

**EXPLORATION OF MACHINE LEARNING ALGORITHMS
AND CLASS IMBALANCE HANDLING ON
PLANT DISEASE DETECTION**

LAPORAN NON-REGULER - SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :
ERVIN ADITYA
22.11.5263

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

**EXPLORATION OF MACHINE LEARNING ALGORITHMS
AND CLASS IMBALANCE HANDLING ON
PLANT DISEASE DETECTION**

LAPORAN NON-REGULER - SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

ERVIN ADITYA

22.11.5263

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

JALUR NON-REGULER

**EXPLORATION OF MACHINE LEARNING ALGORITHMS
AND CLASS IMBALANCE HANDLING ON
PLANT DISEASE DETECTION**

yang disusun dan diajukan oleh

Ervin Aditya

22.11.5263

telah disetujui oleh Dosen Pembimbing
pada tanggal 17 Desember 2025

Dosen Pembimbing.



Rumini, S.Kom., M.Kom.
NIK. 190302246

HALAMAN PENGESAHAN

JALUR NON-REGULER

**EXPLORATION OF MACHINE LEARNING ALGORITHMS
AND CLASS IMBALANCE HANDLING ON
PLANT DISEASE DETECTION**

yang disusun dan diajukan oleh

Ervin Aditya
22.11.5263

Telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arifivanto Hadinegoro, S.Kom., M.T
NIK. 190302289

Subektiingsih, M.Kom
NIK. 190302413

Rumini, S.Kom., M.Kom
NIK. 190302246



Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani., M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ervin Aditya

NIM : 22.11.5263

Menyatakan bahwa Laporan dengan judul berikut:

EXPLORATION OF MACHINE LEARNING ALGORITHMS AND CLASS IMBALANCE HANDLING ON PLANT DISEASE DETECTION

Dosen Pembimbing : Rumini, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 17 Desember 2025

Yang Menyatakan,



Ervin Aditya

HALAMAN PERSEMBAHAN

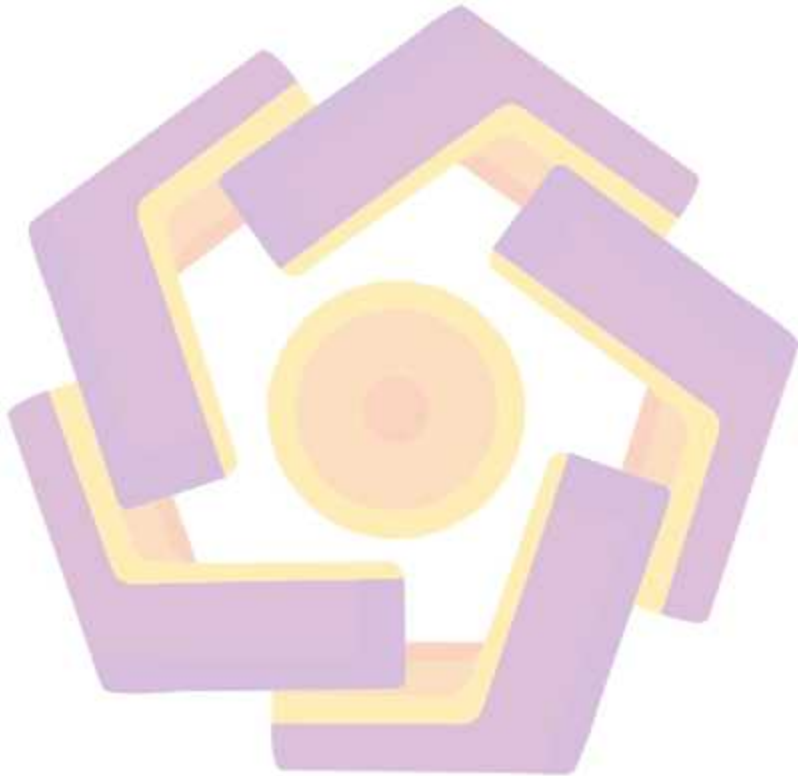
Segala puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala, atas limpahan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini sebagai salah satu syarat untuk memperoleh gelar sarjana. Dengan izin dan ridha-Nya, karya ini dapat terselesaikan setelah melalui proses panjang yang penuh doa, usaha, dan pembelajaran berharga.

Dengan penuh rasa syukur dan kerendahan hati, penulis mempersembahkan karya ini kepada:

1. Dengan segenap rasa syukur, penulis memanjatkan puji dan syukur ke hadirat Allah Subhanahu Wa Ta'ala, atas limpahan rahmat, taufik, serta hidayah-Nya yang telah memberikan kekuatan, kesehatan, dan kemudahan sehingga penulis dapat menyelesaikan penelitian ini dengan baik.
2. Kepada Keluarga saya, khususnya Abang saya Usfri Raranda, yang selalu menjadi sumber kekuatan dan inspirasi dalam kehidupan penulis. Atas segala limpahan kasih sayang yang tak terhingga, doa yang tiada henti, dukungan moril maupun material yang tidak pernah putus, serta pengorbanan dalam mengusahakan segala hal demi menunjang kehidupan dan perjalanan studi penulis.
3. Ibu Rumini, S.Kom, M.Kom, selaku Dosen Pembimbing yang sudah membantu dalam pembuatan laporan penulisan, atas segala bimbingan, arahan, semangat yang diberikan, serta selalu sabar dalam membimbing penulisan laporan hingga terselesaikan penelitian ini.
4. Bapak Ajie Kusuma Wardhana, S.Kom, M.Eng, selaku Dosen yang sudah membantu Riset, atas segala bimbingan, arahan, semangat yang diberikan, serta kesabarannya dalam membimbing penulis hingga terselesaikannya penelitian ini.
5. Teman-teman tongkrongan RAMBE yang selalu memberikan dukungan, terima kasih juga untuk canda dan tawa, serta kebersamaan yang selalu

menguatkan selama masa perkuliahan dan yang selalu dapat di ajak kerjasama apabila ada tugas perkuliahan.

6. Ervin Aditya, terima kasih atas segala usaha, perjuangan, dan keteguhan yang telah membawa karya ini hingga terwujud, menjadi bukti nyata perjalanan Panjang penulis.



KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala, atas limpahan rahmat, taufik, dan hidayah-Nya sehingga laporan kelulusan ini dapat terselesaikan dengan baik. Proses penyusunan laporan ini tidak terlepas dari bimbingan, bantuan, dukungan, serta doa dari berbagai pihak yang telah memberikan kontribusi berarti.

Oleh karena itu, dengan penuh rasa hormat dan penghargaan yang tulus, penulis menyampaikan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusrini, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta
3. Ibu Eli Pujiastuti, M.Kom., selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
4. Ibu Rumini, S.Kom., M.Kom., selaku Dosen Pembimbing, atas segala arahan, bimbingan, dan motivasi yang terus diberikan selama proses penyusunan penelitian ini.
5. Bapak Ajie Kusuma Wardhana, S.Kom., M.Eng., selaku Dosen yang turut membantu dalam penelitian dan eksperimen, serta memberikan masukan, panduan, dan motivasi sehingga penelitian dan penulisan ini dapat terselesaikan dengan baik.

Yogyakarta, 15 Desember 2025

Penulis

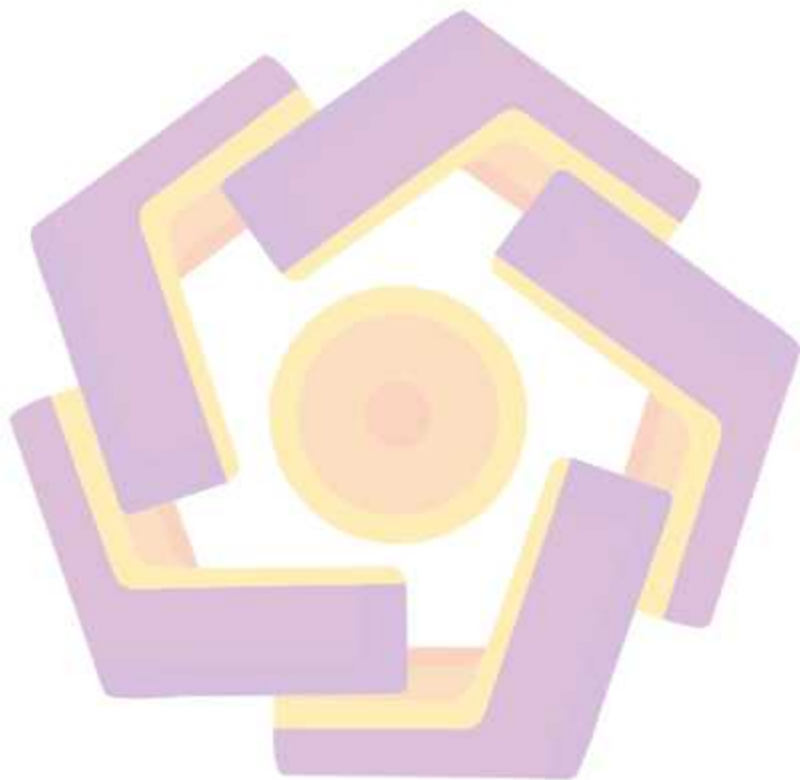
DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN KARYA.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN.....	xiii
DAFTAR ISTILAH.....	xiv
INTISARI.....	xv
<i>ABSTRACT</i>	xvi
Bab I Pendahuluan.....	1
1.1. Gambaran Umum.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah.....	3
1.4. Tujuan.....	3
Bab II Tinjauan Pustaka.....	5
2.1. Studi Literatur.....	5
2.2. Landasan Teori.....	6
2.2.1. Penyakit Tanaman.....	6
2.2.2. Citra Daun Tanaman.....	6
2.2.3. Machine Learning.....	6
2.2.4. Feature Extraction.....	7
2.2.5. Random Undersampling.....	7
2.2.6. Pca (Principal Component Analysis).....	7
2.2.7. Normalisasi Data.....	7
Bab III Metode Penelitian.....	9

3.1. Metode.....	9
3.1.1. Gathering Data	9
3.1.2. Feature Extraction	12
3.1.3. Data Balancing	13
3.1.4. Principal Component Analysis (Pca)	16
3.1.5. Data Preparation.....	16
3.1.6. Data Normalization	17
3.1.7. Models.....	17
3.1.8. Evaluation	18
3.1.9. Perhitungan Algoritma Klasifikasi.....	19
Bab IV Pembahasan.....	22
4.1. Sub Pembahasan	22
4.1.1. Data Balancing	22
4.1.2. Model Performance And Evaluation.....	23
4.1.3. Deployment.....	35
Bab V Kesimpulan	39
5.1. Kesimpulan	39
5.2. Saran.....	40
Referensi	41
Curriculum Vitae	44
Lampiran Dan Bukti Pendukung.....	45
A. Lampiran 1 Letter Of Acceptance (Loa)	45
B. Lampiran 2 Lembar Review	46
C. Lampiran 3 Bukti Terbit/Terindex.....	48
D. Lampiran 4 Bukti Pembayaran.....	48

DAFTAR TABEL

Tabel 3. 1 Perbandingan Teknik Penimbangan Kelas	14
Tabel 3. 2 Informasi Train dan Test Data	17
Table 4. 1 Hasil Perbandingan Model	24

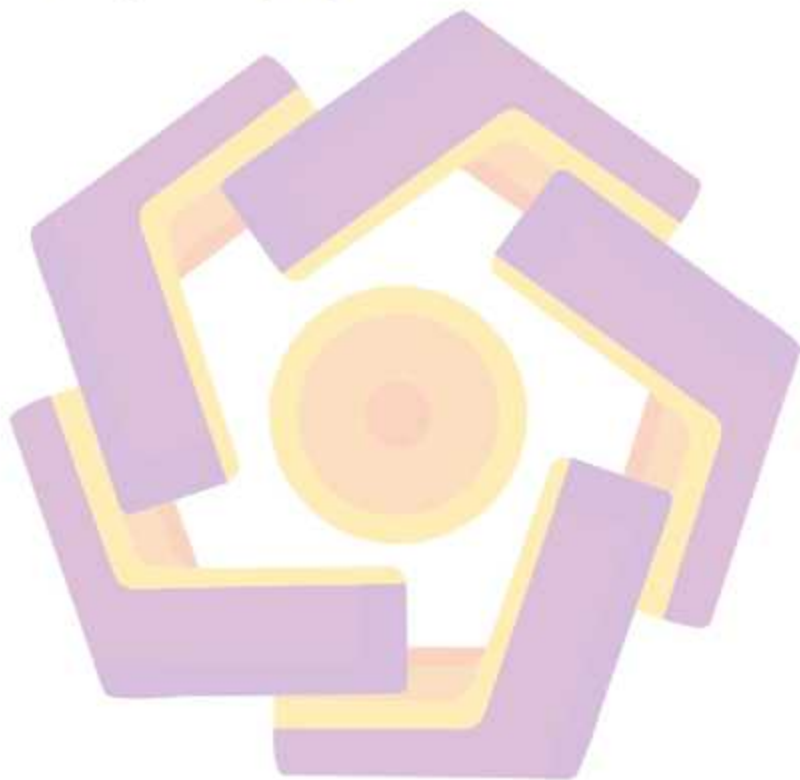


DAFTAR GAMBAR

Gambar 3. 1 Alur penelitian	9
Gambar 3. 2 Distribusi kelas dataset PlantVillage	11
Gambar 3. 3 Representatif daun yang terinfeksi penyakit	11
Gambar 3. 4 Representatif daun sehat	12
Gambar 3. 5 Visualisasi 2D dari ekstraksi fitur	13
Gambar 4. 1 Sebelum <i>Random Undersampling</i>	23
Gambar 4. 2 Sesudah <i>Random Undersampling</i>	23
Gambar 4. 3 Per kelas Akurasi <i>Logistic Regresion</i>	27
Gambar 4. 4 Per kelas Akurasi <i>Support Vector Machine</i>	28
Gambar 4. 5 Per Kelas Akurasi <i>LightGBM</i>	28
Gambar 4. 6 Per Kelas Akurasi <i>Random Forest</i>	29
Gambar 4. 7 Per Kelas Akurasi <i>Decision Tree</i>	29
Gambar 4. 8 Per Kelas Akurasi <i>AdaBoost</i>	30
Gambar 4. 9 Per Kelas Akurasi <i>Naive Bayes</i>	31
Gambar 4. 10 Per Kelas Akurasi <i>K-Nearest Neighbors</i>	31
Gambar 4. 11 Per Kelas Akurasi <i>Ridge Classifier</i>	32
Gambar 4. 12 Per Kelas Akurasi <i>XGBoost</i>	32
Gambar 4. 13 Per Kelas Akurasi <i>Gradient Boosting</i>	33
Gambar 4. 14 Model Performance Train & Test Accuracy	34
Gambar 4. 15 Hasil Klasifikasi Berhasil Mendeteksi Penyakit Tanaman	36
Gambar 4. 16 Hasil Klasifikasi Tidak Mendeteksi Penyakit Tanaman	37
Gambar 4. 17 Hasil Klasifikasi Data Uji Real	38

DAFTAR LAMPIRAN

a.	Lampiran 1 Letter of Acceptance (LOA)	45
b.	Lampiran 2 Lembar Review	46
c.	Lampiran 3 Bukti Terbit/Terindex	48
d.	Lampiran 4 Bukti pembayaran	48



DAFTAR LAMBANG DAN SINGKATAN



SVM	<i>Support Vector Machines</i>
PCA	Principal Component Analysis
CNN	Convolutional Neural Network
ML	Machine Learning
TP	True Positive
FN	False Negative
FP	False Positive
TN	True Negative
LR	Logistic Regression
DT	Decision Tree
RF	Random Forest
GB	Gradient Boosting
AB	Ada Boost
NB	Naive Bayes
KNN	K-Nearest Neighbors
XGB	eXtreme Gradient Boosting
LGBM	Light Gradient Boosting Machine
RC	Ridge Classifier
RUS	Random Under Sampling
DL	Deep Learning
ML	Machine Learning
CNN	Convensional Neural Network

DAFTAR ISTILAH



Vektor	besaran yang mempunyai arah
<i>Eigen Value</i>	akar-akar persamaan
<i>Overfitting</i>	Model terlalu sesuai data latih
<i>Underfitting</i>	Model terlalu sederhana
<i>Cross Validation</i>	Evaluasi model dengan lipatan data
<i>Class Imbalance</i>	Distribusi kelas yang tidak seimbang
<i>Accuracy</i>	Proporsi prediksi yang benar
<i>Precision</i>	Proporsi positif yang benar
<i>Recall</i>	Proporsi positif terdeteksi
<i>Random Under sampling</i>	Mengurangi data kelas mayoritas
Dataset	Kumpulan data

INTISARI

Penyakit daun pada tanaman merupakan ancaman serius bagi produktivitas pertanian, sehingga memerlukan sistem indentifikasi yang akurat dan efisien untuk penanganan tepat waktu. Penelitian ini mengusulkan pendekatan yang memanfaatkan ekstraksi fitur dan transfer learning ResNet50 dan dikombinasikan dengan algoritma machine learning untuk mengenali 38 class daun tanaman, termasuk kategori sehat dan sakit. Setiap gambar diubah menjadi vector fitur berdimensi 2048, diikuti dengan normalisasi dan pengurangan dimensi menggunakan Analisis Komponen Utama (PCA). Untuk mengatasi masalah ketidakseimbangan kelas dalam dataset, dilakukan Random Undersampling untuk memastikan representasi yang seimbang di semua kelas. Sebelas model machine learning dilatih dan dievaluasi menggunakan K-Fold Cross Validation, akurasi, presisi, recall, F1-score, dan ROC AUC score. Hasil menunjukkan bahwa algoritma SVM mencapai kinerja tertinggi dengan akurasi 99,63%, diikuti oleh Logistic Regression dengan akurasi 97,33%, dan LightGBM dengan akurasi 96,25%. Model-model tersebut menunjukkan kemampuan generalisasi yang kuat dalam multi kelas, sementara algoritma seperti AdaBoost dan Decision Tree menghasilkan akurasi yang lebih rendah. Analisis perbandingan akurasi pelatihan dan pengujian menunjukkan seberapa kuat model bekerja serta apakah model mengalami overfitting atau tidak. Temuan ini menekankan potensi penggunaan model CNN pretrained sebagai ekstraktor fitur lalu digabungkan dengan model klasifikasi berpotensi memberikan hasil yang lebih baik dalam menyelesaikan masalah klasifikasi pertanian yang rumit. Untuk penelitian berikutnya, disarankan menggunakan dataset yang lebih besar dan beragam agar hasilnya lebih akurat dan lebih mewakili kondisi nyata, serta integrasi metode yang diusulkan ke dalam platform diagnostik yang mudah diakses untuk mendukung pertanian presisi dan meningkatkan pemantauan kesehatan tanaman.

Kata kunci: Deteksi Penyakit Tanaman, Machine Learning, Under-Sampling, Pre-Trained

ABSTRACT

Plant leaf diseases pose a significant threat to agricultural productivity, necessitating accurate and efficient identification systems for timely intervention. This study proposes an approach that leverages deep feature extraction using a pretrained ResNet50 model combined with machine learning algorithms to recognize 38 classes of plant leaves, including both healthy and diseased categories. Each image was transformed into a 2048-dimensional feature vector, followed by normalization and dimensionality reduction using Principal Component Analysis (PCA). To mitigate the issue of class imbalance in the dataset, random undersampling was applied at the feature level to ensure equal representation across all classes. Eleven machine learning models were trained and evaluated using 5-fold cross-validation, with performance assessed through accuracy, precision, recall, F1-score, and ROC AUC score. Among the evaluated models, the Support Vector Machine (SVM) achieved the highest accuracy of 99.63%, followed by Logistic Regression at 97.33%, and LightGBM at 96.25%. These models demonstrated strong generalization capabilities in multiclass settings, while simpler classifiers like AdaBoost and Decision Tree yielded lower performance. A comparative analysis of training and test accuracy further highlighted model robustness and overfitting tendencies. The findings emphasize the potential of combining pretrained convolutional neural networks for feature extraction with conventional classifiers to address complex agricultural classification tasks. Future work may explore the inclusion of larger and more diverse samples, as well as the integration of the proposed method into accessible diagnostic platforms to support precision farming and enhance crop health monitoring.

Keyword: Plant Disease Detection, Machine Learning, Under-Sampling, Pre-trained ResNet50, CNN