

Implementasi Feature Engineering pada Sistem Klasifikasi Penyakit Tanaman Cabai Menggunakan XGBoost

Okta Gabriel Sinsaku Sinaga*, Euis Oktavianti

^{1,2}Program Studi D4 Teknik Informatika, Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta
Depok, Jawa Barat

okta.gabriel.sinsaku.sinaga.tik22@mhsw.pnj.ac.id, gabrieleycm@gmail.com, euis.oktavianti@tik.pnj.ac.id

Chili plants as a major horticultural commodity in Indonesia are highly vulnerable to leaf diseases such as geminivirus, cercospora leaf spot, and leaf curl virus, which are difficult to detect manually due to overlapping visual symptoms and field variations. This research aims to implement a feature engineering framework utilizing HSV color extraction (mean, standard deviation, skewness), GLCM texture parameters (contrast, homogeneity, energy across four orientations), and Canny edge density, followed by Principal Component Analysis (PCA) dimension reduction. The classification engine is powered by XGBoost, deployed for a mobile-based detection application via a hybrid Waterfall SDLC and CRISP-DM framework. Experiments on internal controlled datasets indicate that the hybrid XGBoost model achieves an accuracy of up to 98.75% without PCA. To optimize memory for mobile deployment, a 5-component PCA configuration was selected as the sweet spot, retaining high discriminative capacity with an accuracy of 96.37% and significantly lowering peak memory usage to 0.032 MB with sub-second inference. However, robustness stress testing using an independent out-of-distribution dataset from Siborongborong fields ($n = 178$ images, Cohen's Kappa = 0.9772) triggered a massive performance drop to 24.72% global accuracy due to domain drift. This failure analysis highlights severe feature distortion caused by unstructured environmental noise (soil, mulch reflections) and illumination variance, proving the limitation of handcrafted features and establishing the critical need for deep learning segmentation in future open-field edge applications.

Keywords: xgboost, feature engineering, plant disease, computer vision, domain drift

Tanaman cabai sebagai komoditas hortikultura utama di Indonesia rentan terhadap penyakit daun seperti geminivirus, cercospora leaf spot, dan leaf curl virus yang sulit dideteksi secara manual akibat gejala visual yang mirip dan variasi lingkungan. Penelitian ini bertujuan mengimplementasikan kerangka kerja feature engineering melalui ekstraksi hibrida fitur warna HSV (mean, std, skewness), tekstur GLCM (contrast, homogeneity, energy pada 4 orientasi), dan kerapatan tepi (Canny density), yang diikuti dengan reduksi dimensi Principal Component Analysis (PCA). Siklus pengembangan sistem mengadopsi integrasi metode Waterfall SDLC dan CRISP-DM pipeline untuk Machine Learning. Berdasarkan eksperimen pada dataset internal terkontrol, model XGBoost hibrida mampu menghasilkan akurasi optimal mencapai 98,75% tanpa PCA. Untuk kebutuhan komputasi gawai mobile, konfigurasi PCA 5 Komponen dipilih sebagai titik optimal (sweet spot) karena mampu mempertahankan akurasi sebesar 96,37% sekaligus mengoptimalkan memori puncak (peak memory) secara ekstrem hingga 0,032 MB dengan waktu inferensi <2 detik. Meskipun demikian, uji ketangguhan (stress testing) menggunakan dataset independen luar laboratorium dari lahan pertanian Siborongborong ($n = 178$ citra, Koefisien Kappa = 0,9772) memicu degradasi performa drastis dengan akurasi global ambruk hingga 24,72%. Analisis kegagalan mengonfirmasi terjadinya domain drift berupa distorsi fitur akibat interferensi background noise (tanah, mulsa plastik) dan illumination variance cahaya matahari, yang menegaskan limitasi handcrafted features serta urgensi integrasi modul segmentasi berbasis deep learning untuk aplikasi di lahan terbuka.

Kata kunci: xgboost, feature engineering, penyakit tanaman, komputer visi, domain drift

I. PENDAHULUAN

Tanaman cabai merupakan komoditas hortikultura strategis di Indonesia dengan luas panen cabai rawit dan cabai merah yang menempati peringkat teratas untuk kategori sayuran semusim. Meskipun memiliki nilai ekonomi tinggi,

produktivitas nasional cabai sering kali terhambat oleh serangan berbagai penyakit infeksius, seperti geminivirus, cercospora, dan leaf curl virus yang menyerang struktur daun secara masif. Deteksi dini secara manual di area persawahan sangat sulit dilakukan karena manifestasi gejala visual awal

antar-penyakit cenderung identik dan penuh variasi noise lapangan.

Penelitian terdahulu telah berupaya menyelesaikan masalah ini menggunakan ekstraksi fitur HSV dan GLCM yang diklasifikasikan dengan *Support Vector Machine* (SVM) dan sukses memperoleh tingkat akurasi sebesar 91,3% [4]. Di samping itu, tren arsitektur *Deep Learning* seperti *Convolutional Neural Network* (CNN) banyak diterapkan karena kemampuannya mengekstraksi ciri secara otomatis dengan akurasi tinggi (>97%). Namun, pendekatan berbasis CNN memiliki keterbatasan signifikan (*knowledge gap*), yakni membutuhkan *resource* komputasi tinggi serta ukuran memori runtime yang besar. Ketika model CNN dipaksakan untuk ditanamkan ke dalam gawai mobile petani yang memiliki spesifikasi perangkat keras terbatas, waktu inferensi menjadi tidak efisien dan memicu latensi tinggi.

Untuk mengisi rumpang penelitian (*scientific gap*) tersebut, diperlukan pendekatan rekayasa fitur (*feature engineering*) hibrida yang ringan namun tetap diskriminatif, yang kemudian dipadukan dengan algoritma *ensemble learning* berbasis data tabular yang efisien seperti XGBoost. Algoritma XGBoost terbukti tangguh dalam memodelkan hubungan non-linier dari representasi matriks statistik citra digital. Keunikan penelitian ini terletak pada integrasi rekayasa tiga komponen fitur utama citra daun (warna HSV, tekstur GLCM, dan kerapatan tepi Canny) yang direduksi dimensinya menggunakan *Principal Component Analysis* (PCA) guna membangun model klasifikasi hibrida yang ringan namun sangat akurat untuk kebutuhan platform mobile. Manuskrip ini bertujuan untuk merancang, mengimplementasikan, dan menguji efisiensi sistem rekayasa fitur hibrida berbasis XGBoost yang diintegrasikan ke dalam backend Rest API FastAPI dan aplikasi mobile Flutter.

II. METODOLOGI PENELITIAN

A. Rancangan dan Tahapan Penelitian

Penelitian ini dijalankan dengan mengadopsi model hibrida linear sekuensial Waterfall untuk pengembangan aplikasi perangkat lunak dan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk siklus pengembangan model kecerdasan buatan. Alur pengerjaan dilakukan secara bertahap, dimulai dari studi literatur, pengumpulan dataset dari repositori publik Kaggle dan Roboflow, pra-pemrosesan citra (*resize*

256×256 piksel dan peningkatan kontras adaptif CLAHE), ekstraksi fitur hibrida, reduksi dimensi dengan PCA, pelatihan model XGBoost dan Random Forest, pengujian performa laboratorium, hingga *deployment* sistem ke dalam arsitektur gawai mobile berbasis Flutter.

B. Pipeline Rekayasa Fitur (Feature Engineering)

Materi pokok dalam tahapan ekstraksi data visual daun cabai dibagi ke dalam tiga domain utama:

- *Ruang Warna HSV*: Citra diubah dari RGB ke HSV guna memisahkan komponen intensitas cahaya (*Value*) dengan warna murni (*Hue* dan *Saturation*). Tiga metrik statistik deskriptif (*mean*, *standard deviation*, *skewness*) dihitung untuk setiap kanal piksel, menghasilkan 9 fitur warna.
- *Tekstur GLCM (Gray-Level Co-occurrence Matrix)*: Karakteristik tekstur permukaan daun diekstrak berdasarkan parameter *contrast*, *homogeneity*, dan *energy* pada empat sudut orientasi spasial ($0^\circ, 45^\circ, 90^\circ, 135^\circ$), menghasilkan 5 fitur tekstur utama.
- *Kerapatan Tepi Canny (Canny Edge Density)*: Deteksi tepi biner dihitung menggunakan operator Canny untuk mengukur kerapatan garis kontur (1 fitur geometri) yang merepresentasikan tingkat kerutan fisik akibat infeksi penyakit *leaf curl*.

C. Transformasi Principal Component Analysis (PCA)

Gabungan fitur hibrida menghasilkan vektor data awal berdimensi 15 ($D = 15$). Guna menghindari *curse of dimensionality* dan memotong biaya komputasi runtime, transformasi linear dilakukan melalui standarisasi nilai z-score menggunakan Persamaan (1):

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Keterangan μ melambangkan rata-rata populasi fitur ke- i , dan σ melambangkan standar deviasi kolom ke- i . Matriks data tersentralisasi ini kemudian didekomposisi secara aljabar linear ke dalam bentuk matriks kovarians Σ berukuran 15×15 menggunakan Persamaan (2):

$$\Sigma = \frac{1}{n-1} Z^T Z \quad (2)$$

Nilai Eigen (λ) dan Vektor Eigen (v) ditentukan lewat persamaan karakteristik $|\Sigma - \lambda I| = 0$ guna memproyeksikan fitur ke ruang dimensi baru yang saling ortogonal dengan mempertahankan

ambang batas akumulasi varians yang optimal untuk komputasi gawai mobile.

III. HASIL DAN PEMBAHASAN

A. Implementasi Sistem dan Pipeline REST API

Model klasifikasi yang telah dilatih diekspor ke dalam format berkas biner kompresi serial model_xgboost_cabai.pkl bersama dengan objek standardisasi dan transformasi PCA memakai pustaka joblib. Sisi server backend dibangun menggunakan framework FastAPI yang melayani permintaan inferensi secara asinkron (*on-the-fly feature engineering*). Aliran biner gambar yang dikirimkan oleh gawai petani melalui protokol HTTP POST diubah menjadi matriks array NumPy secara langsung memakai fungsi cv2.imdecode.

Aplikasi mobile dirancang lintas-platform memakai framework Flutter. Ketika gambar daun cabai diambil melalui kamera atau galeri gawai, aplikasi akan mengirimkan berkas tersebut ke server backend via endpoint REST API, dan hasil diagnosis berupa nama label penyakit beserta nilai tingkat keyakinan (*confidence score*) akan disajikan kembali kepada pengguna dalam komponen antarmuka *Card List* dinamis secara seketika (*real-time*).

B. Hasil Pengujian Performa Klasifikasi (Laboratorium)

Pengujian performa dievaluasi menggunakan metode 5-fold cross validation pada lingkungan laboratorium (*in-distribution data*). Skenario pertama menguji kontribusi murni fitur secara parsial (tanpa PCA) menggunakan variasi *test size* untuk mengukur sensitivitas deskriptif fitur warna (HSV), tekstur (GLCM), dan tepi (Canny) terhadap algoritma utama XGBoost dan pembanding Random Forest. Hasil pengujian parsial disajikan pada Tabel I.

TABEL I. HASIL PENGUJIAN PERFORMA KLASIFIKASI PADA SKENARIO FITUR PARSIAL

Algoritma Klasifikasi	Skenario Fitur	Test Size	Fitur	Akurasi	Presisi	Recall	F1-Score	Kappa Score
XGBoost	Parsial HSV (Warna)	0.2	9	0.93	0.93	0.93	0.93	0.90
Random Forest	Parsial HSV (Warna)	0.2	9	0.91	0.91	0.91	0.91	0.88

XGBoost	Parsial GLCM (Tekstur)	0.2	5	0.78	0.78	0.78	0.78	0.70
Random Forest	Parsial GLCM (Tekstur)	0.2	5	0.77	0.77	0.77	0.77	0.69
XGBoost	Parsial Canny (Tepi)	0.2	1	0.47	0.33	0.47	0.36	0.30
Random Forest	Parsial Canny (Tepi)	0.2	1	0.47	0.33	0.47	0.36	0.30

Data eksperimen pada Tabel I menunjukkan bahwa fitur warna (HSV) menjadi deskriptor visual paling dominan bagi kedua arsitektur model dengan capaian akurasi 0.93 pada XGBoost. Sebaliknya, fitur tekstur (GLCM) menghasilkan performa stagnan (0.77-0.78) karena intensitas keabuan (gray-level) kerutan halus daun sehat dan daun sakit sering kali tumpang tindih (*overlapping variance*). Penurunan performa paling radikal terjadi pada skenario parsial Canny Edge Density (akurasi 0.47) karena parameter berdimensi tunggal tidak memiliki variansi memadai jika berdiri sendiri. Namun, fitur Canny dirancang bukan untuk berdiri mandiri, melainkan bertindak sebagai elemen penciri struktural komplementer untuk mempertegas kerutan infeksi Curl Virus saat digabungkan dalam fitur hibrida.

Skenario kedua melibatkan pengujian Fitur Hibrida Utuh (15 Dimensi) yang dikombinasikan dengan reduksi dimensi PCA. Pengujian ini dirancang secara komprehensif untuk menganalisis efisiensi penggunaan sumber daya melalui paradigma *Computing Complexity*—yang mencakup Kompleksitas Ruang (Peak Memory dalam satuan MegaBytes menggunakan pelacakan dinamis *tracemalloc*) dan Kompleksitas Waktu (Train Time). Hasil komparasi performa dan beban komputasi arsitektur pada test size 0.2 disajikan pada Tabel II.

Algoritma	Konfigurasi	Fitur Akhir	Akurasi	F1-Score	Peak Memory (MB)	Train Time (s)
PCA						
XGBoost	Tanpa PCA	15	0.9875	0.9876	0.7493	0.0078
XGBoost	PCA Explained 0.95	9	0.9800	0.9800	0.5882	0.0087

XGBoost	PCA: 5 Komponen	5	0.9637	0.9638	0.4677	0.0087
XGBoost	PCA: 2 Komponen	2	0.7762	0.7750	0.3346	0.0092
Random Forest	Tanpa PCA	15	0.9850	0.9850	4.1653	0.0288
Random Forest	PCA: 5 Komponen	5	0.9637	0.9638	2.4106	0.0180

Berdasarkan analisis *trade-off* multi-perspektif pada Tabel II, **algoritma XGBoost dengan konfigurasi PCA: 5 Komponen** ditetapkan sebagai titik optimal (*sweet spot*) untuk kebutuhan produksi perangkat mobile. Skenario ini berhasil **mereduksi dimensi fitur hibrida sebesar 66,67%** (dari 15 menjadi 5 fitur akhir). Meskipun ruang fitur dipotong sepertiganya, model XGBoost mampu mempertahankan akurasi tinggi sebesar **0.9637 (96,37%)** dengan nilai degradasi yang sangat minimal dari kondisi Tanpa PCA (hanya turun 2,38%).

Ditinjau dari aspek memori puncak (*Space Complexity*), runtime mencatat bahwa algoritma XGBoost mengonsumsi ruang memori operasional yang sangat kecil dan stabil (**0.4677 MB** pada PCA:5), berkali-kali lipat lebih efisien dibandingkan Random Forest yang membutuhkan alokasi memori masif (**2.4106 MB**). Hal ini dipengaruhi oleh faktor arsitektural; Random Forest (*Bagging*) wajib mengalokasikan memori besar secara simultan untuk menampung ratusan pohon keputusan paralel di dalam RAM, sementara XGBoost (*Boosting*) menyusun pohon korektif secara bertahap sehingga meminimalkan beban memori dinamis sistem, menjadikannya sangat ideal untuk komputasi gawai mobile.

C. Pengujian Ketangguhan Lapangan (Robustness Testing) dan Analisis Domain Drift

Untuk memvalidasi ketahanan model terhadap perubahan lingkungan riil (*out-of-distribution data*), dilakukan pengujian *Stress Testing* menggunakan dataset independen luar laboratorium yang diambil langsung dari lahan pertanian terbuka di Kecamatan Siborongborong, Kabupaten Tapanuli Utara dengan total sampel sebanyak $n = 178$ citra daun cabai (*Cercospora*: 41, *Curl Virus*: 57, Sehat: 30, *Geminivirus*: 50). Berdasarkan perbandingan antara pelabelan mandiri peneliti dan justifikasi langsung oleh Petani Senior selaku pakar domain (*domain expert*), diperoleh nilai Koefisien Kappa Cohen sebesar **0,9772** yang

memberikan pembenaran akademis bahwa label referensi memiliki tingkat kesepakatan hampir sempurna (*almost perfect agreement*).

Hasil pengujian ketangguhan menunjukkan terjadinya degradasi performa global yang sangat masif, di mana **Akurasi Global merosot tajam ke angka 24,72%**. Hasil evaluasi kuantitatif sektoral disajikan pada Tabel III.

Kelas Penyakit	Support (Sampel)	Precision	Recall	F1-Score	Akurasi Global
Cercospora Leaf Spot	41	0.29	0.34	0.31	
Curl Virus (Keriting)	57	0.65	0.23	0.34	24.72%
Healthy Leaf (Sehat)	30	0.16	0.47	0.23	
Geminivirus	50	0.16	0.06	0.09	

Secara matematis, angka akurasi global 24,72% ini setara dengan performa model klasifikasi acak (*random guessing*) pada struktur 4 kelas (di mana probabilitas tebakan acak murni adalah 25%). Meskipun demikian, jika ditinjau per kelas, kelas *Curl Virus* mencatatkan nilai *Precision* tertinggi sebesar **0,65**. Hal ini mengindikasikan bahwa rekayasa geometri tepi Canny memiliki ketangguhan parsial yang baik saat model memprediksi penyakit keriting daun, walaupun nilai *Recall*-nya rendah (0,23) karena banyak data daun keriting asli lapangan yang gagal terjaring oleh sistem. Penurunan drastis pada akurasi global ini dianalisis secara kritis sebagai akibat dari fenomena **Domain Drift** dan **Covariate Shift** pada lingkungan yang tidak terstruktur (*unstructured environment*):

1. **Peringatan Keselarasan Skala Data (*StandardScaler Warning*):** Munculnya peringatan sistem (*UserWarning*) pada log pengujian mengonfirmasi adanya ketidaksesuaian orientasi atau pergeseran distribusi nilai biner pada data lapangan baru saat melewati fungsi *StandardScaler*. Akibatnya, koordinat spasial fitur hibrida baru tersebut meleset jauh dari pola batas keputusan (*decision boundary*) yang telah dipelajari oleh pohon keputusan XGBoost selama masa pelatihan internal.
2. **Kerapuhan Fitur Tradisional Terhadap Latar Belakang Kompleks (*Background Noise*):** Algoritma ekstraksi fitur yang diterapkan mengekstrak seluruh matriks citra

secara global tanpa adanya tahap segmentasi objek daun terlebih dahulu. Pada dataset internal laboratorium, latar belakang citra cenderung bersih dan homogen. Namun, pada citra lapangan Siborongborong, latar belakang daun bersifat sangat kompleks—mencakup objek tanah, bayangan tanaman, mulsa plastik pertanian, dan tumpang tindih daun lain. Akibatnya, nilai kerapatan piksel dari fungsi GLCM (Tekstur) dan Canny (Kerapatan Tepi) menangkap informasi acak dari latar belakang tersebut, sehingga fitur hibrida yang dikirim ke model XGBoost menjadi bias.

3. **Kerentanan Ruang Warna HSV Terhadap Variasi Pencahayaannya Alami (*Illumination Variance*):** Ekstraksi fitur warna menggunakan metrik statistik ruang warna HSV terbukti sangat sensitif terhadap intensitas cahaya matahari *outdoor* (pagi, siang, dan sore). Pantulan cahaya matahari langsung (*overexposure*) atau kondisi cuaca berawan (*underexposure*) di lahan pertanian Siborongborong menggeser nilai *Hue* dan *Saturation* piksel daun sejati. Hal inilah yang menyebabkan performa kelas *Geminivirus* sangat rendah (*F1-Score* 0,09), karena model kesulitan membedakan bimen warna kuning penyakit dengan daun sehat yang terpapar pantulan cahaya terang matahari.

IV. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan sistem klasifikasi penyakit tanaman cabai dengan memanfaatkan kerangka rekayasa fitur hibrida (HSV, GLCM, Canny) yang dipadukan dengan algoritma XGBoost. Hasil pengujian di lingkungan laboratorium menunjukkan performa optimal, di mana model XGBoost mampu mencapai akurasi puncaknya sebesar 98,75% tanpa PCA, dan mempertahankan stabilitas performa sebesar 96,37% saat menggunakan konfigurasi reduksi dimensi PCA 5 Komponen. Ditinjau dari Space Complexity, arsitektur XGBoost terbukti sangat efisien dengan kebutuhan memori puncak (peak memory) hanya sebesar 0,4677 MB, jauh lebih ringan dibandingkan Random Forest (2,4106 MB), sehingga sangat ideal untuk integrasi aplikasi mobile berbasis Flutter dengan waktu inferensi sub-detik (<2 detik).

Namun, uji ketangguhan (stress testing) pada dataset riil lapangan Siborongborong mengungkap

limitasi struktural yang besar dari implementasi *handcrafted features*, di mana akurasi global model ambruk hingga 24,72% akibat fenomena Domain Drift, Background Noise latar belakang pertanian terbuka (mulsa dan tanah), serta Illumination Variance cahaya alami matahari.

Untuk pengembangan selanjutnya, disarankan secara kuat bagi penelitian masa depan untuk mengintegrasikan algoritma berbasis Deep Learning khusus untuk fase lokalisasi objek, seperti Object Detection atau Semantic Segmentation (misalnya menggunakan arsitektur YOLO atau U-Net ringan). Modul ini wajib ditanamkan sebagai tahap preprocessing utama sebelum proses ekstraksi fitur hibrida dijalankan untuk mengisolasi dan memotong (cropping) objek daun cabai utama secara ketat dari background noise lapangan, sehingga kalkulasi parameter matriks warna HSV, tekstur GLCM, dan kerapatan tepi Canny yang dihitung oleh model XGBoost murni berasal dari area daun yang terinfeksi tanpa adanya distorsi lingkungan pertanian terbuka.

V. REFERENSI

VI. Jurnal:

- [1] Kiran SM, Chandrappa DN. Plant Leaf Disease Detection Using Efficient Image Processing and Machine Learning Algorithms. *Journal of Robotics and Control (JRC)*. 2023; 4(6): 840–848.
- [2] Kurniawati R, Nurhayati N, Sanjaya A. Implementasi Metode Canny Edge Detection Untuk Ekstraksi Ciri Geometri Pada Penyakit Daun Tanaman Hortikultura. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*. 2021; 8(4): 789–796.
- [3] Noola DA, Basavaraju DR. Corn Leaf Image Classification Based On Machine Learning Techniques For Accurate Leaf Disease Detection. *International Journal Of Electrical And Computer Engineering (IJECE)*. 2022; 12(3): 2945–2953.
- [4] Sembiring F, Wahyuni S. Analisis Komparasi Fitur Tekstur Dan Bentuk Canny Dalam Klasifikasi Penyakit Tanaman Cabai Berbasis Pohon Keputusan. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi)*. 2023; 7(2): 345–352.
- [5] Sari DP, Hermanto T. Analisis Pola Klorosis Daun Akibat Infeksi Virus Mosaik Pada Tanaman Solanaceae. *Jurnal Fitopatologi Indonesia*. 2022; 18(2): 45–52.
- [6] Mashuri AS, Sunyoto A, Kusnawi K. Klasifikasi Penyakit Pada Daun Cabai Menggunakan Arsitektur VGG16. *Journal of Electrical Engineering and Computer (JEECOM)*. 2024; 6(2): 305–313.
- [7] Winiarti S, Pujiyanta A. Identification of Chili Plant Diseases Based on Leaves Using Hyperparameter Optimization Architecture Convolutional Neural Network. *International Journal of Advanced Computer Science and Applications*. 2024; 15(11): 730–738.
- [8] Nurmadinah N, Wajidi F, Arifin N. DETEKSI PENYAKIT DAUN CABAI MENGGUNAKAN KOMBINASI GLCM DENGAN KLASIFIKASI SVM. *Insect (Informatics and Security)*: *Jurnal Teknik Informatika*. 2025; 11(2): 178–189.

- [9] Studer S, Bui TB, Drescher C, Hanuschkin A, Winkler L, Peters S, Müller KR. Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. *Machine Learning and Knowledge Extraction*. 2021; 3(2): 392–413.
- [10] Ayu S, Taufik M, M R, Hasan A, Widodo CJ, Botek M, Huttni LO. Pengukuran keparahan gejala penyakit kuning dan kandungan nitrogen tanaman cabai terinfeksi Virus Gemini berbasis spektral biosensor. *Agrokompleks*. 2025; 25(1): 10–21.
- [11] Sulandari S, Suseno R, Hidayat SH, Harjosudarmo J, Sosromarsono S. Deteksi dan Kajian Kisaran Inang Virus Penyebab Penyakit Daun Keriting Kuning Cabai. *HAYATI Journal of Biosciences*. 2006; 13(1): 1–6.
- [12] Wiridannisaa N, Kasim NN, Prihatin P. Identification of Begomoviruses Causing Yellow Curl Disease in Chilli in West Java. *Jurnal Biologi Tropis*. 2025; 25(3): 4231–4238.

VII. Proceeding:

- [13] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York. 2016: 785–794. 2-11)

VIII. Textbooks:

- [14] Gonzalez RC, Woods RE. *Digital Image Processing*. Edisi Keempat. New York: Pearson. 2017