

**ANALISIS SENTIMEN ULASAN APLIKASI DANA
MENGUNAKAN SUPPORT VECTOR MACHINE DAN
NAÏVE BAYES**

SKRIPSI

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



disusun oleh :

MUHAMAD YULI PAMUNGKAS

22.11.4924

Kepada

**FALKUTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS SENTIMEN ULASAN APLIKASI DANA
MENGUNAKAN SUPPORT VECTOR MACHINE DAN
NAÏVE BAYES**

SKRIPSI

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



Disusun oleh :

MUHAMAD YULI PAMUNGKAS

22.11.4924

Kepada

**FALKUTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN APLIKASI DANA
MENGUNAKAN SUPPORT VECTOR MACHINE DAN
NAÏVE BAYES**

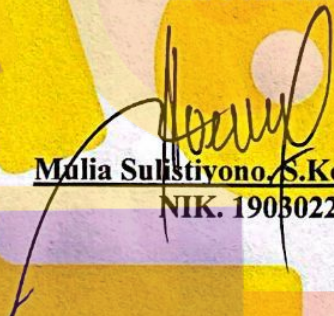
yang disusun dan diajukan oleh

Muhamad Yuli Pamungkas

22.11.4924

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Desember 2025

Dosen Pembimbing,


Mulia Sulistiyono, S.Kom., M.Kom
NIK. 190302248

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN ULASAN APLIKASI DANA
MENGGUNAKAN SUPPORT VECTOR MACHINE DAN
NAÏVE BAYES

yang disusun dan diajukan oleh

Muhamad Yuli Pamungkas

22.11.4924

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Desember 2025

Susunan Dewan Penguji

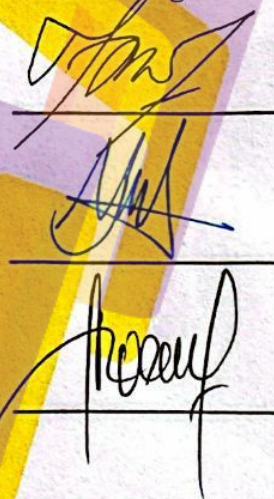
Nama Penguji

Tanda Tangan

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

Yoga Pristyanto, S.Kom., M.Eng
NIK. 190302412

Mulia Sulistiyono, S.Kom., M.Kom
NIK. 190302248



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12-12-2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Muhamad Yuli Pamungkas**
NIM : **22.11.4924**

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Ulasan Aplikasi Dana Menggunakan Support Vector Machine dan Naïve Bayes

Dosen Pembimbing : **Mulia Sulistiyono, 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 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, 12 Desember 2025

Yang Menyatakan,



Muhamad Yuli Pamungkas

HALAMAN PERSEMBAHAN

Pertama saya ucapkan puji syukur kehadirat Allah SWT atas segala nikmat berupa kesehatan, kekuatan dan inspirasi yang sangat banyak dalam proses penyelesaian skripsi ini. Shalawat serta salam selalu terlimpahkan pada Nabi Muhammad SAW. Tiada lembar skripsi yang indah dalam laporan skripsi ini kecuali lembar persembahan.

Untuk karya yang sederhana ini, maka penulis persembahkan untuk :

1. Cinta pertama dan pintu surgaku, Ayahanda Sunar dan Ibunda Siti Chamdanah selaku kedua orang tua saya. Terimakasih atas segala doa dan dukungan yang tidak pernah putus. Memberikan cinta, kasih sayang, doa dan pengorbanan yang mengiringi dalam setiap langkahku untuk menyelesaikan pendidikan ini. Terimakasih sudah berjuang untukku, membesarkanku dan mendidikku sampai mendapat gelar sarjanaku. Semoga Allah SWT senantiasa menjaga kalian sampai melihatku berhasil dengan keputusanku sendiri. Hiduplah lebih lama.
2. Keluarga besar, yang kehangatan dan dukungannya menjadi pelabuhan tenang ditengah segala kebisingan perjalanan. Terimakasih telah mengajarkan arti kebersamaan dan keteguhan.
3. Teruntuk diriku sendiri, yang telah bertahan dan berjuang sejauh ini. Terimakasih sudah tetap melangkah meski sering merasa lelah, ragu, bahkan ingin menyerah. Terimakasih telah mempercayai proses dan tetap menjaga mimpi-mimpi itu tetap hidup. Perjalanan ini menjadi bukti bahwa diriku mampu menghadapi segala rintangan yang datang. Teruslah kuat, teruslah tumbuh dan teruslah menjadi versi terbaik dari dirimu.

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya saya dapat menyelesaikan skripsi ini dengan judul **“Analisis sentimen ulasan aplikasi DANA menggunakan Support Vector machine dan Naive Bayes”**. Penelitian ini disusun untuk memenuhi salah satu persyaratan guna memperoleh gelar sarjana di Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Pada kesempatan ini, saya mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Mulia Sulistiyono, M.Kom, selaku Dosen pembimbing yang telah memberikan bimbingan, arahan, sekaligus motivasi yang telah diberikan selama proses penyusunan skripsi.
2. Tim Dosen Penguji, yang telah memberikan saran, kritik dan masukan berharga dalam penyempurnaan skripsi ini.
3. Semua pihak yang telah membantu dalam penyelesaian skripsi ini, baik secara langsung maupun tidak langsung.
4. Orang tua dan keluarga, yang telah memberikan dukungan moral dan doa selama saya menempuh pendidikan, serta keluarga yang selalu memberikan semangat.

Saya menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang konstruktif sangat saya harapkan demi penyempurnaan di masa yang akan datang.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi perkembangan ilmu pengetahuan.

Yogyakarta, 12 Desember 2025



Muhamad Yuli Pamungkas

DAFTAR ISI

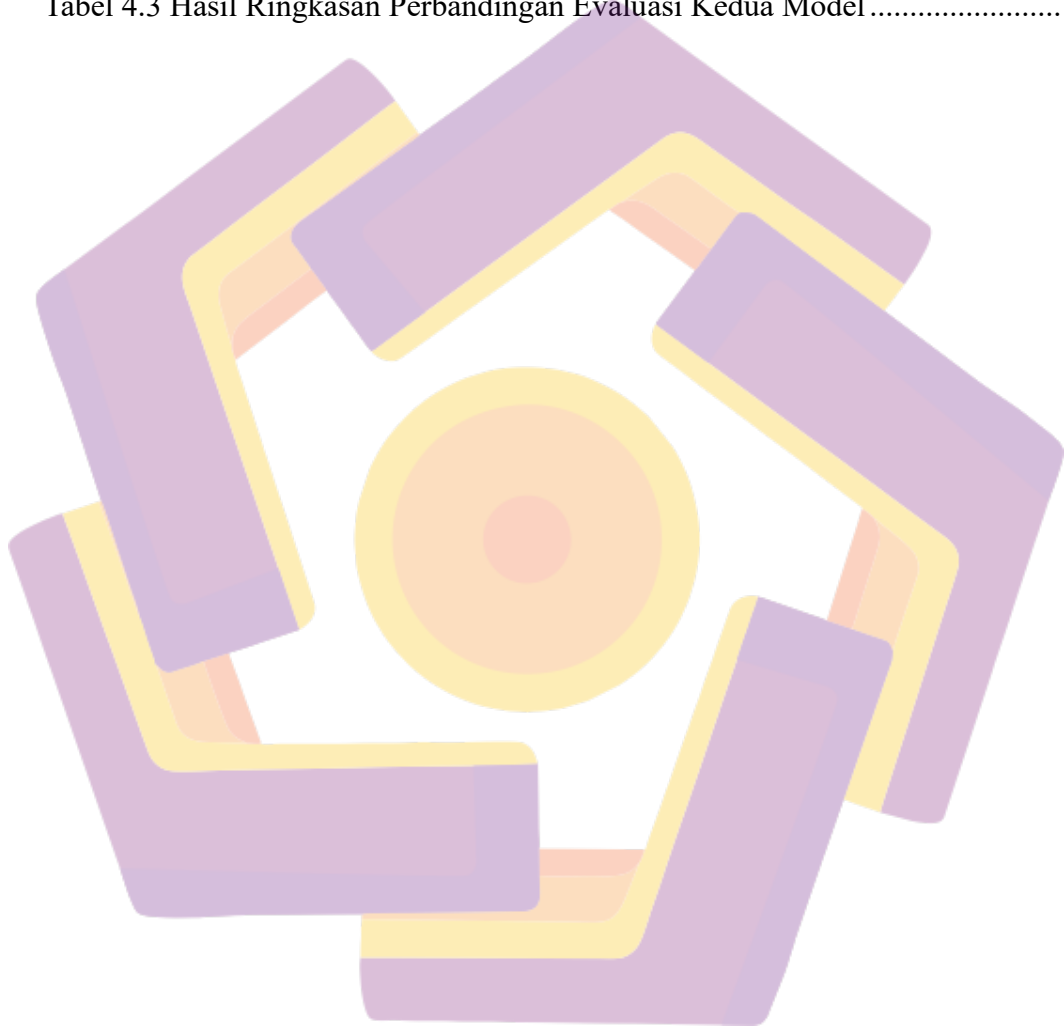
HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xiii
DAFTAR LAMBANG DAN SINGKATAN.....	xiv
DAFTAR ISTILAH	xv
INTISARI.....	xvi
<i>ABSTRACT</i>	xvii
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.....	5
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori	17
2.2.1 Analisis Sentimen.....	17
2.2.2 <i>Natural Language Processing (NLP)</i>	17
2.2.3 Scraping Data	18
2.2.4 Teks Preprocessing	18

2.2.5	TF-IDF Term Frequence and Inverse Document Frequence..	20
2.2.6	Machine Learning.....	21
2.2.7	Support Vector Machine	21
2.2.8	Kernel RBF.....	23
2.2.9	Naive Bayes.....	23
2.2.10	Naive Bayes Multinomial.....	24
2.2.11	Hyperparameter Tuning.....	24
2.2.12	Synthetic Minority Over-Sampling Technique (SMOTE)	26
2.2.13	Evaluasi Kinerja Model.....	26
BAB III METODE PENELITIAN.....		29
3.1	Objek Penelitian	29
3.2	Alur Penelitian.....	29
3.2.1	Pengumpulan Data	30
3.2.2	Pelabelan Sentimen	31
3.2.3	Preprocessing Data	31
3.2.4	Pembagian Dataset	32
3.2.5	Vektorisasi TF-IDF.....	32
3.2.6	Penyeimbang Data SMOTE	33
3.2.7	Pelatihan Model Support Vector Machine.....	33
3.2.8	Pelatihan Model Naive Bayes	33
3.2.9	Evaluasi Model Support Vector Machine dan Naive Bayes...	34
3.3	Alat dan Bahan	34
BAB IV HASIL DAN PEMBAHASAN.....		36
4.1	Dataset yang digunakan	36
4.2	Pelabelan	37
4.3	Preprocessing Data.....	38
4.3.1	<i>Cleaning</i>	38
4.3.2	<i>Case Folding</i>	39
4.3.3	<i>Stopword Removal</i>	40
4.3.4	<i>Tokenizing</i>	41
4.3.5	<i>Steamming</i>	42

4.4	Pembersihan Data Duplikat.....	43
4.5	<i>Exploratory Data Analysis (EDA)</i>	44
4.6	<i>WordCloud</i> Positif.....	45
4.7	<i>WordCloud</i> Negatif.....	46
4.8	Pembagian Data.....	47
4.9	<i>Ekstraksi Fitur TF-IDF</i>	48
4.10	<i>SMOTE (Synthetic Minority Over-Sampling Technique)</i>	49
4.11	Hyperparameter Tuning <i>Support Vector Machine Kernel RBF</i>	51
4.12	Hyperparameter Tuning <i>Naive bayes</i>	52
4.13	Pelatihan dan Evaluasi Model <i>Support Vector Machine Kernel RBF</i> 53	
4.13.1	Pelatihan <i>Support Vector Machine Kernel RBF</i>	53
4.13.2	<i>Confusion Matrix</i> Model <i>Support Vector Machine Kernel RBF</i>	56
4.14	Pelatihan dan Evaluasi Model <i>Naive Bayes</i>	57
4.14.1	Pelatihan Model <i>Naive Bayes</i>	57
4.14.2	<i>Confusion Matrix</i> Model <i>Naive Bayes</i>	59
4.15	Perbandingan Hasil Evaluasi Kedua Model.....	60
4.15.1	Hasil Evaluasi <i>Support Vector Machine</i> kernel RBF dan <i>Naive Bayes</i>	60
4.15.2	Ringkasan dan Kesimpulan Evaluasi Model.....	61
BAB V PENUTUP.....		62
5.1	Kesimpulan.....	62
5.1	Saran.....	63
REFERENSI		64
LAMPIRAN.....		67

DAFTAR TABEL

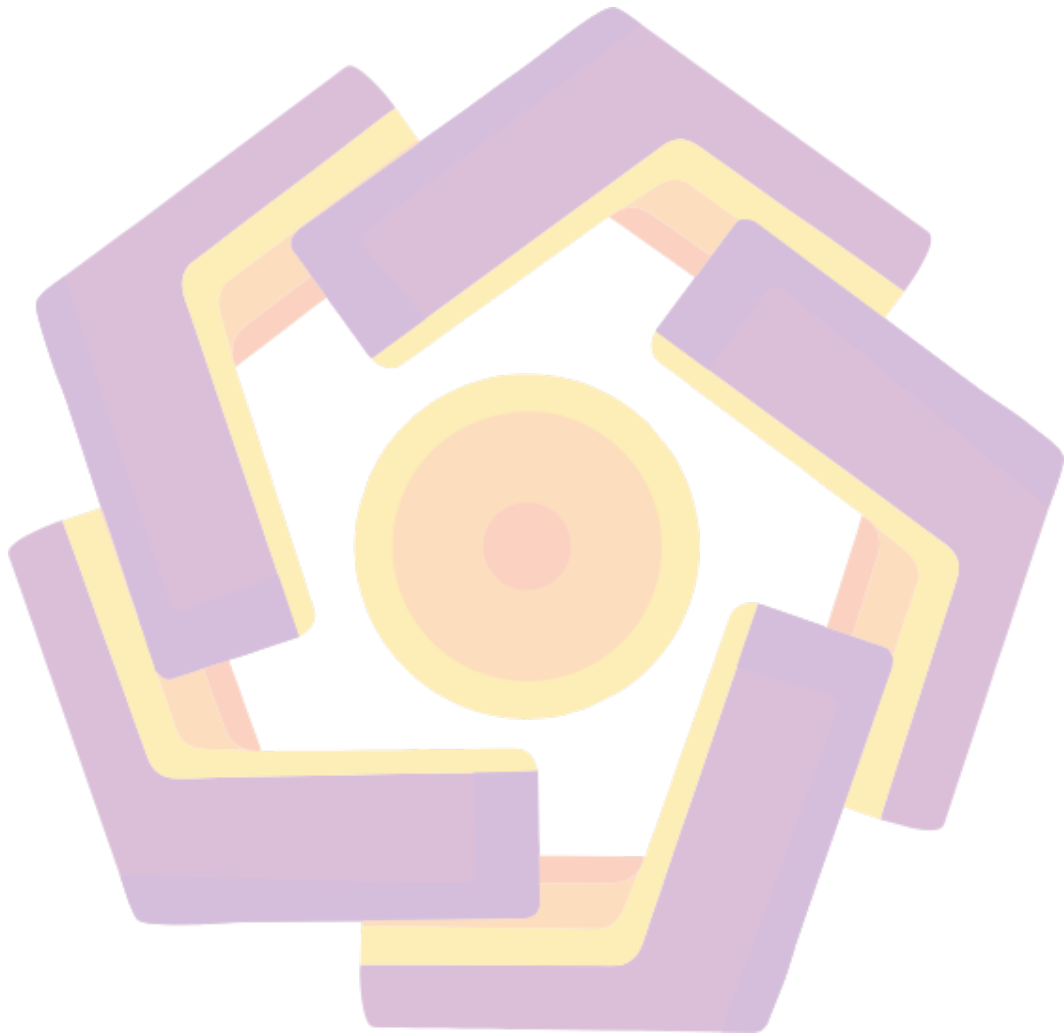
Tabel 2.1 Keaslian Penelitian	10
Tabel 3.1 Hardware	34
Tabel 3.2 Software	35
Tabel 4.1 Hasil <i>Training</i>	60
Tabel 4.2 Hasil <i>Testing</i>	60
Tabel 4.3 Hasil Ringkasan Perbandingan Evaluasi Kedua Model	62



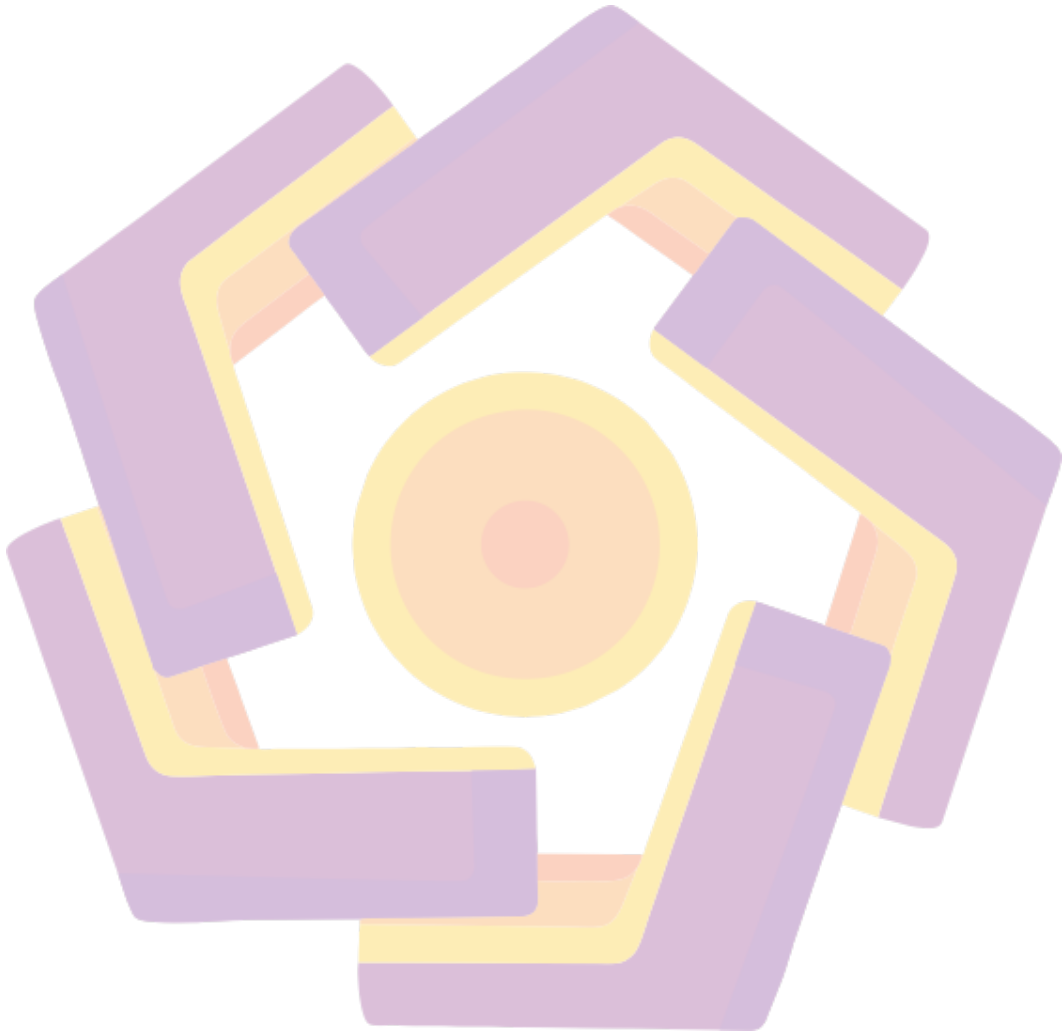
DAFTAR GAMBAR

Gambar 2.1 Konsep SVM Mencari Hyperlane terbaik [13]	22
Gambar 2.2 Representasi <i>Visual Grid-Search</i> [25]	25
Gambar 2.3 <i>Confosion Matrix</i> [27]	26
Gambar 3.1 Alur Penelitian	30
Gambar 4.1 Proses Scraping yang Dipublikasikan oleh Kundi di LinkedIn	36
Gambar 4.2 Tampilan Dataset Hasil <i>Scraping</i>	36
Gambar 4.3 <i>Source Code</i> Pelabelan	37
Gambar 4.4 Tampilan Hasil Palebelan	38
Gambar 4.5 <i>Source Code Cleaning</i>	38
Gambar 4.6 Tampilan Hasil Cleaning	39
Gambar 4.7 <i>Source Code Case Folding</i>	39
Gambar 4.8 Tampilan Hasil Case Folding	40
Gambar 4.9 <i>Source Code Stopword Removal</i>	40
Gambar 4.10 Tampilan Hasil <i>Stopword Removal</i>	41
Gambar 4.11 <i>Source Code Tokenizing</i>	41
Gambar 4.12 Tampilan Hasil <i>Tokenizing</i>	41
Gambar 4.13 <i>Source Code Steaming</i>	42
Gambar 4.14 Tampilan Hasil <i>Steaming</i>	43
Gambar 4.15 <i>Source Code Data Duplikat</i>	43
Gambar 4.16 <i>Source Code EDA</i>	44
Gambar 4.17 Tampilan Hasil <i>EDA</i>	44
Gambar 4.18 <i>Source Code WordCloud</i> Positif	45
Gambar 4.19 Tampilan hasil <i>WordCloud</i> Positif	45
Gambar 4.20 <i>Source Code WordCloud</i> Negatif	46
Gambar 4.21 Tampilan hasil <i>WordCloud</i> Negatif	46
Gambar 4.22 <i>Source Code Pembagian data</i>	47
Gambar 4.23 <i>Source Code Pembagian data</i>	47
Gambar 4.24 Tampilan Hasil Pembagian Data	47
Gambar 4.25 <i>Source Code TF-IDF</i>	48
Gambar 4.26 <i>Source Code TF-IDF</i>	49
Gambar 4.27 Tampilan Hasil TF-IDF	49
Gambar 4.28 <i>Source Code SMOTE</i>	50
Gambar 4.29 Tampilan Hasil Sebelum <i>SMOTE</i>	50
Gambar 4.30 <i>Source Code Sesudah SMOTE</i>	50
Gambar 4.31 Tampilan Hasil Sesudah <i>SMOTE</i>	51
Gambar 4.32 Tampilan Hasil <i>SMOTE</i>	51
Gambar 4.33 <i>Source Code Hyperparameter Tuning SVM</i>	52
Gambar 4.34 <i>Source Code Hyperparameter Tuning Naive Bayes</i>	53
Gambar 4.35 <i>Source Code Pelatihan Model SVM Kernel RBF</i>	54
Gambar 4.36 <i>Source Code Hasil Model SVM Kernel RBF</i>	55
Gambar 4. 37 Visualisasi <i>Confusion Matrix</i> Model <i>Support Vector Machine Kernel RBF</i>	56
Gambar 4.38 <i>Source Code Pelatihan Naive Bayes</i>	58

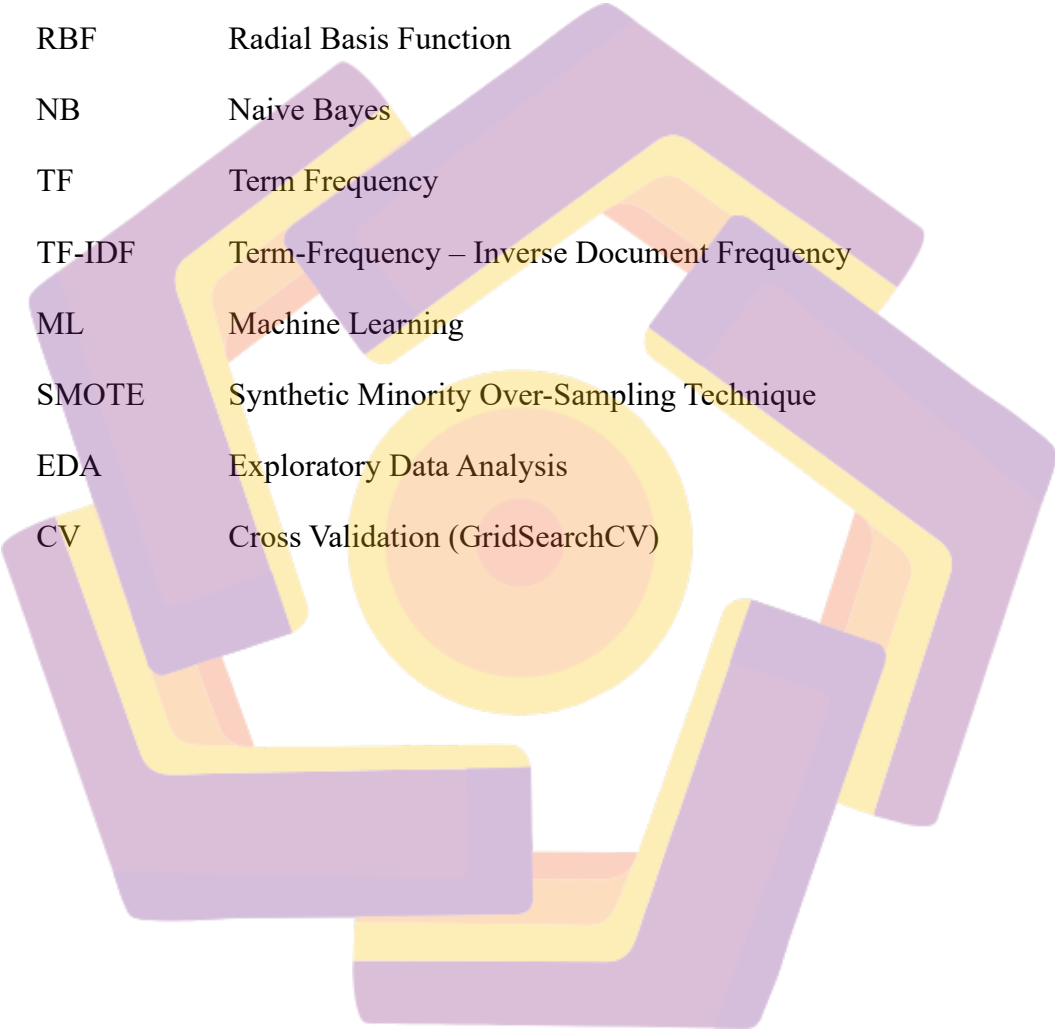
Gambar 4.39 <i>Source Code Hasil Naive Bayes</i>	58
Gambar 4.40 Visualisasi <i>Confusion Matrix Model Naive Bayes Multinomial</i>	59



DAFTAR LAMPIRAN

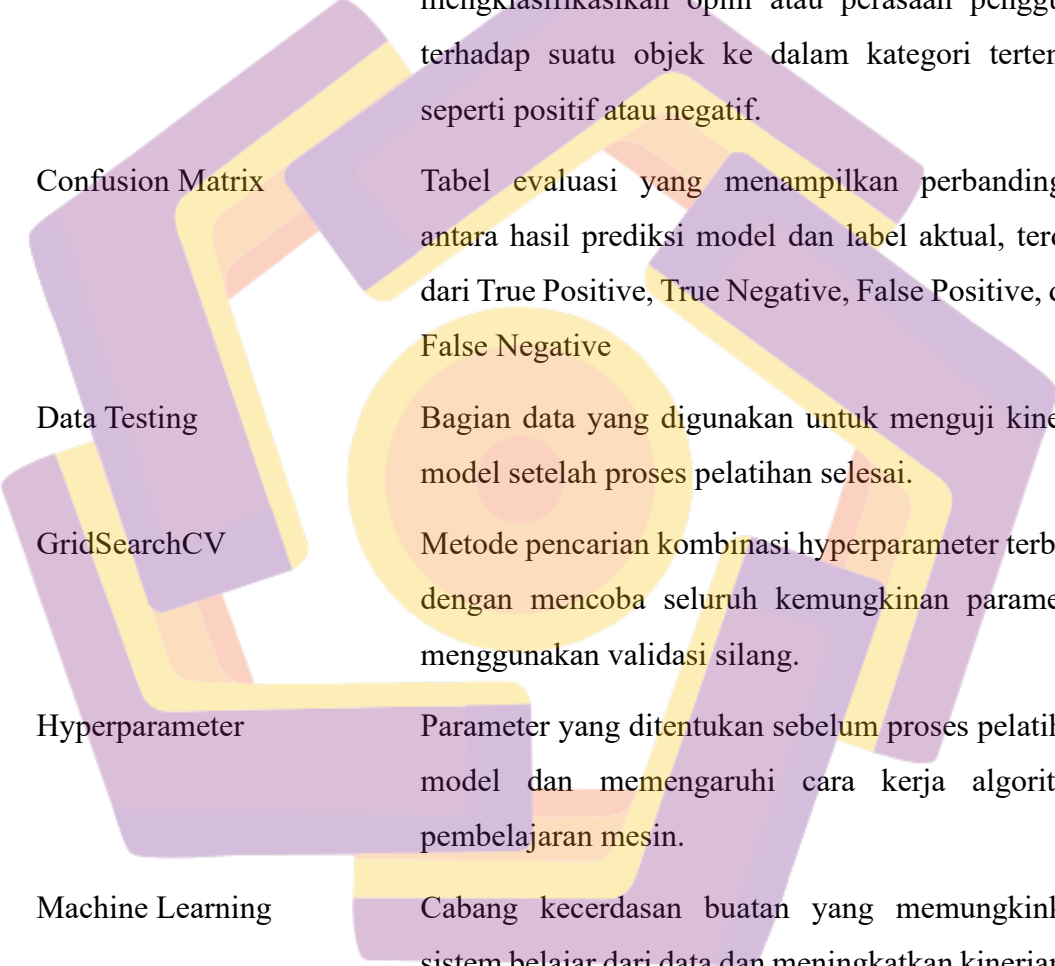


DAFTAR LAMBANG DAN SINGKATAN



SVM	Support Vector Machine
NLP	Natural Language Processing
RBF	Radial Basis Function
NB	Naive Bayes
TF	Term Frequency
TF-IDF	Term-Frequency – Inverse Document Frequency
ML	Machine Learning
SMOTE	Synthetic Minority Over-Sampling Technique
EDA	Exploratory Data Analysis
CV	Cross Validation (GridSearchCV)

DAFTAR ISTILAH



Akurasi (Accuracy)	Ukuran kinerja model yang menunjukkan persentase prediksi yang benar dibandingkan dengan seluruh data yang diuji.
Analisis Sentimen	Proses untuk mengidentifikasi dan mengklasifikasikan opini atau perasaan pengguna terhadap suatu objek ke dalam kategori tertentu, seperti positif atau negatif.
Confusion Matrix	Tabel evaluasi yang menampilkan perbandingan antara hasil prediksi model dan label aktual, terdiri dari True Positive, True Negative, False Positive, dan False Negative
Data Testing	Bagian data yang digunakan untuk menguji kinerja model setelah proses pelatihan selesai.
GridSearchCV	Metode pencarian kombinasi hyperparameter terbaik dengan mencoba seluruh kemungkinan parameter menggunakan validasi silang.
Hyperparameter	Parameter yang ditentukan sebelum proses pelatihan model dan memengaruhi cara kerja algoritma pembelajaran mesin.
Machine Learning	Cabang kecerdasan buatan yang memungkinkan sistem belajar dari data dan meningkatkan kinerjanya tanpa diprogram secara eksplisit.

INTISARI

Penelitian ini bertujuan untuk menganalisis dan membandingkan hasil kerja algoritma Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF) serta Multinomial Naïve Bayes dalam mengklasifikasikan sentimen dari 5.200 ulasan aplikasi DANA yang berbahasa Indonesia. Ulasan-ulasan tersebut dikumpulkan melalui teknik web scraping di platform Google Play Store. Penelitian ini fokus pada dua jenis sentimen, yaitu sentimen positif dan negatif. Sebelum dilakukan klasifikasi, terlebih dahulu dilakukan beberapa proses preprocessing teks, seperti cleaning, case folding, menghilangkan stopwords, tokenizing, dan stemming agar data teks menjadi lebih bersih dan standar. Data kemudian diubah menjadi representasi TF-IDF sehingga teks bisa diubah ke dalam bentuk angka yang menunjukkan tingkat kepentingan kata dalam dokumen. Karena data sentimen tidak seimbang, maka digunakan metode SMOTE (Synthetic Minority Over-Sampling Technique) untuk menyeimbangkan distribusi kelas dan mencegah bias dalam prediksi sentimen. Proses pelatihan model juga dilakukan dengan menambahkan teknik hyperparameter tuning menggunakan GridSearchCV pada kedua algoritma untuk mencari parameter terbaik dan meningkatkan akurasi prediksi. Evaluasi kinerja dilakukan dengan menggunakan metrik Accuracy, Precision, Recall, dan F1-Score. Hasilnya menunjukkan bahwa SVM dengan kernel RBF memberikan performa lebih baik dibandingkan Multinomial Naïve Bayes. SVM RBF memiliki Accuracy sebesar 87,78%, Precision 88,18%, Recall 87,84%, dan F1-Score 87,91%. Sementara itu, Multinomial Naïve Bayes memiliki Accuracy 85,54%, Precision 89,05%, Recall 85,43%, dan F1-Score 86,18%. Dengan demikian, algoritma SVM kernel RBF lebih unggul dan disarankan untuk digunakan dalam analisis sentimen pada bahasa Indonesia, terutama dalam aplikasi keuangan digital seperti DANA karena memberikan hasil yang lebih akurat dan seimbang untuk kedua kelas sentimen.

Kata kunci: Analisis Sentimen, TF-IDF, SMOTE, SVM-RBF, Multinomial Naïve Bayes, GridSearchCV, Google Play Store, DANA.

ABSTRACT

This study aims to analyze and compare the performance of the Support Vector Machine (SVM) algorithm with a Radial Basis Function (RBF) kernel and the Multinomial Naïve Bayes algorithm in classifying sentiment from 5,200 Indonesian-language user reviews of the DANA application. The reviews were collected using web scraping techniques from the Google Play Store platform. This research focuses on two sentiment classes, namely positive and negative. Prior to classification, several text preprocessing steps were applied, including text cleaning, case folding, stopwords removal, tokenization, and stemming to ensure the data were standardized and noise-free. The text data were then transformed into numerical representations using TF-IDF to capture the importance of each term within the documents. Due to class imbalance in the sentiment data, the Synthetic Minority Over-Sampling Technique (SMOTE) was employed to balance the class distribution and reduce prediction bias. Model training was enhanced by applying hyperparameter tuning using GridSearchCV for both algorithms to determine the optimal parameter settings and improve prediction accuracy. Performance evaluation was conducted using Accuracy, Precision, Recall, and F1-Score metrics. The results show that the SVM with RBF kernel outperformed the Multinomial Naïve Bayes algorithm, achieving an Accuracy of 87.78%, Precision of 88.18%, Recall of 87.84%, and an F1-Score of 87.91%. Meanwhile, Multinomial Naïve Bayes achieved an Accuracy of 85.54%, Precision of 89.05%, Recall of 85.43%, and an F1-Score of 86.18%. Therefore, the SVM with RBF kernel is recommended for sentiment analysis tasks in the Indonesian language, particularly for digital financial applications such as DANA, as it provides more accurate and balanced performance across sentiment classes.

Keyword: *Sentiment Analysis, TF-IDF, SMOTE, SVM-RBF, Multinomial Naïve Bayes, GridSearchCV, Google Play Store, DANA*