

**PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE  
dan K-NEAREST NEIGHBORS PADA SINYAL TUBUH PEROKOK**

**SKRIPSI**

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



disusun oleh

**MUHAMMAD ALIF MUSTHOFA**

**20.11.3368**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2024**

**PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE  
dan K-NEAREST NEIGHBORS PADA SINYAL TUBUH PEROKOK**

**SKRIPSI**

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



disusun oleh

**MUHAMMAD ALIF MUSTHOFA**

**20.11.3368**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2024**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE dan K-  
NEAREST NEIGHBORS PADA SINYAL TUBUH PEROKOK**

yang disusun dan diajukan oleh

**Muhammad Alif Musthofa**

**20.11.3368**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 20 Desember 2023

**Dosen Pembimbing,**



**Majid Bahardi, S.Kom., M.Eng**  
**NIK. 190302393**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE dan K-  
NEAREST NEIGHBORS PADA SINYAL TUBUH PEROKOK**

yang disusun dan diajukan oleh

**Muhammad Alif Musthofa**

**20.11.3368**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 20 Desember 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Majid Rahardi, S.Kom., M.Eng**  
NIK. 190302393

**Asro Nasiri, Drs, M.Kom**  
NIK. 190302152

**Haryoko, S.Kom., M.Cs.**  
NIK. 190302286

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 20 Desember 2023

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom.**  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Muhammad Alif Musthofa**  
NIM : **20.11.3368**

Menyatakan bahwa Skripsi dengan judul berikut:

**Perbandingan Algoritma Support Vector Machine dan K-Nearest Neighbors Pada Sinyal Tubuh Perokok**

Dosen Pembimbing : **Majid Rahardi, S.Kom., M.Eng**

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 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. 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 ketidakbenaran 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, 20 Desember 2023

Yang Menyatakan,



8B2FAKX795580921

Muhammad Alif Musthofa

## HALAMAN PERSEMBAHAN

Segala puji dan syukur saya persembahkan kepada Allah subhānahu wa ta'āla yang selalu memberikan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan naskah skripsi ini dengan lancar dan barokah. Naskah skripsi ini saya persembahkan kepada:

1. Kedua orang tua saya yang selalu memberikan dukungan berupa doa dan semangat.
2. Universitas Amikom Yogyakarta sebagai tempat menimba ilmu melanjutkan studi saya.
3. Bapak dosen pembimbing yang telah membimbing saya dalam penyelesaian naskah skripsi ini.
4. Teman – teman saya yang telah membantu saya dalam pemberian arahan ketika saya membutuhkan bantuan.
5. Seluruh pihak yang telah memberikan kontribusi sekecil apapun dalam proses penyelesaian naskah skripsi ini yang tidak dapat saya sebutkan satu per satu.



## KATA PENGANTAR

Dengan puji syukur saya persembahkan kepada Allah subhānahu wa ta'āla, Tuhan Yang Maha Esa, atas karunia-Nya dan rido-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Algoritma Support Vector Machine dan K-Nearest Neighbors Pada Sinyal Tubuh Perokok” yang mana naskah skripsi ini saya ajukan sebagai salah satu syarat kelulusan pada Program Studi S1 Informatika. Penulis juga mengucapkan terima kasih yang sebesar – besarnya kepada semua pihak yang telah berkontribusi membantu menyelesaikan naskah skripsi ini. Penulis sangat menyadari bahwa tanpa adanya bantuan, arahan dan dukungan dari berbagai pihak maka naskah skripsi ini tidak akan selesai.

Selanjutnya penulis juga mengucapkan terima kasih atas dukungan dan bantuannya sehingga dapat sampai pada tahap ini kepada :

1. Muhammad Bunyamin dan Tutik Marchamah selaku kedua orang tua yang selalu memberi doa dan dukungan kepada penulis.
2. Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta.
3. Majid Rahardi, S.Kom., M.Eng selaku Dosen pembimbing yang telah membantu dan memberikan arahan dan bimbingan selama proses pengerjaan skripsi.
4. Segenap Dosen Teknik Komputer yang telah memberikan wawasan kepada penulis selama proses menimba ilmu di Universitas Amikom Yogyakarta.
5. Seluruh Teman – teman S1-02 prodi Informatika Angkatan 2020 yang telah memberikan saran kepada penulis untuk menyelesaikan naskah skripsi.
6. Semua pihak yang telah berkontribusi membantu penulis yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa dengan keterbatasan ilmu yang dimiliki saat ini maka skripsi yang dibuat masih jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang sifatnya membangun dari semua pihak demi memperbaiki laporan penelitian ini.

Yogyakarta, 20 Desember 2023

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN .....	xii
DAFTAR ISTILAH.....	xiii
INTISARI.....	xiv
<i>ABSTRACT</i> .....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.5.1 Manfaat secara teori.....	3
1.5.2 Manfaat secara praktisi.....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA .....	5
2.1 Studi Literatur.....	5
2.2 Dasar Teori.....	12
2.2.1 Data Mining.....	12
2.2.2 Machine Learning.....	12



2.2.3	SMOTE	14
2.2.4	Feature Selection	14
2.2.5	Forward Selection	15
2.2.6	Korelasi Pearson	16
2.2.7	Support Vector Machine	17
2.2.8	K-Nearest Neighbors	20
2.2.9	Confuntion Matrix	22
<b>BAB III METODE PENELITIAN</b>		<b>24</b>
3.1	Objek Penelitian	24
3.2	Alur Penelitian	24
3.3	Alat dan Bahan	26
5.3.1	Data penelitian	26
5.3.2	Alat penelitian	28
<b>BAB IV HASIL DAN PEMBAHASAN</b>		<b>29</b>
4.1	Memuat Dataset	29
4.2	Preprosessing Data	30
4.3	Korelasi Pearson	33
4.4	Forward Selection	35
4.5	Split Data	36
4.6	Pemodelan Data	36
4.7	Evaluasi	37
4.7.1	Evaluasi Algoritma SVM	37
4.7.2	Evaluasi Algoritma KNN	40
4.8	Hasil	42
<b>BAB V PENUTUP</b>		<b>44</b>
5.1	Kesimpulan	44
5.2	Saran	45
<b>REFERENSI</b>		<b>46</b>
<b>LAMPIRAN</b>		<b>50</b>

## DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	8
Tabel 2.2. Tingkat Hubungan Koefisien Korelasi	17
Tabel 2.3. Persamaan Rumus Fungsi Kernel	20
Tabel 2.4. Confusion Matrix	23
Tabel 2.5. Persamaan Rumus Penghitung Performa Klasifikasi	23
Tabel 3.1. Atribut Dataset Body Signal of Smoking	27
Tabel 4.1. Split Dataset	36
Tabel 4.2. Hasil Akurasi Model Algoritma SVM	38
Tabel 4.3. Hasil Akurasi Model Algoritma KNN	41



## DAFTAR GAMBAR

Gambar 2.1. Hyperplane Pemisah	18
Gambar 2.2. Hyperplane Dengan Margin Terbesar	18
Gambar 2.3. Tranformasi Data Menggunakan Fungsi Kernel	20
Gambar 2.4. Klasifikasi K-Nearest Neighbors	21
Gambar 3.1. Alur Penelitian	24
Gambar 4.1. Perbandingan Antara Perokok dan Tidak Perokok	30
Gambar 4.2. Hasil Pengecekan Nilai yang Kosong	30
Gambar 4.3. Hasil Pengecekan Variable Bertipe Objek	31
Gambar 4.4. Code Penggunaan Label Encoder	31
Gambar 4.5. Nilai Unik Sebelum Melalui Proses Label Encoder	32
Gambar 4.6. Nilai Unik Sesudah Melalui Proses Label Encoder	32
Gambar 4.7. Imbalance Data	32
Gambar 4.8. Balance Data	33
Gambar 4.9. Visualisasi Heatmap Korelasi Pearson	34
Gambar 4.10. Hasil korelasi Pearson	34
Gambar 4.11. Hasil Forward Selection	35
Gambar 4.12. Perbedaan Antara Forward Selection Original dan SMOTE	35
Gambar 4.13. Confuntion Matrix Metode SVM dengan Data SMOTE, Kernel Linear dan Forward Selection	39
Gambar 4.14. Record Klasifikasi Metode SVM dengan Data SMOTE, Kernel Linear dan Forward Selection	39
Gambar 4.15. Grafik Tingkat Akurasi dari Nilai K Antara 1 - 25	40
Gambar 4.16. Confuntion Matrix Metode KNN dengan Data SMOTE dan Semua Feature	42
Gambar 4.17. Record Klasifikasi Metode KNN dengan Data SMOTE dan Semua Feature	42
Gambar 4.18. Grafik Tingkat Akurasi SVM dan KNN	43
Gambar 4.19. Grafik Record Time SVM dan KNN	43

## DAFTAR LAMPIRAN

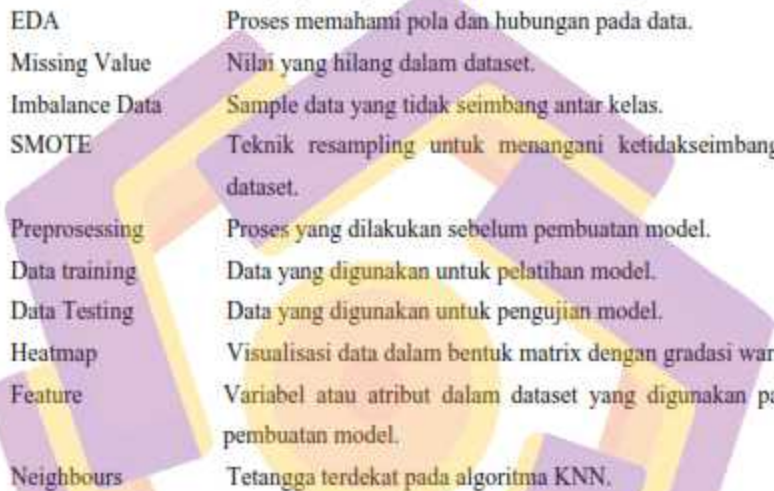
Lampiran 1. Link dataset	50
Lampiran 2. Link Google Colab	50
Lampiran 3. Source Code Memuat Dataset	50
Lampiran 4. Source Code PreProcessing Missing Value	51
Lampiran 5. Source Code PreProcessing Drop Kolom	51
Lampiran 6. Source Code PreProcessing Label Encoder	51
Lampiran 7. Source Code PreProcessing SMOTE Over Sampling	52
Lampiran 8. Source Code Korelasi Pearson	52
Lampiran 9. Source Code Forward Selection	53
Lampiran 10. Source Code Split Data	54
Lampiran 11. Source Code Variabel Independen Semua Feature	54
Lampiran 12. Source Code Variabel Independen Forward Selection	54
Lampiran 13. Source Code Variabel Independen Korelasi Pearson	54
Lampiran 14. Source Code KNN	54
Lampiran 15. Source Code SVM	55
Lampiran 16. Source Code Record Time	55
Lampiran 17. Source Code Confusion Matrix	55
Lampiran 17. Source Code Diagram Venn Output Forward Selection	56
Lampiran 18. Source Code Grafik Tingkat Akurasi SVM dan KNN	56

## DAFTAR LAMBANG DAN SINGKATAN

KNN	K-Nearest Neighbors
SVM	Support Vector Machines
AI	Artificial Intelligence
SRM	Structural Risk Minimization
RBF	Radial Basis Function
CSV	Comma Separated Values
EDA	Exploratory Data Analysis
SMOTE	Synthetic Minority Over-sampling Technique



## DAFTAR ISTILAH



Variabel Dependen	Variabel yang dipengaruhi variabel lain.
Variabel Independen	Variabel yang mempengaruhi variabel lain.
Dataset	Kumpulan data
Cloud	Penyimpanan yang disimpan secara daring.
EDA	Proses memahami pola dan hubungan pada data.
Missing Value	Nilai yang hilang dalam dataset.
Imbalance Data	Sample data yang tidak seimbang antar kelas.
SMOTE	Teknik resampling untuk menangani ketidakseimbangan dataset.
Preprocessing	Proses yang dilakukan sebelum pembuatan model.
Data training	Data yang digunakan untuk pelatihan model.
Data Testing	Data yang digunakan untuk pengujian model.
Heatmap	Visualisasi data dalam bentuk matrix dengan gradasi warna.
Feature	Variabel atau atribut dalam dataset yang digunakan pada pembuatan model.
Neighbours	Tetangga terdekat pada algoritma KNN.



## INTISARI

Data mining merupakan proses penggalian informasi dari data yang besar untuk mendapatkan pola tersembunyi yang mana salah satunya menggunakan machine learning. Machine learning mempunyai beberapa kategori termasuk salah satunya Supervised Learning. Pada Supervised Learning, machine learning dapat melakukan beberapa hal seperti prediksi dan klasifikasi data. Namun terdapat beberapa faktor yang dapat mengurangi performa model. Salah satunya adalah karena terlalu banyak atribut yang digunakan saat pembuatan model. Maka dari itu penggunaan feature selection perlu digunakan untuk menyeleksi atribut yang relevan. Selain itu, pemilihan parameter algoritma yang tepat juga sangat mempengaruhi performa model.

Penelitian ini melakukan klasifikasi pada dataset status perokok menggunakan dua algoritma yaitu Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Saat implementasi metode SVM, peneliti menggunakan beberapa parameter kernel yang berbeda yaitu kernel linear, RBF dan Polynomial. Sedangkan implementasi metode KNN menggunakan parameter nilai  $K = 1 - 25$ . Penelitian ini juga melakukan proses SMOTE untuk mengatasi imbalance class. Selain itu peneliti juga melakukan proses feature selection menggunakan metode forward selection dan Pearson Correlation. Confusion matrix digunakan untuk proses evaluasi performa model yang dibuat.

Hasil evaluasi menunjukkan bahwa metode SVM kernel linear dengan proses SMOTE dan metode forward selection menghasilkan akurasi tertinggi sebesar 76%. Sedangkan akurasi dari metode KNN dengan proses SMOTE dan menggunakan semua feature mendapatkan akurasi tertinggi sebesar 84%. Pada penelitian ini, penggunaan proses SMOTE dapat meningkatkan performa model. Selain itu, penggunaan feature selection juga dapat mempengaruhi hasil akurasi kecuali pada KNN dengan proses SMOTE menggunakan semua feature. Dengan hasil penelitian tersebut diharapkan dapat membantu peneliti lainnya meningkatkan pemahamannya tentang metode pada machine learning sehingga mendapatkan performa model yang baik.

**Kata kunci:** Machine Learning, SMOTE, Feature Selection, Support Vector Machine, K-Nearest Neighbor.

## ***ABSTRACT***

Data mining is the process of extracting information from large data to obtain hidden patterns, one of which uses machine learning. Machine learning has several categories, including Supervised Learning. In Supervised Learning, machine learning can do several things such as prediction and data classification. However, several factors can reduce model performance. One of them is because too many attributes are used when creating the model. Therefore, feature selection needs to be used to select relevant attributes. Apart from that, choosing the right algorithm parameters also greatly influences model performance.

This research classifies the smoker status dataset using two algorithms, namely Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). When implementing the SVM method, researchers used several different kernel parameters, namely linear, RBF, and Polynomial kernels. Meanwhile, the implementation of the KNN method uses parameter values  $K=1-25$ . This research also carries out the SMOTE process to overcome imbalance class. Apart from that, researchers also carried out a feature selection process using the forward selection method and Pearson Correlation. The function matrix is used to evaluate the performance of the model created.

The evaluation results show that the linear kernel SVM method with the SMOTE process and the forward selection method produce the highest accuracy of 76%. Meanwhile, the accuracy of the KNN method with the SMOTE process and using all features obtained the highest accuracy of 84%. In this research, the use of the SMOTE process can improve model performance. Apart from that, the use of feature selection can also affect accuracy results except for KNN where the SMOTE process uses all features. It is hoped that the results of this research can help other researchers improve their understanding of machine learning methods so that they can get good model performance.

***Keyword:*** Machine Learning, SMOTE, Feature Selection, Support Vector Machines, K-Nearest Neighbor.