

# TESIS

**OPTIMASI PERFORMA MODEL KLASIFIKASI  
MENGUNAKAN TEKNIK *HYPERPARAMETER TUNING*  
PADA ALGORITMA *K-NEAREST NEIGHBORS*,  
*SUPPORT VECTOR MACHINE*, DAN *RANDOM FOREST***



Oleh :

**IZZA AFKARINA**

**20206052007**

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

**PROGRAM STUDI INFORMATIKA  
PROGRAM MAGISTER FAKULTAS SAINS DAN  
TEKNOLOGI  
UIN SUNAN KALIJAGA**

**YOGYAKARTA  
2024**

# PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini:

Nama : Izza Afkarina

NIM : 20206052007

Jenjang : Magister

Program Studi : Informatika

menyatakan bahwa tesis saya yang berjudul “Optimasi Performa Model Klasifikasi Menggunakan Teknik *Hyperparameter Tuning* Pada Algoritma *K-Nearest Neighbors*, *Support Vector Machine*, Dan *Random Forest*” merupakan hasil penelitian saya sendiri, tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang sepengetahuan saya, tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 9 Mei 2024

Penulis,



Izza Afkarina

NIM. 20206052007

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

# PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama : Izza Afkarina

NIM : 20206052007

Jenjang : Magister

Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan benar-benar bebas dari plagiasi. Jika di kemudian hari terbukti melakukan plagiasi, maka saya siap ditindak sesuai ketentuan hukum yang berlaku.

Yogyakarta, 16 Mei 2024

Penulis,



Izza Afkarina

NIM. 20206052007

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

# LEMBAR PENGESAHAN



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
FAKULTAS SAINS DAN TEKNOLOGI  
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

## PENGESAHAN TUGAS AKHIR

Nomor : B-702/Un.02/DST/PP.00.9/05/2024

Tugas Akhir dengan judul : Optimasi Performa Model Klasifikasi Menggunakan Teknik Hyperparameter Tuning Pada Algoritma K-Nearest Neighbours, Support Vector Machine dan Random Forest

yang dipersiapkan dan disusun oleh:

Nama : IZZA AFKARINA, S.T.  
Nomor Induk Mahasiswa : 20206052007  
Telah diujikan pada : Kamis, 16 Mei 2024  
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

### TIM UJIAN TUGAS AKHIR



Ketua Sidang

Dr. Ir. Bambang Sugiantoro, S.Si., M.T., IPM., ASEAN Eng.  
SIGNED

Valid ID: 664bf7546da29



Penguji I

Dr. Agung Fatwanto, S.Si., M.Kom.  
SIGNED

Valid ID: 664ab87e8389



Penguji II

Dr. Ir. Sumarsono, S.T., M.Kom.  
SIGNED

Valid ID: 664ac765c2636



Yogyakarta, 16 Mei 2024  
UTN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.  
SIGNED

Valid ID: 664ac3ed3a69b

# NOTA DINAS PEMBIMBING



Universitas Islam Negeri Sunan Kalijaga

## SURAT PERSETUJUAN TUGAS AKHIR

Hal : Persetujuan Tugas Akhir  
Lamp : -

Kepada  
Yth. Dekan Fakultas Sains dan Teknologi  
UIN Sunan Kalijaga  
di-  
Yogyakarta

*Assalamualaikum wr. wb.*

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa tugas tesis Saudara :

Nama : Izza Afkarina  
NIM : 20206052007  
Judul Tesis : Optimasi Performa Model Klasifikasi Menggunakan Teknik *Hyperparameter Tuning*  
Pada Algoritma *K-Nearest Neighbors*, *Support Vector Machine*, dan *Random Forest*

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Program Studi Magister Informatika UIN Sunan Kalijaga untuk diajukan dalam rangka memperoleh gelar Magister Informatika.

*Wassalamualaikum wr. wb.*

Yogyakarta, 13 Mei 2024

Pembimbing

**Dr. Ir. Bambang Sugiantoro, S.SI., M.T., IPM.**

NIP: 19751024 200912 1 002

## ABSTRAK

Klasifikasi merupakan proses identifikasi model atau pola yang mampu menggambarkan dan memisahkan kelas dalam dataset. Tujuannya untuk memungkinkan model melakukan prediksi terhadap objek dengan label kelas yang belum diketahui. Klasifikasi dapat menyelesaikan beberapa permasalahan di antaranya adalah: *data classification*, *image classification*, *text categorization* dan lain-lain. Beberapa metode yang dapat digunakan untuk menyelesaikan permasalahan klasifikasi diantaranya adalah: K-NN, SVM dan *random forest*. Tolak ukur seberapa baik performa model klasifikasi adalah dengan melakukan beberapa pengujian seperti pengujian nilai akurasi, presisi, *recall* dan *f-measure*. Meningkatkan performa model klasifikasi dapat dilakukan dengan mengatur *hyperparameter*. Namun, berbagai algoritma klasifikasi yang tersedia memiliki set *hyperparameter* yang berbeda. Oleh karena itu, proses penentuan *hyperparameter* yang optimal melalui teknik *hyperparameter tuning* diperlukan untuk mendapatkan *hyperparameter* yang dapat memaksimalkan performa model. Tujuan dari penelitian ini adalah menerapkan teknik *hyperparameter tuning grid search* dan *random search* pada beberapa model klasifikasi yang memiliki kemampuan menangani data berdimensi besar, yaitu K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), dan *random forest*. *Grid search* dan *random search* digunakan pada penelitian ini karena kedua metode tersebut merupakan metode yang paling mudah untuk diimplementasikan dibandingkan metode lain yang lebih kompleks namun metode ini tetap memiliki performa yang baik saat digunakan untuk pencarian *hyperparameter* optimal. Teknik *hyperparameter tuning grid search* dan *random search* yang diterapkan menggunakan *CrossValidation*, sehingga performa yang baik pada proses pelatihan model juga tercermin pada proses pengujian. Data yang digunakan pada penelitian ini merupakan data yang berasal dari situs UCI Machine Learning dengan judul "*Predict Students' Dropout and Academic Success*". Dataset tersebut berisi 4424 *record* dengan 36 *features* dan 1 target. Dilakukan beberapa kali pengujian klasifikasi terhadap algoritma yang digunakan dengan menerapkan teknik *hyperparameter tuning grid search* dan *random search*. *Grid search* dan *random search* berhasil meningkatkan akurasi algoritma K-NN dan *random forest* dibandingkan dengan tanpa optimasi. Sedangkan akurasi model SVM tidak ada peningkatan meskipun menggunakan

*hyperparameter* optimal dari hasil *tuning*. Akurasi yang dihasilkan semua model sudah cukup baik, namun selalu ada ruang untuk peningkatan performa dengan mengeksplorasi teknik pemodelan yang lebih canggih mengingat kompleksitas data yang digunakan cukup tinggi.

**Kata Kunci :** K-NN, SVM, *random forest*, *grid search* dan *random search*



## ABSTRACT

Classification is the process of identifying models or patterns capable of depicting and segregating classes within a dataset. Its aim is to enable models to predict objects with unknown class labels. Classification can address various issues including data classification, image classification, text categorization, protein structure predictions, and more. Several methods can be utilized to tackle classification problems, such as K-NN, SVM, and random forest. Evaluating the performance of a classification model is done through multiple tests like accuracy, precision, recall, and f-measure. Enhancing the performance of a classification model can be achieved by adjusting hyperparameters. However, different classification algorithms come with distinct sets of hyperparameters. Thus, determining optimal hyperparameters through techniques like hyperparameter tuning is necessary to obtain parameters that maximize model performance. The objective of this research is to apply grid search and random search hyperparameter tuning techniques to several classification models capable of handling high-dimensional data, namely K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), and random forest. Grid search and random search are chosen for their simplicity in implementation compared to more complex methods, yet they maintain good performance when seeking optimal hyperparameters. The hyperparameter tuning techniques applied here use CrossValidation, ensuring that good performance in model training is also reflected in the testing process. The dataset used in this research is sourced from the UCI Machine Learning repository titled "Predict Students' Dropout and Academic Success." It comprises 4424 records with 36 features and 1 target variable. Multiple classification tests were conducted on the algorithms used by applying grid search and random search hyperparameter tuning techniques. Grid search and random search successfully improved the accuracy of the K-NN and random forest algorithms compared to unoptimized models. However, there was no improvement in SVM model accuracy despite using optimal hyperparameters from tuning.

**Keywords :** K-NN, SVM, *random forest, grid search and random search*



## KATA PENGANTAR

Puji syukur kehadiran Allah Yang Maha Esa, atas segala rahmat dan karunia-Nya sehingga tesis yang berjudul "Optimasi Performa Model Klasifikasi Menggunakan Teknik *Hyperparameter tuning* Pada Algoritma *K-Nearest neighbors*, *Support vector machine*, dan *Random forest*" ini dapat diselesaikan dengan baik. Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Informatika di Universitas Islam Negeri Sunan Kalijaga.

Penyusunan tesis ini tidak lepas dari bantuan, dukungan, dan kerjasama dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, baik secara langsung maupun tidak langsung. Bantuan yang diberikan, mulai dari bimbingan akademis, dukungan moril, hingga fasilitas dan informasi yang diperlukan untuk penelitian ini, sangatlah berarti bagi penulis.

Kami berharap adanya kritik dan saran yang membangun untuk perbaikan di masa mendatang, mengingat tidak ada sesuatu yang sempurna tanpa saran yang membangun. Semoga tesis ini dapat memberikan manfaat dan kontribusi positif bagi pengembangan ilmu pengetahuan, khususnya dalam bidang teknologi.

Yogyakarta, 17 Mei 2024

Izza Afkarina

# DAFTAR ISI

<b>PERNYATAAN KEASLIAN .....</b>	<b>ii</b>
<b>PERNYATAAN BEBAS PLAGIASI.....</b>	<b>iii</b>
<b>LEMBAR PENGESAHAN.....</b>	<b>iv</b>
<b>NOTA DINAS PEMBIMBING .....</b>	<b>v</b>
<b>ABSTRAK.....</b>	<b>vi</b>
<b>ABSTRACT .....</b>	<b>viii</b>
<b>KATA PENGANTAR .....</b>	<b>ix</b>
<b>DAFTAR ISI .....</b>	<b>x</b>
<b>DAFTAR TABEL.....</b>	<b>xii</b>
<b>DAFTAR GAMBAR .....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan Penelitian .....	5
1.5 Manfaat Penelitian .....	5
<b>BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI.....</b>	<b>7</b>
2.1 Kajian Pustaka .....	7
2.2 Landasan Teori.....	9
2.2.1 Klasifikasi.....	9
2.2.2 <i>K-Nearest Neighbor (K-NN)</i> .....	10
2.2.3 <i>Support vector machine (SVM)</i> .....	12
2.2.4 <i>Random forest</i> .....	14
2.2.5 <i>Hyperparameter tuning</i> .....	16

2.2.6	<i>Grid search</i> .....	16
2.2.7	<i>Random search</i> .....	16
2.2.8	<i>Confussion Matrix</i> .....	17
<b>BAB III METODOLOGI PENELITIAN</b> .....		<b>19</b>
3.1	Tahap Identifikasi .....	19
3.2	Pengumpulan Data .....	20
3.3	Pemodelan Data .....	20
3.4	Penyimpulan Hasil .....	22
<b>BAB IV PEMBAHASAN</b> .....		<b>23</b>
4.1	Pengumpulan Data .....	23
4.2	Pemodelan Data .....	25
4.2.1	Prapemrosesan Data .....	32
4.2.2	Klasifikasi Algoritma K-NN .....	35
4.2.3	Klasifikasi Algoritma SVM .....	42
4.2.4	Klasifikasi Algoritma <i>Random forest</i> .....	48
4.3	Hasil Evaluasi .....	54
<b>BAB V PENUTUP</b> .....		<b>59</b>
5.1	Kesimpulan .....	59
5.2	Saran .....	59
<b>DAFTAR PUSTAKA</b> .....		<b>61</b>
<b>DAFTAR LAMPIRAN</b> .....		<b>64</b>
<b>DAFTAR RIWAYAT HIDUP</b> .....		<b>96</b>

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 2. 1 Tabel <i>Confusion matrix</i> .....	17
Tabel 4. 1 Daftar Kelompok Atribut .....	24
Tabel 4. 2 Dataset .....	26
Tabel 4. 3 <i>Features</i> dengan Data Kategorik .....	29
Tabel 4. 4 <i>Features</i> dengan Data Numerik .....	31
Tabel 4. 5 Data Latih <i>Scaling</i> .....	32
Tabel 4. 6 Data Uji <i>Scaling</i> .....	33
Tabel 4. 7 Hasil Evaluasi K-NN .....	37
Tabel 4. 8 Parameter Terbaik K-NN dengan <i>Grid search</i> .....	39
Tabel 4. 9 Hasil Evaluasi K-NN dengan <i>Grid search</i> .....	39
Tabel 4. 10 Parameter Terbaik K-NN dengan <i>Random search</i> .....	41
Tabel 4. 11 Hasil Evaluasi K-NN dengan <i>Random search</i> .....	41
Tabel 4. 12 Hasil Evaluasi SVM .....	44
Tabel 4. 13 Parameter Terbaik SVM dengan <i>Grid search</i> .....	45
Tabel 4. 14 Hasil Evaluasi SVM dengan <i>Grid search</i> .....	45
Tabel 4. 15 Parameter Terbaik SVM dengan <i>Random search</i> .....	47
Tabel 4. 16 Hasil Evaluasi SVM dengan <i>Random search</i> .....	47
Tabel 4. 17 Hasil Evaluasi <i>Random forest</i> .....	49
Tabel 4. 18 Parameter Terbaik <i>Random forest</i> dengan <i>Grid search</i> ...	51
Tabel 4. 19 Hasil Evaluasi <i>Random forest</i> dengan <i>Grid search</i> .....	51
Tabel 4. 20 Parameter Terbaik <i>Random forest</i> dengan <i>Random search</i> .....	52
Tabel 4. 21 Hasil Evaluasi <i>Random forest</i> dengan <i>Random search</i> ....	53
Tabel 4. 22 Peningkatan Performa Model .....	58

## DAFTAR GAMBAR

Gambar	Halaman
Gambar 2. 1 Klasifikasi K-NN dengan 1, 3 dan 5 tetangga terdekat ..	12
Gambar 2. 2 Klasifikasi SVM .....	12
Gambar 2. 3 Klasifikasi <i>Random forest</i> .....	15
Gambar 3. 1 Metodologi Penelitian.....	19
Gambar 3. 2 <i>Flowchart</i> Model Klasifikasi.....	21
Gambar 3. 1 Metodologi Penelitian.....	19
Gambar 3. 2 <i>Flowchart</i> Model Klasifikasi.....	21
Gambar 4. 1 Pengecekan <i>Missing Value</i> .....	25
Gambar 4. 2 Distribusi Kelas <i>Dropout, Enrolled</i> dan <i>Graduate</i> .....	26
Gambar 4. 3 <i>Association Matrix</i> .....	29
Gambar 4. 4 <i>Correlation Ratio</i> .....	31
Gambar 4. 5 Visualisasi <i>Confusion matrix</i> K-NN.....	38
Gambar 4. 6 Visualisasi <i>Confusion matrix</i> K-NN dengan <i>Grid search</i> .....	40
Gambar 4. 7 Visualisasi <i>Confusion matrix</i> K-NN dengan <i>Random</i> <i>search</i> .....	42
Gambar 4. 8 Visualisasi <i>Confusion matrix</i> SVM .....	44
Gambar 4. 9 Visualisasi <i>Confusion matrix</i> SVM dengan <i>Grid search</i>	46
Gambar 4. 10 Visualisasi <i>Confusion matrix</i> SVM dengan <i>Random</i> <i>search</i> .....	48
Gambar 4. 11 Visualisasi <i>Confusion matrix</i> <i>Random forest</i> .....	50
Gambar 4. 12 Visualisasi <i>Confusion matrix</i> <i>Random forest</i> .....	51
Gambar 4. 13 Visualisasi <i>Confusion matrix</i> <i>Random forest</i> dengan <i>Random search</i> .....	53

Gambar 4. 14 Grafik Perbandingan Akurasi Model.....	54
Gambar 4. 15 Grafik Perbandingan Presisi Model.....	55
Gambar 4. 16 Grafik Perbandingan Sensitivitas Model.....	56
Gambar 4. 17 Grafik Perbandingan F-1 <i>Score</i> Model.....	57



## DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1 Daftar Atribut <i>Dataset</i> .....	64
Lampiran 2 Lima Puluh <i>Record</i> Data Semua Atribut .....	73
Lampiran 3 Lima Puluh <i>Record</i> Data dari <i>Features</i> dengan Data Kategorik .....	86
Lampiran 4 Lima Puluh <i>Record</i> Data dari <i>Features</i> dengan Data Numerik .....	88
Lampiran 5 Lima Puluh Data Latih Normalisasi.....	90
Lampiran 6 Lima Puluh Data Uji Normalisasi .....	93



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Salah satu topik utama dalam *data mining* atau *machine learning* adalah klasifikasi (Aryani & Wijayanto, 2021). Klasifikasi merupakan proses identifikasi model atau pola yang mampu menggambarkan dan memisahkan kelas dalam dataset. Tujuannya untuk memungkinkan model melakukan prediksi terhadap objek dengan label kelas yang belum diketahui. Proses pembentukan model bergantung pada analisis dari data pelatihan. Model hasil klasifikasi dapat digunakan untuk melakukan klasifikasi dan meramalkan tren data di masa mendatang (Widjiyati, 2021).

Klasifikasi dapat menyelesaikan beberapa permasalahan diantaranya adalah: *data classification*, *image classification*, *text categorization*, *protein structure predictions* dan lain-lain. Model terbaik dari klasifikasi dapat merepresentasikan hubungan atau pola antar data dengan label. Permasalahan utama dari klasifikasi adalah bagaimana suatu model dapat memisahkan data-data berdasarkan kelas atau labelnya (Fajri & Primajaya, 2023). Beberapa metode yang dapat digunakan untuk menyelesaikan permasalahan klasifikasi diantaranya adalah: K-NN, *Naïve Bayes*, SVM, *random forest*, ADABOOST, dan lain-lain. Setiap metode memiliki kelebihan dan kekurangannya masing-masing. Tolak ukur seberapa baik performa model klasifikasi adalah dengan melakukan beberapa pengujian seperti pengujian nilai akurasi, presisi, *recall* dan *f-measure* (Azis et al., 2020).



Meningkatkan performa model klasifikasi dapat dilakukan dengan mengatur *hyperparameter*. Metode pencarian *hyperparameter* optimal yang dapat digunakan antara lain *grid search*, *random search*, *bayesian search*, algoritma genetika dan *Particle Swarm Optimization* (PSO) (Toha et al., 2022). Penelitian oleh (Sunarya & Haryanti, 2022) dilakukan dengan membandingkan beberapa teknik optimasi *hyperparameter tuning* yaitu *grid search*, *random search* dan *bayesian search* pada algoritma *random forest* menunjukkan bahwa kinerja *random forest* yang dioptimasi baik menggunakan *grid search*, *random search* maupun *bayesian search* mendapatkan nilai akurasi yang lebih baik dibandingkan dengan algoritma *random forest* tanpa optimasi.

Hasil penelitian yang telah disebutkan sebelumnya menunjukkan bahwa penyetelan *hyperparameter* yang optimal memiliki dampak ketika digunakan untuk meningkatkan performa model klasifikasi. Namun, berbagai algoritma klasifikasi yang tersedia memiliki *set hyperparameter* yang berbeda. Oleh karena itu, proses penentuan *hyperparameter* yang optimal melalui teknik *hyperparameter tuning* seperti *grid search*, *random search*, atau *optimasi bayesian* diperlukan untuk mendapatkan *hyperparameter* terbaik yang dapat memaksimalkan performa model.

Tujuan dari penelitian ini adalah menerapkan teknik *hyperparameter tuning grid search* dan *random search* pada beberapa model klasifikasi yang memiliki kemampuan menangani data berdimensi besar, yaitu *K-Nearest neighbors* (K-NN), *Support vector machine* (SVM), dan *random forest*. K-NN unggul dalam mengklasifikasikan data yang besar secara efisien dengan hasil yang

akurat (Aditya Brahmantha et al., 2024). K-NN tidak mempertimbangkan distribusi data, sehingga fleksibel ketika digunakan dalam berbagai kasus (Muhtar et al., 2024). SVM memiliki kinerja akurat dengan kemampuan generalisasi yang baik dan dasar teori yang kuat (Saputra et al., 2023). Sementara *random forest* digunakan karena kelebihanannya dalam mengatasi data pelatihan dalam jumlah besar secara efisien dan mampu menghasilkan *error* yang relatif rendah (Yanti et al., 2023). *Grid search* dan *random search* digunakan pada penelitian ini karena kedua metode tersebut merupakan metode yang paling mudah untuk diimplementasikan dibandingkan metode lain yang lebih kompleks namun metode ini tetap memiliki performa yang baik saat digunakan untuk pencarian *hyperparameter optimal*. Beberapa teknik *hyperparameter tuning* seperti *grid search* dan algoritma genetika, *random search* dan PSO telah dibandingkan pada penelitian sebelumnya dan di dapatkan bahwa *grid search* dan *random search* lebih unggul dalam pencarian *hyperparameter optimal* (Fajri & Primajaya, 2023).

Teknik *hyperparameter tuning grid search* dan *random search* yang diterapkan menggunakan *CrossValidation*. *CrossValidation* digunakan untuk memperkirakan keakuratan kinerja sebuah model (Muliawan & Dazki, 2023), sehingga performa yang baik pada proses pelatihan model juga tercermin pada proses pengujian. Data yang digunakan pada penelitian ini merupakan data yang berasal dari situs *repository UCI Machine Learning* dengan judul "*Predict Students' Dropout and Academic Success*". Dataset tersebut berisi 4424 *record* dan 37 atribut dengan 36 *features* dan 1 *target*. Dataset dikumpulkan dengan tujuan untuk berkontribusi dalam mengurangi angka putus

sekolah dan kegagalan akademik di perguruan tinggi. Mengidentifikasi siswa yang beresiko putus sekolah dan gagal pada tahap awal perkuliahan dengan menggunakan *machine learning* dapat dilakukan sebagai strategi saat pengambilan keputusan bagi pihak-pihak yang terkait (Realinho et al., 2022).

## 1.2 Rumusan Masalah

Penyetelan *hyperparameter* memiliki dampak ketika digunakan untuk optimasi performa model klasifikasi. Namun, setiap algoritma klasifikasi memiliki *set hyperparameter* yang berbeda. Bagaimana menerapkan teknik *hyperparameter tuning grid search* dan *random search* pada algoritma klasifikasi *K-Nearest neighbors* (K-NN), *Support vector machine* (SVM), dan *random forest* untuk mendapatkan *hyperparameter* yang optimal sehingga dapat meningkatkan performa dari masing-masing model?

## 1.3 Batasan Masalah

Batasan masalah yang ditetapkan pada penelitian ini diantaranya adalah:

1. Teknik *hyperparameter tuning* yang dibandingkan adalah *grid search* dan *random search*.
2. *Hyperparameter* yang diselidiki dari setiap algoritma:
  - a. K-NN: jumlah tetangga, skema pembobotan tetangga, dan metrik jarak
  - b. SVM: gamma dan parameter penalti untuk kesalahan klasifikasi
  - c. *Random forest*: jumlah pohon dan kedalaman maksimum setiap pohon

3. Metrik evaluasi yang digunakan untuk mengukur performa model adalah akurasi, presisi, *recall* dan *f1-score*

#### **1.4 Tujuan Penelitian**

Tujuan penelitian berdasarkan rumusan masalah diatas adalah menerapkan teknik *hyperparameter tuning grid search* dan *random search* pada algoritma klasifikasi *K-Nearest neighbors* (K-NN), *Support vector machine* (SVM), dan *random forest* untuk mendapatkan *hyperparameter* yang optimal sehingga dapat meningkatkan performa dari masing-masing model.

#### **1.5 Manfaat Penelitian**

Kontribusi penelitian dengan judul "Optimasi Performa Model Klasifikasi Menggunakan Teknik *Hyperparameter tuning* Pada Algoritma *K-Nearest neighbors*, *Support vector machine*, dan *Random forest*" adalah sebagai berikut:

1. Penerapan model *K-Nearest neighbors* (K-NN), *Support vector machine* (SVM) dan *Random forest*: Penelitian ini berkontribusi dengan membangun model klasifikasi untuk memprediksi kumpulan data keberhasilan akademik. Masalahnya dirumuskan sebagai tugas klasifikasi tiga kategori. Dengan membangun model tersebut, penelitian ini memberikan pemahaman yang lebih baik tentang struktur dan pola dalam *dataset* tersebut.
2. Penerapan teknik *hyperparameter tuning grid search* dan *random search*: Penelitian ini memberikan hasil penerapan dari teknik optimasi, yaitu *grid search* dan *random search*. Dengan menerapkan kedua teknik tersebut, penelitian ini memberikan

wawasan tentang seberapa baik masing-masing teknik dalam mengoptimalkan model yang dibangun.

3. Landasan bagi penelitian selanjutnya: Hasil dan temuan dari penelitian ini dapat menjadi dasar bagi penelitian di masa depan yang mencakup pengembangan metode atau algoritma baru, penggabungan dengan teknik optimasi lainnya atau penyesuaian algoritma yang ada untuk mengatasi tantangan kompleksitas data yang terus berkembang.

# BAB V

## PENUTUP

### 5.1 Kesimpulan

Penelitian ini menerapkan teknik *hyperparameter tuning grid search* dan *random search* pada algoritma klasifikasi *K-Nearest neighbors* (K-NN), *Support vector machine* (SVM), dan *random forest* untuk mendapatkan *hyperparameter* yang optimal menggunakan dataset "*Predict Students' Dropout and Academic Success*". Dataset berisi 4424 *record* dan 37 atribut dengan 36 *features* dan 1 *target*. Target dalam dataset memiliki 3 kelas yaitu: *dropout*, *enrolled* dan *graduate*.

Dilakukan beberapa kali pengujian klasifikasi terhadap algoritma yang digunakan dengan menerapkan teknik *hyperparameter tuning grid search* dan *random search*. *Grid search* dan *random search* berhasil meningkatkan akurasi algoritma K-NN dan *random forest* dibandingkan dengan tanpa optimasi. Sedangkan akurasi model SVM tidak ada peningkatan meskipun menggunakan *hyperparameter* optimal dari hasil *tuning*. Akurasi yang dihasilkan semua model sudah cukup baik, namun selalu ada ruang untuk peningkatan performa dengan mengeksplorasi teknik pemodelan yang lebih canggih mengingat kompleksitas data yang digunakan cukup tinggi hingga dapat mengungguli performa model saat ini.

### 5.2 Saran

Beberapa saran yang dapat dilakukan untuk penelitian selanjutnya sebagai pengembangan dari penelitian ini adalah sebagai berikut:

- a. Dataset yang digunakan memiliki kelas mayoritas hingga menghasilkan *imbalance data*. Saran yang dapat dilakukan adalah mengatasi masalah tersebut baik dengan pendekatan data maupun algoritma. Pada pendekatan data dapat dilakukan dengan menggunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE) atau *ADaptive SYNthetic* (ADASYN). Untuk pendekatan algoritma, dapat menggunakan algoritma yang menyertakan langkah-langkah penyeimbangan data saat melakukan pemodelan.
- b. Dilakukan penggabungan kelas *graduate* dan *enrolled* hingga menghasilkan kelas *biner* untuk memprediksi apakah mahasiswa akan *dropout* atau tidak. Hal ini akan lebih sesuai dengan tujuan awal yaitu mengidentifikasi siswa yang beresiko putus sekolah dan gagal pada tahap awal perkuliahan.



## DAFTAR PUSTAKA

- Adinata, F. D., & Arifin, J. (2022). Klasifikasi Jenis Kelamin Wajah Bermasker Menggunakan Algoritma Supervised Learning. *Jurnal Media Informatika Budidarma*, 6(1), 229–235. <https://doi.org/10.30865/mib.v6i1.3377>
- Aditya Brahmantha, G. P., Utami, E., & Yaqin, A. (2024). Klasifikasi Genre Anime Berdasarkan Sinopsis Menggunakan Algoritma K-Nearest Neighbors. *Jurnal Manajemen Informatika & Sistem Informasi (MISI)*, 7(1), 15–24.
- Andini, E., Faisal, M. R., Herteno, R., Nugroho, R. A., Abadi, F., & Muliadi. (2022). Peningkatan Kinerja Prediksi Cacat Software Dengan Hyperparameter Tuning Pada Algoritma Klasifikasi Deep Forest. *Jurnal MNEMONIC*, 5(2), 119–127. <http://sistemasi.ftik.unisi.ac.id>
- Aryani, Y., & Wijayanto, A. W. (2021). Klasifikasi Pengembalian Radar dari Ionosfer Menggunakan SVM, Naive Bayes dan Random Forest. *Komputika : Jurnal Sistem Komputer*, 10(2), 111–117. <https://doi.org/10.34010/komputika.v10i2.4347>
- Azis, H., Tangguh Admojo, F., & Susanti, E. (2020). Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah. *Techno.COM*, 19(3), 286–294.
- Erdiansyah, U., Irmansyah Lubis, A., & Erwansyah, K. (2022). Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kulit. *Jurnal Media Informatika Budidarma*, 6(1), 208–214. <https://doi.org/10.30865/mib.v6i1.3373>
- Fajri, M., & Primajaya, A. (2023). Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search. *Journal of Applied Informatics and Computing (JAIC)*, 7(1), 10–15. <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Fathurrohman, M., Lambang, R. L., & Susilo, D. D. (2019). Diagnosa Kerusakan Bantalan Bola Menggunakan Metode Support Vector



- Machine. *Mekanika: Majalah Ilmiah Mekanika*, 18(1), 14–21.
- Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. <https://doi.org/10.34148/teknika.v10i1.311>
- Madaerdo Sotarjua, L., & Budhi Santoso, D. (2022). Perbandingan Algoritma KNN, Decision Tree, dan Random Forest Pada Data Imbalanced Class Untuk Klasifikasi Promosi Karyawan. *Jurnal Insitek*, 7(2), 192–200.
- Muhtar, A. F., Wibawa, P. D., & Kallista, M. (2024). Klasifikasi Kualitas Sungai Air Menggunakan Metode Pembelajaran Mesin k-Nearest Neighbour. *E-Proceeding of Engineering*, 11(1), 313–317.
- Muliawan, J., & Dazki, E. (2023). Analisis Sentimen Pemindahan Ibu Kota Negara Indonesia Menggunakan Tiga Algoritma: Naive Bayes, KNN, dan Random Forest. *Jurnal Teknik Informatika (JUTIF)*, 4(5), 1227–1236. <https://doi.org/10.52436/1.jutif.2023.4.5.347>
- Pramudhyta, N. A., & Rohman, M. S. (2024). Perbandingan Optimasi Metode Grid Search dan Random Search dalam Algoritma XGBoost untuk Klasifikasi Stunting. *Jurnal Media Informatika Budidarma*, 8(1), 19–29. <https://doi.org/10.30865/mib.v8i1.6965>
- Realinho, V., Machado, J., Baptista, L., & Martins, M. V. (2022). Predicting Student Dropout and Academic Success. *Data*, 7(146), 1–17. <https://doi.org/10.3390/data7110146>
- Ridho, W. A., Wuryandari, T., & Hakim, A. R. (2023). Perbandingan Metode K-Nearest Neighbor Dan Support Vector Machine Pada Status Penerimaan Bantuan Dari Pemerintah. *Jurnal Gaussian*, 12(3), 372–381. <https://doi.org/10.14710/j.gauss.12.3.372-381>
- Saputra, K., Zuriati, & Suryanto. (2023). Perbandingan Kinerja Fungsi Kernel Algoritma Support Vector Machine Pada Klasifikasi Penyakit Padi. *Jurnal Teknika*, 17(1), 119–131.
- Sudianto, S. (2022). Analisis Kinerja Algoritma Machine Learning Untuk Klasifikasi Emosi. *Building of Informatics, Technology and Science (BITS)*, 4(2), 1027–1034.

<https://doi.org/10.47065/bits.v4i2.2261>

- Sunarya, U., & Haryanti, T. (2022). Perbandingan Kinerja Algoritma Optimasi pada Metode Random Forest untuk Deteksi Kegagalan Jantung. *Jurnal Rekayasa Elektrika*, 18(4), 241–247. <https://doi.org/10.17529/jre.v18i4.26981>
- Swara Darwanto, A. R., Viarindita, T. L., & Widyaningsih, Y. (2021). Analisis Regresi Logistik Binomial dan Algoritma Random Forest pada Proses Pengklasifikasian Penyakit Ginjal Kronis. *Jurnal Statistika Dan Aplikasinya*, 5(1), 1–14.
- Toha, A., Purwono, P., Gata, W., & Toha, A. (2022). Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV. *Buletin Ilmiah Sarjana Teknik Elektro*, 4(1), 12–21. <https://doi.org/10.12928/biste.v4i1.6079>
- Vaganza Junus, C. Z., Tarno, & Kartikasari, P. (2022). Klasifikasi Menggunakan Metode Support Vector Machine dan Random Forest Untuk Deteksi Awal Risiko Diabetes Melitus. *Jurnal Gaussian*, 11(3), 386–396. <https://doi.org/10.14710/j.gauss.11.3.386-396>
- Wibowo, M., Rizieq, M., & Djafar, F. (2023). Perbandingan Metode Klasifikasi Untuk Deteksi Stress Pada Mahasiswa di Perguruan Tinggi. *Jurnal Media Informatika Budidarma*, 7(1), 153–159. <https://doi.org/10.30865/mib.v7i1.5182>
- Widjiyati, N. (2021). Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval. *Jurnal Janitra Informatika Dan Sistem Informasi*, 1(1), 1–7. <https://doi.org/10.25008/janitra.v1i1.118>
- Yanti, C. P., Eva Agustini, N. W., Rahayu GInantra, N. L., & Putri Wulandari, D. A. (2023). Perbandingan Metode K-NN Dan Metode Random Forest Untuk Analisis Sentimen pada Tweet Isu Minyak Goreng di Indonesia. *Jurnal Media Informatika Budidarma*, 7(2), 756–765. <https://doi.org/10.30865/mib.v7i2.5900>