang sesuai dengan tujuan penelitian.

4.2.1.1 Pengumpulan Dataset

Penelitian ini menggunakan kumpulan data yang diperoleh dari hasil analisis laboratorium untuk menemukan kadar gula dan kandungan lainnya dalam madu lebah tanpa sengat. Ini dilakukan dengan menggunakan nilai absorbansi cahaya dari



© Hak Cipta milik 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

data spektrum. Alat spektrofotometer digunakan untuk menghasilkan spektrum dengan mengukur intensitas serapan pada panjang gelombang yang berbeda. Proses pengumpulan data merupakan tantangan khusus karena data semacam ini tidak tersedia secara luas di *repository* publik dan proses pengujian laboratorium membutuhkan sumber daya yang besar. Oleh karena itu, untuk mendapatkan data yang relevan, penelitian ini membutuhkan upaya tambahan, termasuk bekerja sama dengan pihak tertentu. Selanjutnya, data yang dikumpulkan digunakan dalam proses pelatihan dan pengujian model.

x	WaterContent	Diastase	Hydroxymethylfurfural	Reducing Sugar	Saccharose	Acidity	357	357,5	358	
Rangkas	30,29	0	5,49	67,15	69,83	21,59	14,2	14	13,6	
Rangkas	29,68	0	5,82	67,53	69,56	21,45	16,1	15,9	15,1	
Rangkas							18,1	18,1	16,7	
Rangkas							19,4	18,6	18	
Rangkas							21	20	19,6	
x	WaterContent	Diastase	Hydroxymethylfurfural	Reducing Sugar	Saccharose	Acidity	357	357,5	358	358,5
Lampung	30,1	0	3,24	67,46	69,7	18,27	24,9	24,2	24,9	27,7
Lampung	29,69	0	3,15	67,22	69,02	18,92	32,5	32,6	33,1	36,7
Lampung							34,7	35,5	35,4	39,6
Lampung							37	37	36,9	41,6
Lampung							38,8	38,5	38,9	43,1
Lampung							39,3	39,1	39	43,8
Lampung							41,9	41,6	41,4	46,1
Lampung							42,6	42,5	42,6	47,7
Lampung							42,4	42,9	42,6	47,8
Lampung							43,4	43,1	42,9	47,5
Lampung							41,6	41,7	41,1	46

Gambar 4. 1 Dataset daerah Lampung dan Rangkasbitung

Gambar 4.1 menunjukkan data yang digunakan dalam penelitian ini berisi data hasil pengukuran spektrum dari dua lokasi, yaitu Rangkasbitung dan Lampung. Data ini memiliki struktur tabular, di mana:

- Setiap baris merepresentasikan hasil satu sampel madu
- Kolom setelah *acidity* merepresentasikan panjang gelombang cahaya (satuan nanometer) dari 357 hingga 750 nm, dengan interval sebesar 0,5 nm
- Kolom Reducing sugar digunakan sebagai variabel target dalam pemodelan, sementara kolom-kolom spektrum (dari 357.0 hingga 750.0) berfungsi sebagai fitur.

Kolom-kolom berikut adalah parameter madu yang ada pada dataset dalam 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

- *Water Content* (Kadar Air): Persentase air dalam madu; indikator stabilitas dan risiko fermentasi.
- *Diastase*: Aktivitas enzim alami; indikator kesegaran madu dan perlakuan panas.
- *Hydroxymethylfurfural* (HMF): Senyawa yang terbentuk akibat panas/penuaan; indikator kerusakan kualitas madu.
- *Reducing Sugar* (Gula Pereduksi): Total fruktosa dan glukosa; indikator kemurnian dan kematangan madu. Ini adalah variabel target utama yang diprediksi.
- *Saccharose* (Sukrosa): Gula disakarida; kadar tinggi mengindikasikan pemalsuan.
- *Acidity* (Keasaman): Tingkat keasaman madu; memengaruhi rasa dan indikator potensi fermentasi.

4.2.1.2 Pelatihan Model

Penelitian ini menggunakan algoritma *Support Vector Regression* (SVR) dengan kernel *Radial Basis Function* (RBF) pada tahap pelatihan model. Keputusan memilih model dengan kernel-nya didasarkan pada kemampuan untuk menangani hubungan non-linear yang kompleks, yang sering terjadi pada data berbasis spektrum.

Teknik seleksi *forward selection* digunakan agar model hanya menggunakan fitur yang paling penting. Dengan menggunakan metode ini, fitur ditambahkan secara bertahap berdasarkan kontribusinya terhadap performa model. Dengan cara ini, kompleksitas model dapat dikendalikan dan risiko overfitting dapat diminimalkan, terutama dalam situasi di mana jumlah fitur sangat besar, seperti pada data spektral. Selama proses validasi model, pengujian dilakukan dengan beberapa rasio pembagian data latih dan data uji, yaitu 90:10, 80:20, 70:30, dan 60:40. Rasio-rasio ini digunakan untuk menilai konsistensi dan kinerja model terhadap berbagai skenario distribusi data. Dengan melakukan evaluasi pada berbagai rasio, dapat dianalisis bagaimana perubahan proporsi data memengaruhi akurasi model.

Evaluasi dilakukan menggunakan metrik performa seperti koefisien determinasi (R^2), *root mean squared error* (RMSE), *mean squared error* (MSE), *mean absolute*



Hak Cipta :

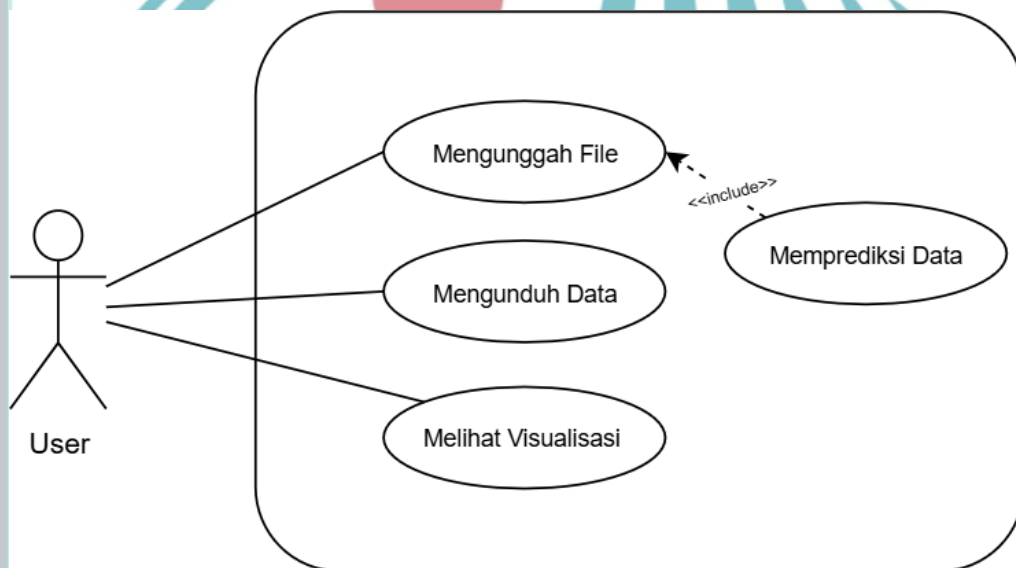
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penulisan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

error (MAE), dan *mean absolute percentage error* (MAPE), untuk memperoleh gambaran yang komprehensif mengenai kemampuan model. Analisis ini membantu dalam menentukan rasio optimal yang menghasilkan performa terbaik pada model.

4.2.2 Perancangan Web

4.2.2.1 Use Case Diagram

Use case diagram dibuat dan dikembangkan berdasarkan kebutuhan fungsional yang sudah dijelaskan sebelumnya pada tabel 4. diagram *use case* menjelaskan interaksi yang bisa dilakukan antara aktor dan sistem.



Gambar 4. 2 Use Case Diagram

Ditunjukkan pada gambar 4.2 menunjukkan *use case diagram* web untuk penelitian ini. Peran aktor dalam web ini mengunggah *file* yang nanti secara otomatis akan melakukan prediksi oleh sistem, setelah itu hasil data yang diprediksi dapat diunduh dan dilihat visualisasinya.

4.2.2.2 Activity Diagram

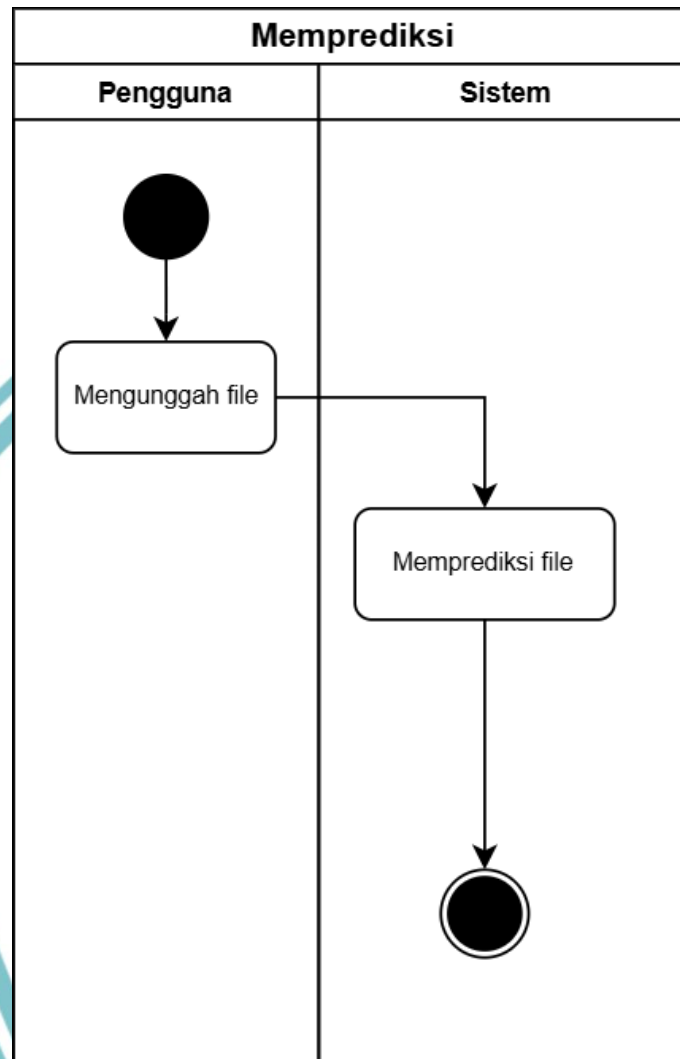
Activity diagram digunakan untuk menggambarkan urutan aktivitas yang terjadi dalam sistem, baik dari sisi pengguna maupun sistem itu sendiri. Diagram ini menyajikan alur proses secara visual dari awal hingga akhir, mulai dari tindakan pengguna seperti mengunggah file hingga sistem melakukan prediksi 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

menampilkan hasil. Hal ini memudahkan proses analisis, pengembangan, dan validasi sistem.



Gambar 4.3 Activity Diagram Mengunggah dan Prediksi file

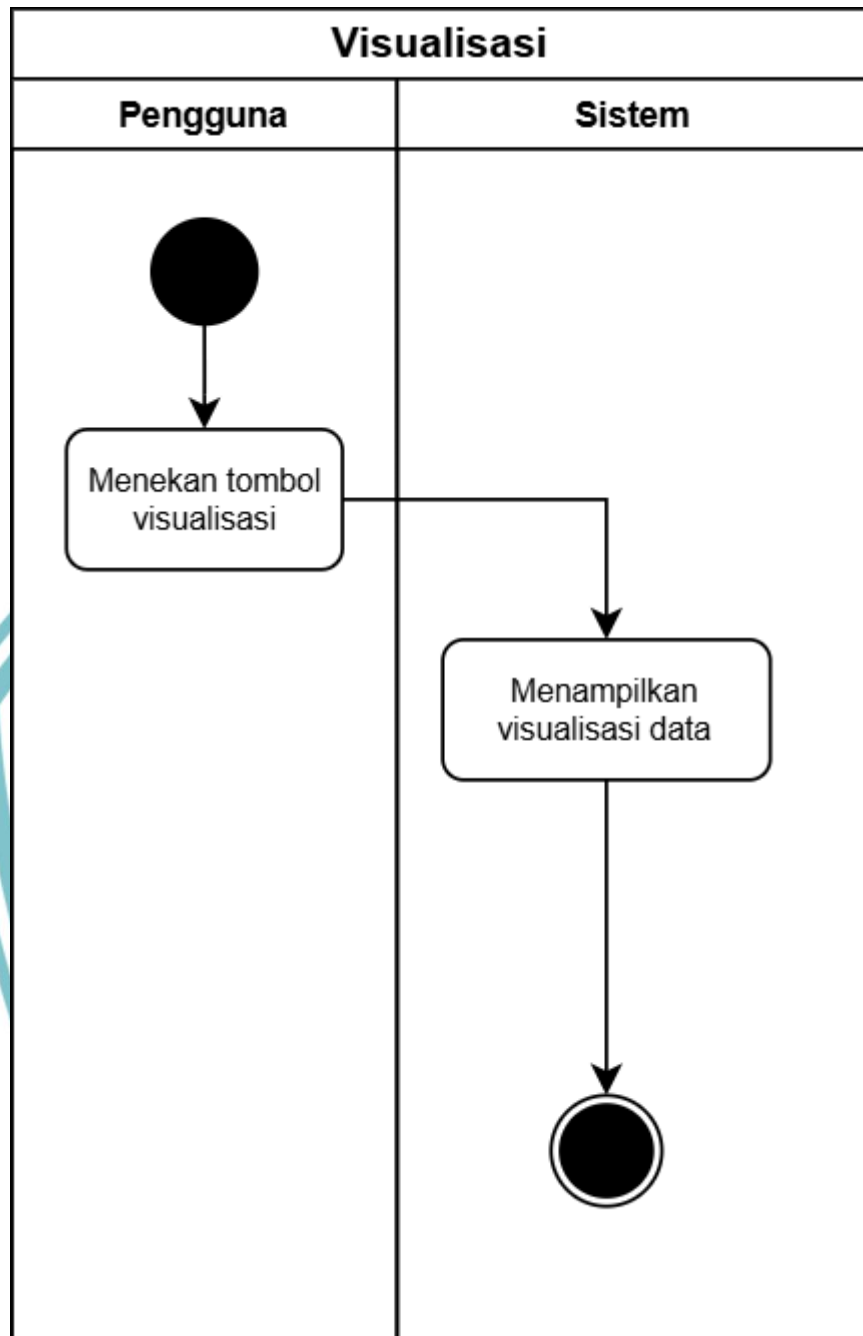
Ditunjukkan pada Gambar 4.3 *activity diagram* saat pengguna ingin melakukan prediksi. Pengguna hanya perlu mengunggah *file* yang ingin diprediksi oleh sistem, setelah itu sistem secara otomatis akan melakukan prediksi. Hasil prediksi dan data yang telah diunggah akan ditampilkan secara bersamaan.



© Hak Cipta milik 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta



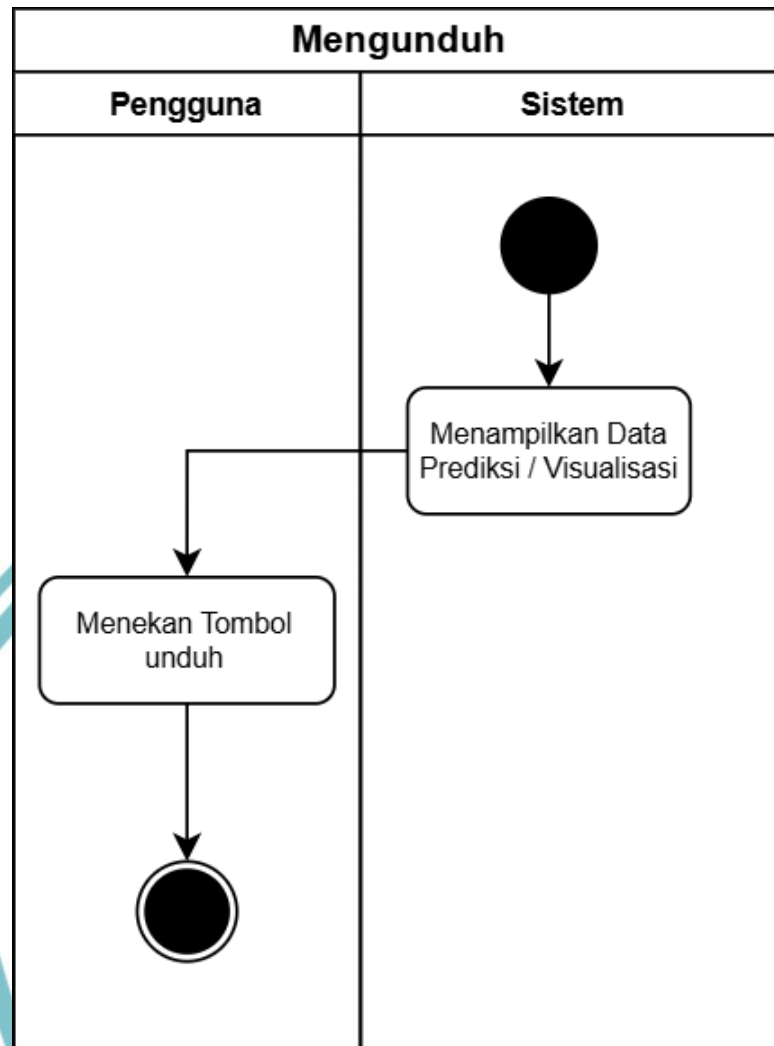
Gambar 4. 4 Activity Diagram Visualisasi Data

Ditunjukkan pada Gambar 4.4 *activity diagram* saat ingin melihat hasil visualisasi dari data yang diprediksi oleh sistem. Pengguna perlu menekan tombol visualisasi untuk agar bisa melihat hasil visualisasinya.



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 Politeknik Negeri Jakarta



Gambar 4. 5 Activity Diagram Mengunduh Data

Ditunjukkan pada gambar 4.5 menunjukkan Activity Diagram untuk mengunduh file yang sudah diprediksi dengan menekan tombol yang ada pada tabel. File yang diunduh akan disimpan dalam *folder* local dalam format CSV.

4.3 Implementasi Sistem

Tahapan implementasi dalam penelitian ini difokuskan pada penerapan model dan sistem yang telah dirancang. Pertama, model regresi yang digunakan untuk memprediksi kandungan gula madu berdasarkan data spektrum diterapkan. Selanjutnya, model tersebut diintegrasikan ke dalam sebuah platform berbasis web, yang memungkinkan pengguna mengunggah data, mendapatkan hasil prediksi secara otomatis, dan menggunakan fitur untuk melihat dan mengunduh 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

4.3.1 Implementasi Model Regresi

- Business Understanding

Model regresi yang dikembangkan bertujuan untuk memprediksi kandungan gula madu lebah tanpa sengat dengan memanfaatkan informasi dari data spektrum cahaya pada rentang panjang gelombang 357 hingga 725,5 nm. Dalam penelitian ini, digunakan metode Support Vector Regression (SVR), yang dikenal mampu menangani pola data yang rumit dan tidak linear. Pemilihan SVR didasarkan pada kemampuannya dalam menghasilkan prediksi yang stabil dan akurat meskipun dengan jumlah data aktual yang terbatas. Agar model tetap efisien tanpa kehilangan akurasi, fitur-fitur yang paling penting dipilih secara selektif sebelum digunakan dalam pelatihan model karena keterbatasan data aktual dan jumlah fitur spektral yang sangat besar. Model ini dirancang agar dapat memberikan estimasi kandungan gula pada madu secara optimal berdasarkan pola spektral yang tersedia.

- Data Understanding

Penelitian ini menggunakan dataset spektrum UV-Vis yang diperoleh dari pengukuran sampel madu menggunakan alat spektrofotometer portabel. Pengukuran dilakukan dalam rentang panjang gelombang 357 hingga 725.5 nm untuk merekam nilai serapan cahaya dari masing-masing sampel. Secara keseluruhan, terdapat 200 sampel yang digunakan, masing-masing terdiri dari 100 sampel madu yang berasal dari dua wilayah berbeda, yaitu Lampung dan Rangkasbitung. Setiap spektrum yang dihasilkan merepresentasikan karakteristik optik madu yang menjadi dasar untuk proses analisis dan prediksi kandungan gula.



© Hak Cipta milik 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 Politeknik Negeri Jakarta

```
[26] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Columns: 745 entries, x to 725.5
dtypes: float64(744), object(1)
memory usage: 1.1+ MB

df.shape

(200, 745)
```

Gambar 4. 6 Informasi Dataset

Gambar 4.6 menunjukkan bahwa dataset memiliki 200 sampel data dan 745 kolom yang terdiri dari 744 kolom bertipe numerik dan 1 kolom label.

```
df.isnull().sum()
```

	0
x	0
WaterContent	196
Diastase	196
Hydroxymethylfurfural	196
Reducing Sugar	196
...	...
723.5	0
724	0
724.5	0
725	0
725.5	0

745 rows × 1 columns

Gambar 4. 7 Informasi missing values

Gambar 4.7 menunjukkan bahwa dari 200 sampel di dataset ini, terdapat 196 data yang tidak memiliki nilai (*missing values*), tetapi setiap data spektrum memiliki data yang lengkap dan dapat digunakan untuk pembuatan model.



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 Politeknik Negeri Jakarta

	x	WaterContent	Diastase	Hydroxymethylfurfural	Reducing Sugar	Saccharose	Acidity	357	357.5	358
0	Lampung	30.10	0.0	3.24	67.46	69.70	18.27	24.9	24.2	24.9
1	Lampung	29.69	0.0	3.15	67.22	69.02	18.92	32.5	32.6	33.1
100	Rangkas	30.29	0.0	5.49	67.15	69.83	21.59	14.2	14.0	13.6
101	Rangkas	29.68	0.0	5.82	67.53	69.56	21.45	16.1	15.9	15.1

Gambar 4. 8 Informasi Data

Gambar 4.8 menunjukkan hanya terdapat empat (4) data hasil uji laboratorium yang tersedia, yaitu dua dari daerah Lampung dan dua dari Rangkasbitung. Salah satu kendala utama adalah biaya yang tinggi untuk melakukan pemeriksaan laboratorium. Selain itu, pembuatan model prediksi juga menjadi tantangan karena keterbatasan data. Untuk mengatasi masalah tersebut, digunakan teknik *Ridge Imputation* sebagai metode imputasi atau pengisian nilai yang hilang.

- Data Preparation

Dataset penelitian ini memiliki keterbatasan, yaitu pada sedikitnya jumlah data yang terisi penuh yang dapat mempengaruhi keakuratan dan keoptimalan hasil prediksi model ini. Untuk meningkatkan kualitas dataset dan memungkinkan proses pemodelan dapat dilakukan, digunakan teknik imputasi pada data yang mengandung nilai kosong. Dalam penelitian ini, metode imputasi yang digunakan adalah *Ridge Imputation*.

Metode imputasi regresi *Ridge* (*impute_by_spectrum*) digunakan untuk mengisi data yang hilang. Metode ini memanfaatkan hubungan antara variabel *Reducing Sugar* yang tidak hilang dengan variabel-variabel spektrum yang lengkap. Cara kerjanya adalah sebagai berikut:

- Model regresi Ridge dilatih menggunakan data *Reducing Sugar* yang tersedia (4 sampel) sebagai variabel target dan data spektrum sebagai variabel prediktor.
- Setelah model terlatih, model tersebut kemudian digunakan untuk memprediksi nilai *Reducing Sugar* pada 196 sampel yang memiliki nilai hilang,



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 Politeknik Negeri Jakarta

	x	WaterContent	Diastase	Hydroxymethylfurfural	Reducing Sugar	Saccharose	Acidity	357
0	Lampung	30.100000	0.0	3.240000	67.460000	69.700000	18.270000	24.9
1	Lampung	29.690000	0.0	3.150000	67.220000	69.020000	18.920000	32.5
2	Lampung	29.434129	0.0	3.171630	67.218447	68.743290	19.084296	34.7
3	Lampung	29.380330	0.0	3.118426	67.136102	68.601983	19.234625	37.0
100	Rangkas	30.290000	0.0	5.490000	67.150000	69.830000	21.590000	14.2
101	Rangkas	29.680000	0.0	5.820000	67.530000	69.560000	21.450000	16.1
102	Rangkas	29.904486	0.0	5.598811	67.223550	69.492202	21.757941	18.1
103	Rangkas	29.745797	0.0	5.658430	67.261851	69.362174	21.826831	19.4

Gambar 4. 9 Dataset setelah *Ridge Imputation*

Gambar 4.9 menunjukkan bahwa dataset yang memiliki nilai kosong, sudah terisi dengan teknik *Ridge Imputation*. Dengan teknik ini, nilai-nilai yang hilang diperkirakan dan diisi berdasarkan pola hubungan antara fitur-fitur spektral yang tersedia, sehingga menghasilkan dataset yang lebih lengkap dan siap untuk digunakan dalam tahap pelatihan model prediksi.

• Modelling

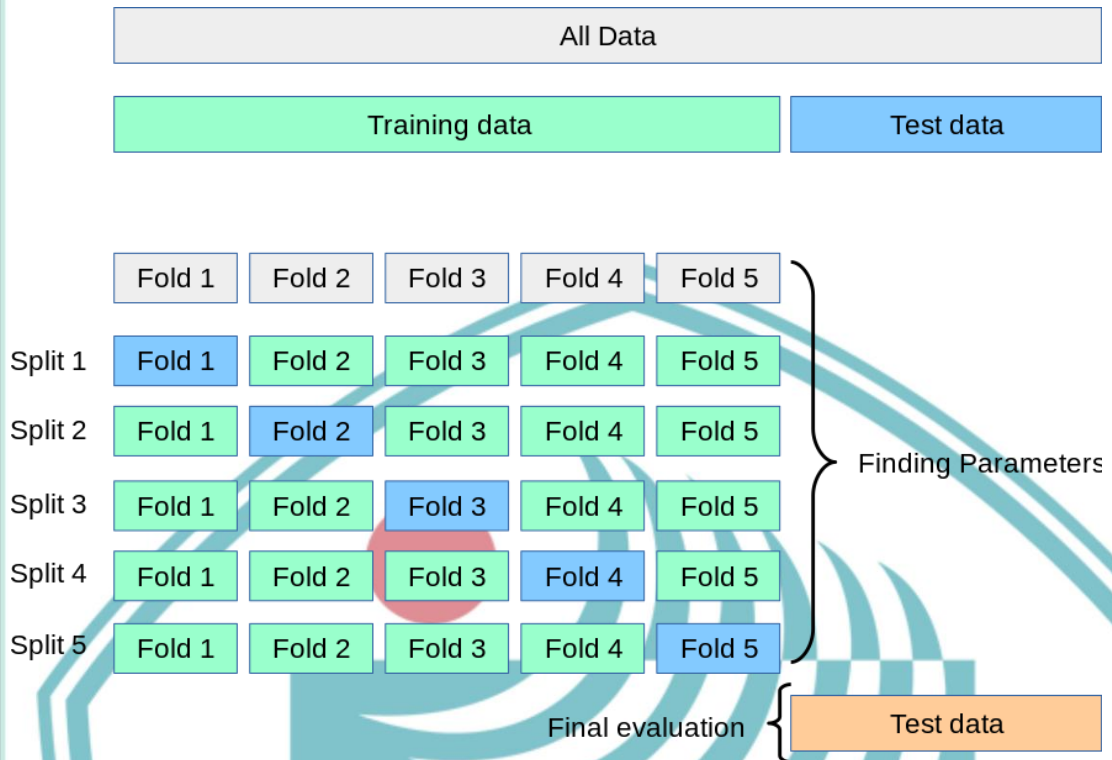
Langkah selanjutnya adalah masuk ke tahap modelling, yang dapat dilakukan setelah proses *pre-processing*. Pada bagian ini, tujuan utamanya adalah membangun sebuah model prediktif dengan memanfaatkan dataset yang telah melalui tahap pembersihan dan persiapan. Algoritma yang dipilih dalam pembuatan model ini adalah *Support Vector Regression (SVR)*, karena dianggap mampu menangani data yang memiliki pola kompleks.

Tahapan pengujian *modelling* dilakukan dengan cara membagi dataset menjadi data pelatihan dan data pengujian, lalu dilanjutkan dengan penerapan *forward selection* untuk memilih fitur yang paling berpengaruh terhadap hasil prediksi. Setelah fitur terbaik diperoleh, Model kemudian dievaluasi menggunakan teknik *k-fold cross-validation* dengan variasi nilai *k*, yaitu 4, 5, 6, 8, dan 10.



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 Politeknik Negeri Jakarta



Gambar 4. 10 K-fold

Sumber: seikit.learn

Gambar 4.10 menunjukkan metodologi evaluasi model yang digunakan dalam penelitian ini, yaitu kombinasi pendekatan *hold-out* dan validasi silang K-Fold. Dataset dibagi menjadi dua bagian utama yaitu data latih untuk proses pembangunan dan optimasi model, serta data uji yang disimpan untuk evaluasi akhir. Data latih kemudian divalidasi menggunakan K-Fold Cross-Validation (dalam hal ini 5-Fold), di mana setiap *fold* secara bergiliran menjadi data validasi.

Pendekatan ini dilakukan untuk menilai konsistensi dan kemampuan model terhadap berbagai pembagian data. Dalam setiap skenario, data dibagi menjadi k subset, di mana model dilatih pada k-n bagian dan diuji pada bagian yang tersisa secara bergiliran. Hasil evaluasi dari masing-masing fold dikombinasikan untuk mendapatkan metrik performa yang lebih representatif, seperti R^2 , RMSE, MAE, dan MAPE. Penggunaan cross-validation ini membantu menghindari ketergantungan model terhadap pembagian data tertentu dan meningkatkan keakuratan dalam memprediksi.



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 Politeknik Negeri Jakarta

Pengujian terhadap berbagai fungsi *kernel* dalam algoritma *Support Vector Regression* (SVR) dilakukan untuk mengetahui jenis *kernel* yang paling sesuai dengan karakteristik data yang dipakai. Pengujian ini dievaluasi berdasarkan empat metrik yaitu R^2 , RMSE, MAE, MAPE.

Tabel 8. Perbandingan Kernel SVR

Kernel	R^2	RMSE	MAE	MAPE
Linear	0.7990	0.0654	0.0562	0.0840
Poly	0.7982	0.0644	0.0536	0.0800
RBF	0.8273	0.0592	0.0454	0.0677

Tabel 8 menunjukkan bahwa fungsi *kernel* RBF (*Radial Basis Function*) memberikan performa terbaik disbanding dengan *kernel* lainnya. *Kernel* RBF memperoleh nilai R^2 tertinggi sebesar 0.8273, disertai RMSE, MAE, dan MAPE yang paling rendah, yang menunjukkan bahwa model dengan *kernel* ini memiliki tingkat kesalahan prediksi yang lebih kecil. Oleh karena itu RBF dipilih dalam proses pemodelan SVR.

```

X = df.iloc[:, 7:] # Data spektral
y = df['Reducing Sugar']
  
```

Gambar 4. 11 Inisiasi Target dan Fitur

Langkah selanjutnya adalah menetapkan variabel X dan variabel Y. Variabel X berisi data spektrum panjang gelombang cahaya dari sampel madu yang digunakan sebagai fitur, sedangkan variabel Y merupakan kandungan reducing sugar yang menjadi target prediksi dalam penelitian ini. Pada gambar 4.11 dijelaskan variabel X memilih semua kolom mulai dari indeks ke-7 hingga akhir, dan variabel Y memilih *reducing sugar* sebagai target.



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 Politeknik Negeri Jakarta

Forward selection digunakan untuk memilih fitur-fitur spektrum yang paling relevan secara bertahap, dengan menambahkan satu per satu fitur yang memberikan peningkatan performa model secara signifikan, proses ini juga bergantung pada estimator yang tepat. *Forward selection* menggunakan *estimator Ridge Regression* karena menyediakan koefisien yang jelas untuk diinterpretasikan selama proses seleksi.

```
# Pipeline
def create_pipeline():
    return Pipeline([
        ('scaler', StandardScaler()),
        ('feature_selector', SequentialFeatureSelector(
            estimator=Ridge(alpha=1.0, random_state=42),
            n_features_to_select='auto',
            direction='forward',
            scoring='r2',
            tol=0.001,
            cv=3,
            n_jobs=-1
        )),
        ('model', SVR(kernel='rbf', C=1.0, epsilon=0.1))
    ])
```

Gambar 4. 12 Pipeline

Pembuatan pipeline pada gambar 4.12 dilakukan untuk menggabungkan beberapa tahapan pemrosesan data dan pelatihan model ke dalam satu alur yang sistematis. Pipeline terdiri dari tiga langkah utama, yaitu standarisasi data, seleksi fitur menggunakan metode *forward selection*, dan pemodelan menggunakan *Support Vector Regression* (SVR) dengan *kernel* RBF. Model SVR dilatih untuk melakukan prediksi kandungan *reducing sugar*. Dengan pendekatan ini, proses seleksi dan pelatihan dilakukan membantu mencegah overfitting dengan hanya menggunakan fitur-fitur yang paling berkontribusi terhadap hasil prediksi.



© Hak Cipta milik 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 Politeknik Negeri Jakarta

```
def evaluate_feature_selection(X, y, fold_list=[4, 5, 6, 8, 10]):
    scoring = {
        'r2': 'r2',
        'rmse': 'neg_root_mean_squared_error',
        'mae': 'neg_mean_absolute_error',
        'mape': mape_scorer
    }

    for k in fold_list:
        print(f"\n Evaluasi {k}-Fold CV : ")
        start_time = time.time()

        cv = KFold(n_splits=k, shuffle=True, random_state=42)
        pipeline = create_pipeline()

        results = cross_validate(
            pipeline, X, y,
            cv=cv,
            scoring=scoring,
            n_jobs=-1,
            return_estimator=True
        )

        # Hasil evaluasi
        r2_mean = np.mean(results['test_r2'])
        rmse_mean = -np.mean(results['test_rmse'])
        mae_mean = -np.mean(results['test_mae'])
        mape_mean = -np.mean(results['test_mape'])

        print(f" {k}-Fold selesai dalam {time.time() - start_time:.2f}s")
        print(f" R2 : {r2_mean:.4f}")
        print(f" RMSE : {rmse_mean:.4f}")
        print(f" MAE : {mae_mean:.4f}")
        print(f" MAPE : {mape_mean:.2f}%")

        # Tampilkan fitur terpilih dari estimator terbaik (berdasarkan R2)
        best_model_idx = np.argmax(results['test_r2'])
        selected_features = X.columns[results['estimator'][best_model_idx]
                                   .named_steps['selector'].get_support()]
        print(f" Fitur terpilih: {selected_features.tolist()[:10]} ...")
```

Gambar 4. 13 K-fold cross validation

Proses modelling menggunakan dengan k-fold cross validation sebagai metrik evaluasinya. Prosedur K-Fold tersebut diterapkan dengan variasi nilai k (4, 5, 6, 8, dan 10) untuk menguji konsistensi performa model terhadap skema pembagian data yang berbeda. K-fold cross validation untuk SVR *kernel* RBF dengan metode *forward selection* ditunjukkan pada gambar 4.13 diatas.



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 Politeknik Negeri Jakarta

Tabel 9. SVR dengan Forward Selection

Fold	R ²	RMSE	MAE	MAPE	Fitur Terpilih (Frekuensi)
4	0.8417	0.0578	0.0432	0.064	435.5 (3), 436 (4), 520 (2), 520.5 (2), 522 (3)
5	0.8395	0.0571	0.0431	0.064	435.5 (4), 436 (4), 520 (3), 520.5 (2), 522 (4)
6	0.8697	0.0519	0.0398	0.059	435.5 (6), 436 (6), 520 (4), 520.5 (3), 522 (6)
8	0.8618	0.0527	0.0415	0.061	435.5 (7), 436 (7), 520 (5), 520.5 (2), 522 (7)
10	0.8508	0.0532	0.0409	0.061	435.5 (9), 436 (9), 520 (7), 520.5 (3), 522 (9)

Tabel 9 menunjukkan perbandingan kinerja model SVR RBF dengan *forward selection* pada berbagai skema *cross validation*. Dari serangkaian pengujian, teridentifikasi bahwa konfigurasi K=6 adalah yang paling optimal. Hal ini dibuktikan dengan pencapaian nilai R² tertinggi dengan 0.8697, RMSE 0.0519 dan MAE 0.0398 dibandingkan dengan jumlah lipatan (K) lainnya. Fitur-fitur yang paling konsisten terpilih selama proses *cross validation* dengan berbagai jumlah fold (4, 5, 6, 8, dan 10) ditunjukkan melalui frekuensi kemunculannya. Lima fitur yang paling sering muncul pada fold ke-6 adalah 435.5, 436, 520, 520.5, dan 522.

Tabel 10. SVR tanpa Forward Selection

Fold	R ²	RMSE	MAE	MAPE
4	0.8362	0.0591	0.0447	0.07%
5	0.8273	0.0592	0.0454	0.07%
6	0.8533	0.0549	0.0427	0.06%
8	0.8478	0.0549	0.0429	0.06%
10	0.8269	0.0570	0.0435	0.06%

Tabel 10 menunjukkan perbandingan kinerja model SVR RBF tanpa *forward selection* pada berbagai skema *cross validation*. Dari serangkaian pengujian, teridentifikasi bahwa konfigurasi K=6 adalah yang paling optimal. Hal ini dibuktikan dengan pencapaian nilai R² tertinggi dengan 0.8533, RMSE 0.0549 dan MAE 0.0427 dibandingkan dengan jumlah lipatan (K) lainnya.



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 Politeknik Negeri Jakarta

- *Evaluation*

Tahap evaluasi dilakukan dengan membandingkan model prediksi SVR dengan *forward selection* dan model tanpa *forward selection*.

Tabel 11. Perbandingan Model

<i>K-Fold</i>	SVR dengan FS			SVR tanpa FS		
	R ²	RMSE	MAE	R ²	RMSE	MAE
4	0.8417	0.0578	0.0432	0.8362	0.0591	0.0447
5	0.8395	0.0571	0.0431	0.8273	0.0592	0.0454
6	0.8697	0.0519	0.0398	0.8533	0.0549	0.0427
8	0.8618	0.0527	0.0415	0.8478	0.0549	0.0429
10	0.8508	0.0532	0.0409	0.8269	0.0570	0.0435

Tabel 11 menunjukkan bahwa model dengan *forward selection* menunjukkan performa yang lebih baik dibandingkan model tanpa seleksi fitur di semua nilai *fold*. Nilai R² cenderung lebih tinggi, sedangkan nilai RMSE, MAE lebih rendah, yang menandakan bahwa model dengan seleksi fitur mampu memberikan prediksi yang lebih akurat dan stabil secara khusus pada *fold* ke-6.

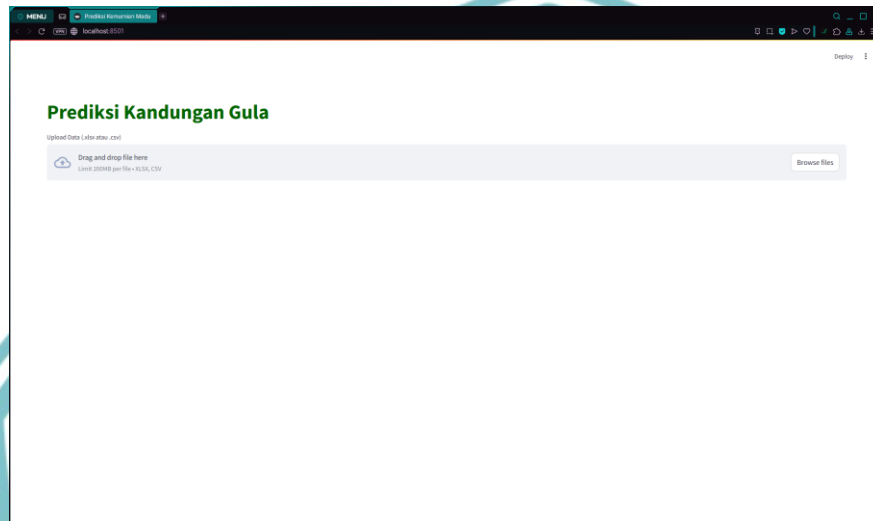
- *Deployment*

Integrasi model regresi ke dalam sebuah aplikasi *web* menjadi inti dari tahap *deployment*. *Web* yang dikembangkan berfungsi sebagai antarmuka bagi pengguna untuk melakukan prediksi kandungan gula pada madu. Mekanismenya melibatkan pengunggahan data spektrum oleh pengguna, yang kemudian diproses secara otomatis oleh model untuk menghasilkan prediksi. Pengguna dapat melihat dan mengunduh data hasil, baik dalam bentuk tabel terstruktur maupun visualisasi grafis, untuk penerapan praktis hasil penelitian.



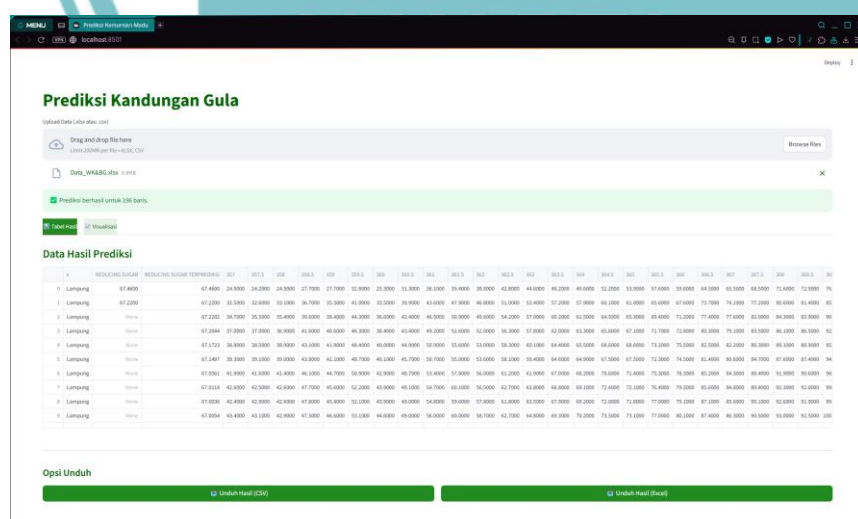
4.3.2 Implementasi Web

Gambar 4.14 menunjukkan halaman yang akan ditemui oleh pengguna pertama kali, di halaman ini hanya terdapat satu tombol yaitu untuk mengunggah data yang ingin diprediksi.



Gambar 4. 14 Halaman Web

Gambar 4.14 menunjukkan tahap setelah pengguna mengunggah data. Data yang diunggah akan secara otomatis diprediksi dan ditampilkan bersamaan dengan data yang diunggah. Output prediksi bisa diunduh dalam format csv atau excel dengan menekan tombol unduh dibawah tabel hasil.



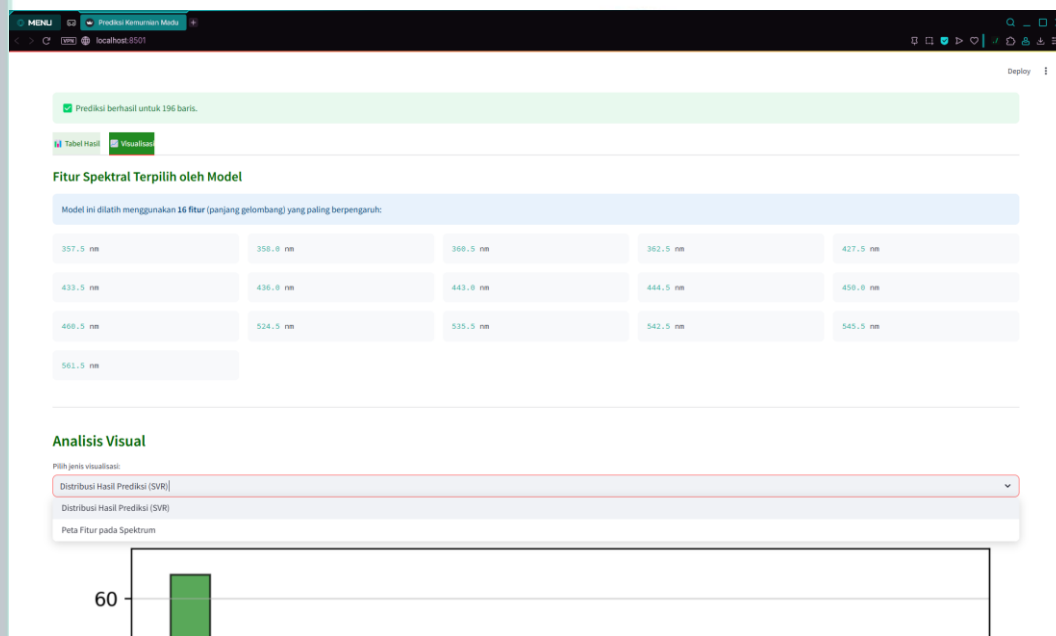
Gambar 4. 15 Web Setelah mengunggah Data



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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Gambar 4.15 menunjukkan halaman visualisasi yang dapat diakses dengan hanya berpindah *tab*, di halaman ini mempunyai tabel fitur terpilih dan gambar analisis visual. Hasil visualisasi juga dapat diunduh sebagai gambar, kecuali tabel fitur fitur yang terpilih.



Gambar 4. 16 Halaman Visualisasi

4.4 Pengujian

Tahapan pengujian dalam penelitian ini berfokus pada evaluasi model prediksi yang dikembangkan untuk memperkirakan kandungan gula pada madu lebah tanpa sengat. Langkah ini bertujuan untuk menilai performa model secara menyeluruh serta mengenali kelemahan yang masih terdapat dalam sistem, sehingga dapat dilakukan penyempurnaan guna meningkatkan akurasi dan ketepatan prediksi pada implementasi selanjutnya.

4.4.1 Deskripsi Pengujian

Pengujian dalam penelitian ini mencakup dua komponen utama. Pertama, dilakukan pengujian black box pada sistem web guna memastikan seluruh fitur berjalan sebagaimana mestinya sesuai dengan rancangan fungsional. Pengujian ini difokuskan pada interaksi pengguna dengan aplikasi, mulai dari mengunggah data hingga proses prediksi dan mengunduh hasil. Kedua, dilakukan evaluasi terhadap



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 Politeknik Negeri Jakarta

performa model prediksi. Evaluasi tersebut mencakup perbandingan antara model yang menggunakan teknik seleksi fitur *Forward Selection* dan yang tidak menggunakannya. Tujuan dari pengujian ini adalah untuk mengetahui pengaruh masing-masing pendekatan terhadap tingkat akurasi model dan kestabilan hasil prediksi.

4.4.1.1 *Blackbox Testing*

Pengujian dilakukan untuk memastikan bahwa seluruh fungsi dalam sistem web prediksi kandungan gula pada madu lebah tanpa sengat berjalan sebagaimana mestinya. Metode black box digunakan karena berfokus pada pengujian perilaku sistem dari sudut pandang pengguna akhir, tanpa memperhatikan struktur kode di baliknya. Setiap skenario dirancang agar mencerminkan tindakan nyata pengguna, seperti mengunggah data, melihat hasil prediksi, mengunduh file, hingga berpindah antar *tab* visualisasi.

Tabel 12. Skenario Blackbox Testing

Kode	Kasus	Skenario	Harapan
BB-1	Mengunggah Data	Pengguna berhasil mengunggah data dengan format .csv atau .xlsx	Sistem menampilkan data yang diunggah
BB-2	Prediksi Data	Pengguna berhasil melakukan prediksi dengan mengunggah data	Data yang telah diunggah akan diprediksi lalu ditampilkan hasil prediksi
BB-3	Mengunduh Data	Pengguna berhasil menekan tombol unduh	Data hasil prediksi, visualisasi dapat tersimpan di folder lokal
BB-4	Melihat Visualisasi Data	Pengguna berhasil berpindah ke <i>tab</i> visualisasi	Hasil visualisasi dapat dilihat oleh pengguna



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 Politeknik Negeri Jakarta

4.4.2 Prosedur Pengujian .

4.4.2.1 Pengujian Mandiri

Pengujian mandiri dilakukan untuk mengevaluasi secara menyeluruh, meliputi pengujian fungsionalitas melalui metode black-box serta pengukuran kinerja model berdasarkan akurasi.

Penilaian terhadap kemampuan model dalam memprediksi nilai sebenarnya dilakukan melalui pengukuran kinerja dengan beberapa metrik evaluasi. Metrik kuantitatif yang digunakan dalam penelitian ini adalah koefisien determinasi (R^2), *root mean squared error* (RMSE), *mean squared error* (MSE), *mean absolute error* (MAE), dan *mean absolute percentage error* (MAPE), rumusnya dijelaskan sebagai berikut:

- R^2 (*Coefficient of Determination*) mengukur seberapa besar proporsi varians dari variabel dependen (target) yang dapat dijelaskan oleh variabel independen melalui model. Semakin mendekati 1, maka semakin baik model dalam menjelaskan data.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (7)$$

$SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ = jumlah kuadrat dari selisih nilai antara aktual dan prediksi.

$SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$ = jumlah kuadrat dari selisih nilai aktual terhadap rata-rata nilai aktual.

- MSE (Mean Squared Error) Rata-rata kuadrat selisih antara nilai aktual dan prediksi. Nilai lebih kecil menunjukkan model lebih akurat.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

$|y_i - \hat{y}_i|$ = Nilai absolut dari selisih antara nilai aktual dan prediksi.



© Hak Cipta milik 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

$\hat{y}_i$ = Nilai prediksi ke-i

y_i = Nilai aktual ke-i

n = Jumlah total observasi

$\sum$ = Penjumlahan dari seluruh observasi

- RMSE (Root Mean Squared Error) Akar kuadrat dari MSE, digunakan agar hasil memiliki satuan yang sama dengan data aslinya

$$RMSE = \sqrt{MSE} \quad (9)$$

- MAE (Mean Absolute Error) Menghitung rata-rata dari selisih absolut antara nilai aktual dan nilai prediksi.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

$|y_i - \hat{y}_i|$ = Nilai absolut dari selisih antara nilai aktual dan prediksi.

- MAPE (Mean Absolute Percentage Error) Mengukur rata-rata kesalahan prediksi dalam bentuk persentase dari nilai aktual.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (11)$$

4.4.3 Data Hasil Pengujian

4.4.3.1 Blackbox Testing

Pengujian fungsionalitas sistem dilakukan menggunakan pendekatan black-box, di mana sebuah fitur dinyatakan berhasil jika output aktual yang dihasilkan sistem sesuai dengan output yang diharapkan. Hasil pengujian pada web prediksi dirangkum pada Tabel 13.



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 Politeknik Negeri Jakarta

Tabel 13. Hasil Blackbox Testing

Kode	Skenario	Aktual	Hasil
BB-1	Pengguna berhasil mengunggah data dengan format .csv atau .xlsx	Sistem menampilkan data yang diunggah	Sesuai
BB-2	Pengguna berhasil melakukan prediksi dengan mengunggah data	Sistem menampilkan data yang diunggah dan diprediksi	Sesuai
BB-3	Pengguna berhasil menekan tombol unduh	Sistem berhasil mengunduh	Sesuai
BB-4	Pengguna berhasil berpindah ke <i>tab</i> visualisasi	Sistem menampilkan data visualisasi	Sesuai

4.4.3.2 Model Performance Testing

Perbandingan performa antara model Support Vector Regression (SVR) dengan dan tanpa penerapan teknik Forward Selection (FS) menggunakan validasi silang K-Fold (K = 4, 5, 6, 8, dan 10) dapat dilihat pada tabel 4.13. Hasil evaluasi menunjukkan bahwa SVR dengan FS memberikan hasil yang lebih baik dibandingkan tanpa FS, khususnya dalam hal nilai R^2 yang lebih tinggi serta nilai RMSE dan MAE yang lebih rendah.

Tabel 14. Perbandingan Model

MODEL	R^2	RMSE	MAE
SVR - FS	0.8697	0.0519	0.0398
SVR	0.8533	0.0549	0.0427

Implementasi model SVR juga dilakukan pada dataset yang berbeda dengan melakukan perbandingan SVR-FS dengan SVR. Dataset ini diambil dari daerah Sukabumi dan Rongkasbitung. Dataset ini memiliki karakteristik yang sama kecuali



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 Politeknik Negeri Jakarta

angka spektrum dan angka kualitas dari sampel madu. Tabel 15 menunjukkan bahwa model dengan SVR-FS mempunyai hasil metrik yang lebih tinggi daripada model SVR biasa. Terbukti bahwa model dengan FS membantu dalam Tingkat akurasi prediksi model.

Tabel 15 Perbandingan Model dataset berbeda

MODEL	R^2	RMSE	MAE
SVR - FS	0.9793	0.0725	0.0595
SVR	0.9714	0.0837	0.0570

Pengujian performa model akhir dilakukan dengan menggunakan data unseen yang diperoleh dari pembagian dataset menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%. Pendekatan ini bertujuan untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya selama proses pelatihan. Hasil prediksi pada data uji kemudian dianalisis menggunakan metrik evaluasi seperti R^2 , RMSE, MAE, dan MAPE untuk menilai akurasi. Hasil evaluasi metrik pada *data unseen* bisa dilihat pada tabel 16

Tabel 16. Evaluasi pada *data unseen*

MODEL	R^2	RMSE	MAE	MAPE
SVR - FS	0.8671	0.0475	0.0380	0.06

4.4.4 Analisis Data Pengujian

Analisis data pengujian dimulai dengan tahap penanganan missing value pada dataset menggunakan metode Ridge Regression Imputation, , sehingga kualitas dan jumlah data yang tersedia meningkat secara substansial tanpa mengorbankan konsistensi data.



© Hak Cipta milik 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin Politeknik Negeri Jakarta

Tahap pemodelan dilakukan menggunakan algoritma Support Vector Regression (SVR) dengan pendekatan feature selection berbasis Ridge Forward Selection. Proses pemilihan fitur ini tidak dilakukan secara manual, tetapi menggunakan strategi otomatis melalui parameter auto, disertai evaluasi melalui validasi silang (cross-validation) untuk memastikan performa model yang stabil dan tidak overfitting.

Penerapan metode Forward Selection (FS) terbukti secara signifikan meningkatkan performa prediktif dari model SVR RBF. Ketika dievaluasi menggunakan skema 6-Fold Cross-Validation yang terbukti paling optimal, model yang dilengkapi dengan FS mampu mencapai nilai rata-rata R^2 sebesar 0.8697, yang lebih tinggi dibandingkan model tanpa FS yang hanya mencapai R^2 0.8533. Peningkatan ini juga tercermin pada RMSE yang lebih rendah (0.0519 berbanding 0.0549) dan MAE yang lebih kecil (0.0398 berbanding 0.0427).

Untuk memvalidasi temuan dari proses K-Fold Cross-Validation, sebuah pengujian akhir dilakukan pada *unseen data* dengan skema 80% data latih dan 20% data uji. Pemilihan rasio ini didasarkan pada skema 6-Fold CV yang sebelumnya terbukti paling optimal. Hasil evaluasi pada data ini menunjukkan performa yang sangat baik, dengan R^2 sebesar 0.8671, RMSE sebesar 0.0475, MAE sebesar 0.0380, dan MAPE hanya 0.06. Hal ini membuktikan bahwa model tidak hanya bekerja baik pada data latih, namun juga pada data uji terpisah. Terdapat keterbatasan cakupan data yang perlu diperhatikan.

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 Politeknik Negeri Jakarta

BAB V

PENUTUP

5.1 KESIMPULAN

Penelitian ini berhasil membangun model prediksi kandungan gula pada madu lebah tanpa sengat menggunakan algoritma Support Vector Regression (SVR) yang dioptimalkan melalui metode seleksi fitur Forward Selection (FS). Evaluasi dilakukan menggunakan berbagai konfigurasi K-Fold Cross-Validation ($K = 4$ hingga 10), dan diperoleh bahwa 6-Fold memberikan hasil paling optimal. Model yang menggunakan FS secara konsisten menunjukkan performa lebih baik dibandingkan model tanpa FS. Pada evaluasi 6-Fold, model dengan FS mencapai R^2 sebesar 0.8697, RMSE 0.0519, dan MAE 0.0398. Hasil ini juga didukung oleh pengujian akhir pada data uji, di mana model mencatatkan R^2 sebesar 0.8671 dan MAPE sebesar 0.06, menandakan kemampuan generalisasi yang baik.

Namun demikian, perlu dicatat bahwa data lengkap hanya tersedia pada empat sampel, sehingga proses imputasi untuk mengisi data yang hilang berpotensi menyebabkan overfitting. Oleh karena itu, hasil ini sebaiknya diinterpretasikan secara hati-hati dan perlu divalidasi lebih lanjut dengan jumlah data riil yang lebih banyak.

5.2 SARAN

Saran untuk penelitian ini antara lain:

1. Penerapan model lain seperti Random Forest, XGBoost, atau Deep Learning juga dapat menjadi opsi pembanding untuk menguji sejauh mana potensi maksimal SVR dalam prediksi
2. Penelitian selanjutnya disarankan untuk memperbanyak jumlah data yang memiliki informasi lengkap, khususnya data yang memuat nilai kandungan gula sebenarnya. Hal ini penting untuk mengurangi risiko overfitting, serta meningkatkan akurasi dan validitas model dalam melakukan prediksi terhadap kandungan gula madu..



DAFTAR PUSTAKA

Abdulazeez Adeshina, B. (2022) 'A Review on FastAPI as a Backend Web Framework', International Journal of Computer Science and Information Technology, 14(3), pp. 1–11. Available at: <https://doi.org/10.5121/ijcsit.2022.14301>.

Alida, F. and Mustikasari, M. (2020) 'Analisis Pengaruh Fungsi Kernel pada Support Vector Machine untuk Klasifikasi Topik Skripsi', Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI), 9(2), pp. 162–171. Available at: <https://doi.org/10.23887/janapati.v9i2.23932>.

A. T. Prasetyo, G. Senoaji, dan M. F. Hidayat, "Inventarisasi hasil hutan bukan kayu lebah tanpa sengat (Stingless Bee) di kawasan Stasiun Percobaan Universitas Bengkulu Tahura Bengkulu Tengah," *Journal of Global Forest and Environmental Science*, vol. 2, no. 3, Desember 2022. E-ISSN: 2809-9346.

Bargam, B. et al. (2024) 'Evaluation of the SVR and Random Forest Models Accuracy for Streamflow Prediction under a Data-Scarce Basin in Morocco', Water, 16(5), 652. Available at: <https://doi.org/10.3390/w16050652>.

Chan, Y. Y. (2021) 'System Development Life Cycle (SDLC) is a conceptual model used in project management that describes the stages involved in an information system development project from an initial feasibility study through maintenance of the completed application'. ResearchGate. Available at: https://www.researchgate.net/figure/System-Development-Life-Cycle-SDLC-is-a-conceptual-model-used-in-project-management_fig1_353974950.

C. Mandang, D. Wuisan, dan J. Mandagi, "Penerapan Metode RAD dalam Merancang Aplikasi Web Proyek PLN UIP Sulbagut," *JOINTER: Journal of Informatics Engineering*, vol. 1, no. 2, pp. –, 27 Des. 2020. [Online]. Tersedia: <https://doi.org/10.53682/jointer.v1i02.18>

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 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 Politeknik Negeri Jakarta

Dash, S., Mohapatra, S., & Nayak, J. (2020). A survey on machine learning in Internet of Things: Algorithms, strategies, and applications. *Internet of Things*, 12, 100314. <https://doi.org/10.1016/j.iot.2020.100314>

Duan, Z. et al. (2025) 'Research on Support Vector Regression Short-Time Traffic Flow Prediction Model for Secondary Roads Based on Associated Road Analysis', *Journal of Advanced Transportation*, 2025, pp. 1–11. Available at: <https://doi.org/10.1155/2025/1569427>.

Ghumatkar, A. and Date, H. (2023) 'Comparative Study of Software Development Life Cycle Models', *NeuroQuantology*, 21(5), pp. 883–890. Available at: <https://doi.org/10.48047/nq.2023.21.5.nq30068>.

Hassoun, A., Regular, A. and Nsor-Atindana, J. (2020) 'Front-face fluorescence of honey of different botanic origin: A case study from Tuscany (Italy)', *Applied Sciences*, 10(5), 1776. Available at: <https://doi.org/10.3390/app10051776>.

I. G. P. Mangku, I. G. B. Udayana, H. A. Hidalgo, A. R. Nicolas, dan M. B. Fresnido, "Development strategy and existing qualities of honeybee in beekeepers group Badung Regency-Bali," **Journal of Agriculture and Crops**, vol. 9, no. 3, pp. 365–371, May 2023. [Online]. Available: <https://doi.org/10.32861/jac.93.365.371>

Kamalov, F., Elnaffar, S., Cherukuri, A. and Jonnalagadda, A. (2021) 'Forward feature selection: empirical analysis', *Computers & Electrical Engineering*, 93, 107238. Available at: <https://doi.org/10.1016/j.compeleceng.2021.107238>.

Khorasani, M. et al. (2022) 'An overview of Streamlit for data science', 2022 12th International Conference on Computer and Knowledge Engineering (ICCKE), pp. 237–242. Available at: <https://doi.org/10.1109/ICCKE57176.2022.9960209>.



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 Politeknik Negeri Jakarta

Lumbanraja, F. et al. (2019) 'IMPLEMENTASI METODE SUPPORT VECTOR MACHINE DALAM PREDIKSI PERSEBARAN DEMAM BERDARAH DI KOTA BANDAR LAMPUNG', Jurnal Komputasi, 7. Available at: <https://doi.org/10.23960/komputasi.v7i2.2426>.

Lumbanraja, F. R., Syahputra, R. and Arisandi, D. (2019) 'Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes dalam Klasifikasi Data Mining', Jurnal Sistem Komputer dan Informatika (JSON), 1(1), pp. 29-34. Available at: <https://doi.org/10.30865/json.v1i1.103>.

M. Shanahan and M. Spivak, "Resin Use by Stingless Bees: A Review," Insects, vol. 12, no. 8, p. 719, Aug. 2021, doi: 10.3390/insects12080719.

Nurdin, J. et al. (2024) 'A Review of Feature Selection Techniques for Machine Learning', Journal of Data Science and Its Applications, 7(1), pp. 1-11. Available at: <https://doi.org/10.21108/jdsa.2024.7.1.1>.

Oğuz, E. and Tilda, R. (2022) 'A Churn Prediction Model in the Telecommunication Sector: A Case Study Based on CRISP-DM', 2022 International Conference on Data Science and Applications (ICONDATA), pp. 1–8. Available at: <https://doi.org/10.1109/ICONDATA56269.2022.9902302>.

Oktavianti, C., Agustin, Y. and Sari, P. (2023) 'Penerapan CRISP-DM untuk Analisis Pola Penjualan Produk Menggunakan Algoritma Apriori', Jurnal Teknologi dan Sistem Informasi (JTSI), 4(2), pp. 315–323.

Parbat, D. and Chakraborty, M. (2020) 'A python based support vector regression model for prediction of COVID19 cases in India', Chaos, Solitons & Fractals, 138, p. 109942. Available at: <https://doi.org/10.1016/j.chaos.2020.109942>.



© Hak Cipta milik 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 Politeknik Negeri Jakarta

Raschka, S., Patterson, J. and Nolet, C. (2020) 'Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence'.

Razavi, S. M. and Esmaeilzadeh Kenari, R. (2023) 'Support vector regression (SVR) method for paddy growth phase modeling using Sentinel-1 imagery data', *Media Statistika*, 16(1), pp. 25-36. Available at: <https://doi.org/10.14710/medstat.16.1.25-36>.

Retnoningsih, E. and Pramudita, R. (2020) 'Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python', *BINA INSANI ICT JOURNAL*, 7(2), p. 156. Available at: <https://doi.org/10.51211/biict.v7i2.1422>.

Sayusti, Tiara & Raffiudin, Rika & Kahono, Sih & Nagir, Teguh. (2020). Stingless bees (Hymenoptera: Apidae) in South and West Sulawesi, Indonesia: morphology, nest structure, and molecular characteristic. *Journal of Apicultural Research*. 60. 143-156. 10.1080/00218839.2020.1816272.

Schröer, C., Kruse, F. and Gómez, J. M. (2021) 'A systematic literature review on applying CRISP-DM process model', *Procedia Computer Science*, 181, pp. 526–534. Available at: <https://doi.org/10.1016/j.procs.2021.01.199>.

Senarath, C. (2021) 'A Comparative Study on Agile and Waterfall Methodologies in Software Development', *Asian Journal of Research in Computer Science*, 10(2), pp. 1–16. Available at: <https://doi.org/10.9734/ajrcos/2021/v10i230236>.

Suhandy, D., Kusumiyati, M., Yulia, M. and Al Riza, D. F. (2024) 'Discrimination between real and fake honey using portable fluorescence spectroscopy and simca', *IOP Conference Series: Earth and Environmental Science*, 1290(1), 012003. Available at: <https://doi.org/10.1088/1755-1315/1290/1/012003>.



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 Politeknik Negeri Jakarta

Wahid, A. (2020) 'Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi', Jurnal Ilmu-ilmu Informatika dan Manajemen STMIK, No. November, pp. 1-5.

Wang, L., Wang, Z. and Liu, A. (2022) 'A review of machine learning-based regression methods for predictive modeling', Journal of Big Data, 9, 1. Available at: <https://doi.org/10.1186/s40537-021-00550-4>.

Winarno, F.G. (2020) PANDUAN ANALISIS KEMURNIAN MADU. Gramedia Pustaka Utama.

Wurijanto, T., Setiawan, H. B. and Tjandrarini, A. B. (2022) 'Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit yang Berisiko Menggunakan Algoritma Support Vector Machine', Jurnal Ilmiah Scroll: Jendela Teknologi Informasi, 10(1), pp. 27-35.

Yong, Y. K., Law, J. X. and Tan, N. H. (2022) 'A review on stingless bee honey: Its properties, prospects and challenges', Food Research International, 158, 111326. Available at: <https://doi.org/10.1016/j.foodres.2022.111326>.

Zhang, F. and O'Donnell, L.J. (2020) 'Chapter 7 - Support vector regression', in A. Mechelli and S. Vieira (eds) Machine Learning. Academic Press, pp. 123–140. Available at: <https://doi.org/10.1016/B978-0-12-815739-8.00007-9>.