

**OPTIMASI ALGORITMA MACHINE LEARNING DENGAN
HYPERPARAMETER TUNING GRID SEARCH CV PADA
KLASIFIKASI PENYAKIT JANTUNG**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

RAHMAN HAKIM

21.11.4416

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**OPTIMASI ALGORITMA MACHINE LEARNING DENGAN
HYPERPARAMETER TUNING GRID SEARCH CV PADA
KLASIFIKASI PENYAKIT JANTUNG**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

RAHMAN HAKIM

21.11.4416

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

OPTIMASI ALGORITMA MACHINE LEARNING DENGAN
HYPERPARAMETER TUNING GRID SEARCH CV PADA
KLASIFIKASI PENYAKIT JANTUNG

yang disusun dan diajukan oleh

Rahman Hakim

21.11.4416

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18 Desember 2024

Dosen Pembimbing

Robert Maroo, S.T., M.T., Ph.D.
NIK. 190302228

HALAMAN PENGESAHAN

SKRIPSI

OPTIMASI ALGORITMA MACHINE LEARNING DENGAN
HYPERPARAMETER TUNING GRID SEARCH CV PADA
KLASIFIKASI PENYAKIT JANTUNG

yang disusun dan diajukan oleh

Rahman Hakim
21.11.4416

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Desember 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ike Vervanti, M.Kom
NIK. 190302237

Ferian Fauzi Abdullahi, S.Kom., M.Kom.
NIK. 190302276

Robert Murey, S.E., M.E., Ph.D.
NIK. 190302228

Naskah ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
tanggal 18 Desember 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fata, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini:

Nama mahasiswa : **Rahman Hakim**
NIM : **21.11.4416**

Menyatakan bahwa Skripsi dengan judul berikut:

**Optimasi Algoritma Machine Learning dengan Hyperparameter Tuning
Grid Search CV pada Klasifikasi Penyakit Jantung**

Dosen Pembimbing: **Robert Marcus, S.T., M.T., Ph.D.**

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diikutkan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian 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. Peningkat hasil 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 ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan penarikan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Desember 2024

Yang Menyatakan

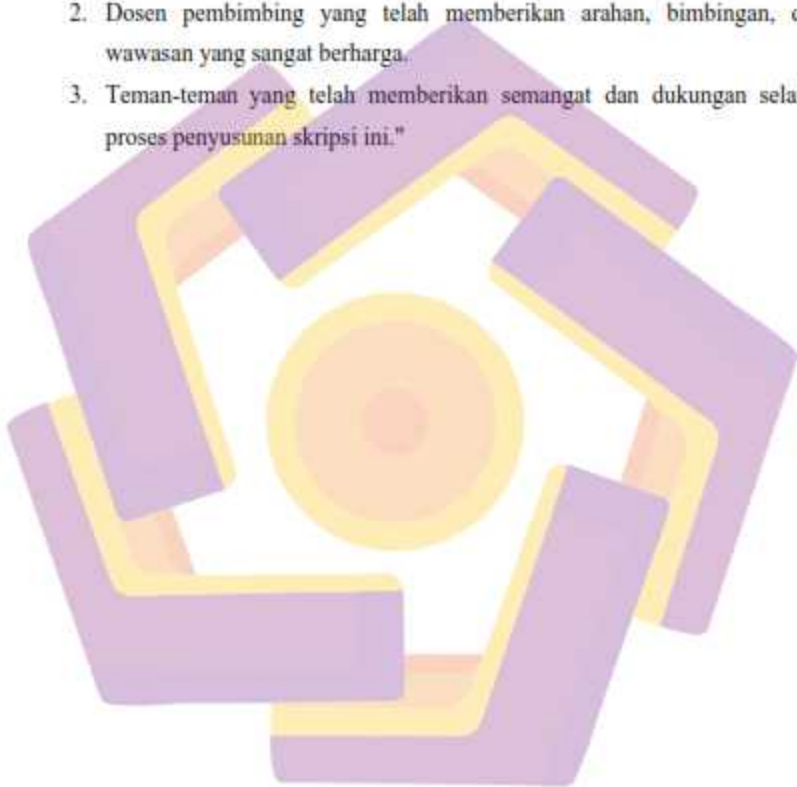


Rahman Hakim

HALAMAN PERSEMBAHAN

Skripsi ini dipersembahkan kepada:

1. Orang tua saya yang telah memberikan dukungan, cinta, dan doa yang tak ternilai.
2. Dosen pembimbing yang telah memberikan arahan, bimbingan, dan wawasan yang sangat berharga.
3. Teman-teman yang telah memberikan semangat dan dukungan selama proses penyusunan skripsi ini."



KATA PENGANTAR

Puji Syukur kehadiran Tuhan Yang Maha Esa, atas berkat, rahmat, dan hidayah-Nya, sehingga penulis dapat melaksanakan dan menyelesaikan tugas akhir skripsi dengan judul “Optimasi Algoritma Machine Learning dengan Hyperparameter Tuning Grid Search CV Pada Klasifikasi Penyakit Jantung”. Skripsi ini disusun guna memenuhi syarat memperoleh gelar sarjana di Universitas Amikom Yogyakarta.

Pada kesempatan kali ini, penulis ingin menyampaikan ucapan terimakasih kepada semua pihak yang telah membantu serta membimbing penulis dalam melaksanakan dan menyusun skripsi ini, kepada:

1. Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta, yang telah memberikan ijin penulis untuk menempuh pendidikan sarjana di Universitas Amikom Yogyakarta.
2. Ibu Windha Mega Pradya D, M.Kom selaku Kepala Prodi SI Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah memberikan arahan dan dorongan semangat dalam menyusun skripsi ini.
3. Bapak Robert Marco, M.T., Ph.D selaku dosen pembimbing yang selalu senantiasa memotivasi, menyarankan, mengarahkan dan membimbing kepada penulis selama penyusunan skripsi.
4. Bapak Trisanto Ari Aji, M.Kom selaku dosen wali yang selalu memberikan arahan dan dukungan selama studi di kampus Universitas Amikom Yogyakarta.
5. Kedua Orang Tua yang selalu memberikan motivasi, semangat, doa dan arahnya kepada penulis.
6. Rekan-rekan mahasiswa dari prodi informatika, yang telah membantu penulis dalam memberikan motivasi, dukungan dan berdiskusi selama proses penulisan skripsi.

7. Bolo Tampil Soenarmi Kos dan Saryanto Grup yang selalu memberikan semangat saat keadaan sedang kesulitan dalam mengerjakan penelitian ini.

Penulis menyadari bahwa skripsi ini jauh dari kesempurnaan, baik secara materi maupun penyajiannya. Oleh karena itu, penulis mengharapkan adanya kritik dan saran yang membangun untuk penyusunan skripsi agar kedepannya bisa lebih baik. Semoga skripsi ini dapat bermanfaat bagi penulis ataupun pembaca khususnya dalam bidang machine learning dan kesehatan.

Yogyakarta, 14 Desember 2024

Penulis,

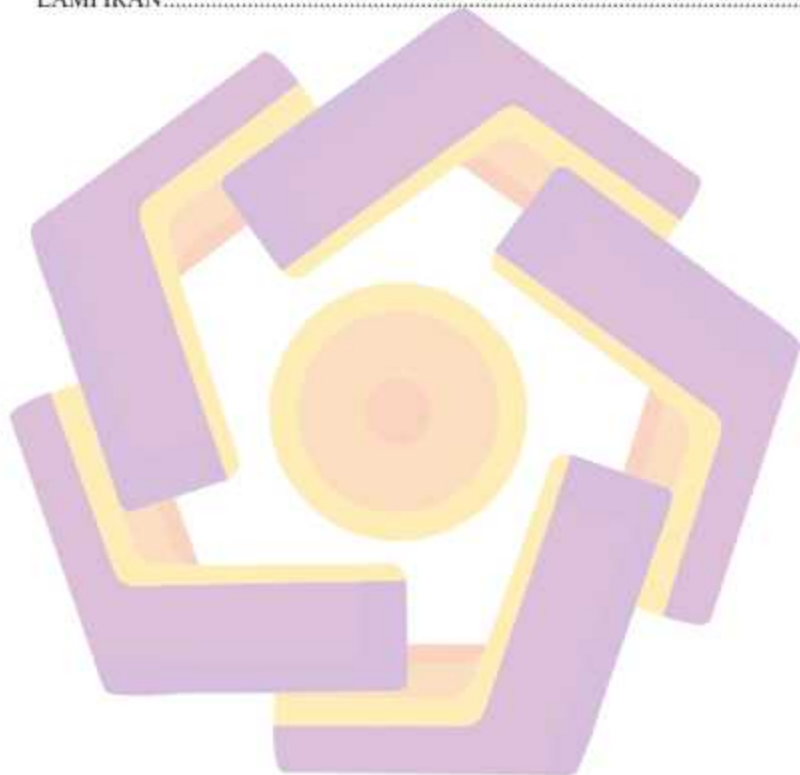


DAFTAR ISI

HALAMAN JUDUL	1
HALAMAN PERSEMBAHAN	5
KATA PENGANTAR	6
DAFTAR TABEL	11
DAFTAR GAMBAR	12
DAFTAR LAMPIRAN	13
DAFTAR LAMBANG DAN SINGKATAN	14
DAFTAR ISTILAH	15
INTISARI	16
ABSTRACT	17
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penelitian	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Landasan Teori	14
2.2.1 Big Data	14
2.2.2 Data Mining	14
2.2.3 Klasifikasi	16
2.2.4 Machine Learning	16
2.2.5 K-Nearest Neighbors	17
2.2.6 Random Forest	18
2.2.7 Extreme Gradient Boosting	20
2.2.8 Decision Tree	21

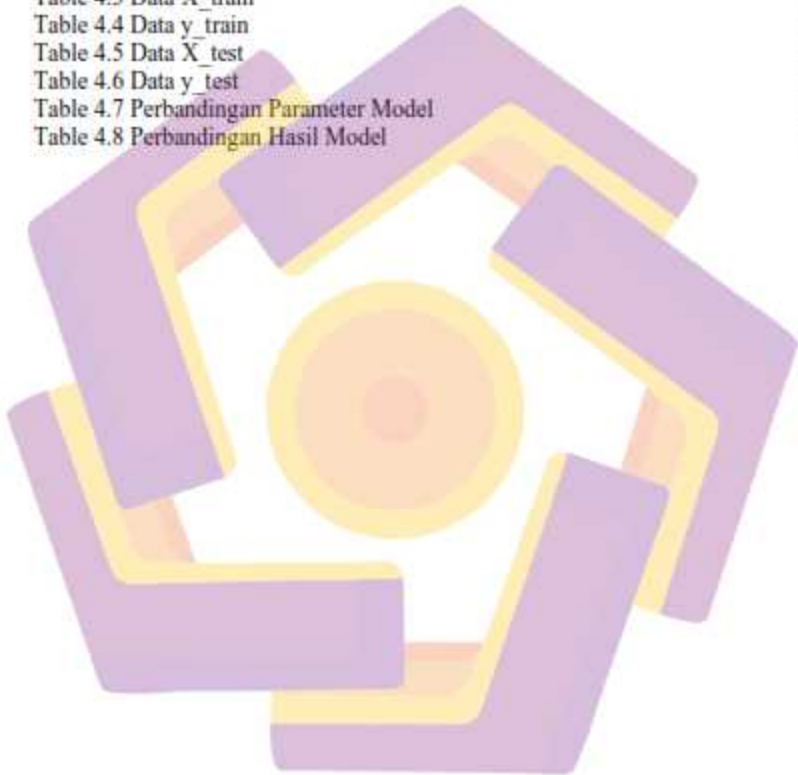
2.2.9	Hyperparameter Tuning	22
BAB III METODOLOGI PENELITIAN		23
3.1	Object Penelitian.....	23
3.2	Tahapan Penelitian.....	24
3.2.1	Pengumpulan Data	25
3.2.2	Preprocessing Data.....	25
3.2.3	Splitting Data	26
3.2.4	Pemodelan dan Hyperparameter Tuning.....	27
3.2.5	Evaluasi Model	27
3.3	Alat & Bahan.....	28
3.3.1	Deskripsi Data.....	29
3.3.2	Perangkat Lunak	30
3.3.3	Perangkat Keras	31
BAB IV HASIL DAN PEMBAHASAN		32
3.1	Dataset	32
3.2	Preprocessing Data	32
4.2.1	Menghapus Fitur (Drop)	32
4.2.2	Label Encoding	33
4.2.3	Missing Value	33
4.2.4	Handling Imbalance Data.....	35
3.3	Splitting Data	37
3.4	Pemodelan dan Hyperparameter Tuning	38
4.4.1	Model K – Nearest Neighbors	38
4.4.2	Model Decision Tree.....	40
4.4.3	Model Random Forest.....	41
4.4.4	Model <i>XGBOOST</i>	43
4.4.5	Parameter Model	44

3.5 Evaluasi Model	45
BAB V Kesimpulan dan Saran	51
5.1 Kesimpulan	51
5.2 Saran	51
REFERENSI	52
LAMPIRAN.....	58



DAFTAR TABEL

Table 2.1 Keaslian Penelitian	8
Table 3.2.5 Confusion Matrix	28
Tabel 3.3.1 Deskripsi Dataset	29
Table 4.1 Data sebelum di Encoding	33
Table 4.2 Data setelah di Encoding	33
Table 4.3 Data X_{train}	37
Table 4.4 Data y_{train}	37
Table 4.5 Data X_{test}	38
Table 4.6 Data y_{test}	38
Table 4.7 Perbandingan Parameter Model	45
Table 4.8 Perbandingan Hasil Model	48



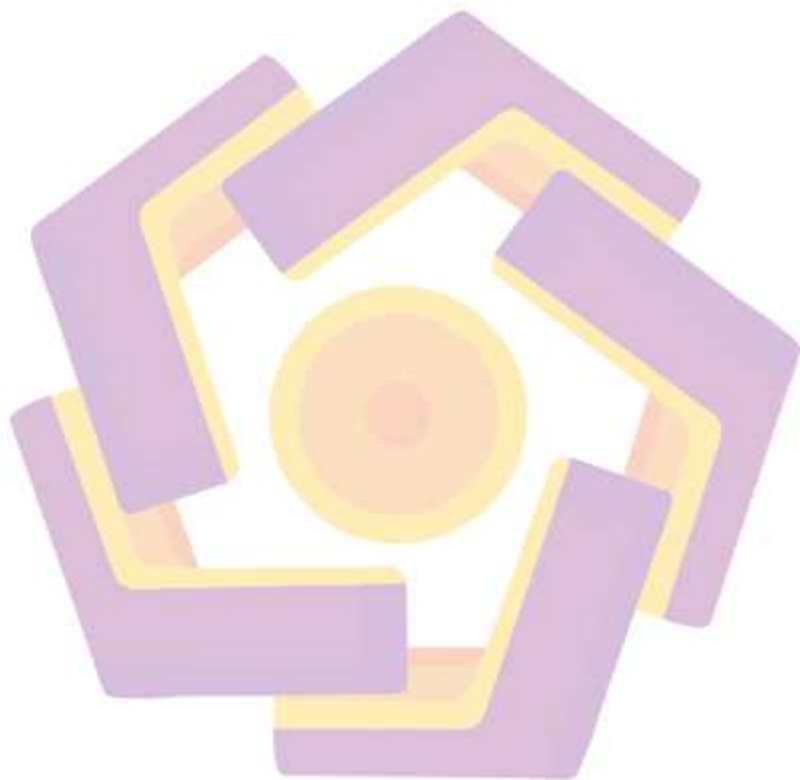
DAFTAR GAMBAR

Gambar 2.2.2 Knowledge Discovery in Database	14
Gambar 2.2.6 Random Forest	19
Gambar 2.2.8.1 Grid Search CV	22
Gambar 3.2 Flowchart Penelitian	24
Gambar 4.1 Dataset	32
Gambar 4.2 Missing Value	34
Gambar 4.4 Hasil Handling Missing Value	35
Gambar 4.5 Sebelum dilakukan SMOTE	35
Gambar 4.6 Proses SMOTE	36
Gambar 4.7 Setelah dilakukan SMOTE	36
Gambar 4.8 Confusion Matrix KNN	46
Gambar 4.9 Confusion Matrix Decision Tree	46
Gambar 4.10 Confusion Matrix XGBOOST	47
Gambar 4.11 Confusion Matrix Random Forest	47
Gambar 4.12 Perbandingan Accuracy Model	48



DAFTAR LAMPIRAN

Lampiran 1. Perhitungan Confusion Matrics	52
Lampiran 2. Source Code	54



DAFTAR LAMBANG DAN SINGKATAN



Ω	Fungsi Regularisasi
L	Fungsi pelatihan yang hilang
θ	Parameter model
S	Himpunan kasus
A	Atribut
AI	Artificial Intelligence
KDD	Knowledge Discovery in Databases
SMOTE	Synthetic Minority Over-sampling Technique
KNN	K-Nearest Neighbors
k	Jumlah tetangga terdekat pada KNN ($n_neighbors$)
CV	Cross Validation
TP	True Positif
TN	True Negatif
FP	False Positif
FN	False Negatif
XGBOOST	Extreme Gradient Boosting

DAFTAR ISTILAH



Overfitting	Model terlalu menghafal data pelatihan sehingga ketika dilakukan pada data baru hasilnya kurang
Bias	Hasil Perbedaan antara nilai aktual dengan prediksi yang dihasilkan oleh model
Error term	Kesalahan antara hasil prediksi dengan nilai actual
Optimasi	Proses meningkatkan kinerja model dengan penyesuaian parameter
Parameter	Nilai yang dipelajari model untuk memetakan input ke output secara optimal.
Hyperparameter	Penentuan nilai parameter sebelum dilakukan pelatihan model
Kelas	Kategori atau label yang menjadi target prediksi model
Fitur/atribut	Variable yang digunakan untuk membuat prediksi
Oversampling	Meningkatkan jumlah kelas minoritas agar jumlahnya sama dengan kelas mayoritas

INTISARI

Penyakit jantung adalah salah satu penyakit yang berbahaya dan memiliki tingkat resiko kematian tinggi di seluruh dunia. Jumlah kasus yang terjadi tiap tahunnya berdasarkan WHO tercatat 17.8 juta kematian akibat penyakit jantung, dengan 12.934.931 kasus terjadi pada tahun yang sama. Jumlah kasus yang tinggi karena kurangnya metode untuk memprediksi adanya penyakit jantung secara tepat. Maka dari itu, untuk membantu ahli jantung dalam mendiagnosis, merencanakan pengobatan dan memberikan perawatan yang sesuai. Dalam era digital, teknik machine learning cocok digunakan untuk mengatasi hal tersebut. Machine learning menjadi alat yang kuat untuk melakukan klasifikasi dan prediksi. Namun, ada masalah baru yang muncul yaitu penggunaan teknik machine learning umumnya masih melakukan pencarian parameternya dilakukan secara manual dan belum ada hubungan antara output algoritma dengan *hyperparameter tuning*. Oleh karena itu, untuk mengatasi hal tersebut diperlukan teknik pencarian parameter yang optimal untuk mempermudah menemukan kombinasi parameter terbaik dari masing-masing algoritma. Metode untuk menemukan parameter terbaik salah satunya adalah Grid Search CV. Penelitian ini bertujuan untuk mengoptimasi kinerja algoritma KNN, Random Forest, XGBOOST, dan Decision Tree dengan metode Grid Search CV pada klasifikasi penyakit jantung. Penelitian menunjukkan bahwa implementasi *hyperparameter tuning* menggunakan Grid Search CV berhasil meningkatkan kinerja model machine learning. Namun, terdapat perbedaan dari tingkat peningkatannya. Dilihat dari hasil accuracy, algoritma KNN mengalami peningkatan sebesar 1%, Decision Tree mengalami peningkatan sebesar 7%, XGBOOST mengalami peningkatan sebesar 4% dan Random Forest mengalami peningkatan sebesar 7%. Pada saat sebelum tuning, KNN memiliki performa terbaik dengan akurasi 89%. Setelah tuning, KNN dan XGBoost mencapai akurasi tertinggi sebesar 90%. Hasil penelitian diharapkan membantu bidang kesehatan dalam mendeteksi dini penyakit jantung secara akurat, sehingga dapat meningkatkan kualitas hidup pasien. Selain itu, penelitian ini menambahkan wawasan tentang algoritma dan teknik tuning yang dapat menjadi referensi untuk penelitian selanjutnya.

Kata kunci: Machine Learning, Klasifikasi, Hyperparameter Tuning, Grid Search CV, Optimasi

ABSTRACT

Heart disease is one of the most dangerous diseases and has a high risk of death worldwide. The number of cases that occur each year according to WHO is 17.8 million deaths from heart disease, with 12,934,931 cases occurring in the same year. The high number of cases is due to the lack of methods to predict the presence of heart disease precisely. Therefore, to assist cardiologists in diagnosing, planning treatment and providing appropriate care. In the digital age, machine learning techniques are suitable for this purpose. However, there is a new problem that arises, namely the use of machine learning techniques generally still search for parameters manually and there is no relationship between algorithm output and hyperparameter tuning, therefore, to overcome this, an optimal parameter search technique is needed to make it easier to find the best parameter combination from each algorithm. One method to find the best parameters is Grid Search CV. This study aims to optimize the performance of the KNN, Random Forest, XGBOOST, and Decision Tree algorithms with the Grid Search CV method in the classification of heart disease. The research shows that the implementation of hyperparameter tuning using Grid Search CV successfully improved the performance of machine learning models. However, there are differences in the level of improvement. Based on the accuracy results, the KNN algorithm experienced an increase of 1%, Decision Tree showed an improvement of 7%, XGBoost had an increase of 4%, and Random Forest saw an improvement of 7%. Before tuning, KNN had the best performance with an accuracy of 89%. After tuning, both KNN and XGBoost achieved the highest accuracy of 90%. The results of this research are expected to assist the healthcare field in the early detection of heart disease accurately, thus improving patients' quality of life. Furthermore, this study provides additional insights into algorithms and tuning techniques that can serve as references for future research.

Keywords: Machine Learning, Classification, Hyperparameter Tuning, Grid Search CV, Optimization.