

**Optimasi Analisis Sentimen Terhadap Aplikasi Pencari Kerja
Menggunakan Perbandingan Algoritma Machine Learning**



Skripsi

Zaki Prastiansyah

00000048831

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**Optimasi Analisis Sentimen Terhadap Aplikasi Pencari Kerja Menggunakan
Perbandingan Algoritma Machine Learning**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer

ZAKI PRASTIANSYAH

00000048831

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

i

Optimasi Analisis Sentimen ..., Zaki Prastiansyah, Universitas Multimedia Nusantara

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : ZAKI PRASTIANSYAH

Nomor Induk Mahasiswa : 00000048831

Program Studi : SISTEM INFORMASI

Skripsi dengan judul:

Optimasi Analisis Sentimen Terhadap Aplikasi Pencari Kerja Menggunakan Perbandingan Algoritma Machine Learning

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 5 Juni 2025



(ZAKI PRASTIANSYAH)

HALAMAN PERSETUJUAN

SKRIPSI dengan judul

Optimasi Analisis Sentimen Terhadap Aplikasi Pencari Kerja Menggunakan Perbandingan Algoritma Machine Learning

Oleh

Nama : ZAKI PRASTIANSYAH
NIM : 00000048831
Program Studi : SISTEM INFORMASI
Fakultas : TEKNIK INFORMATIKA

Telah disetujui untuk diajukan pada
Sidang Ujian Skripsi Universitas Multimedia Nusantara
Tangerang, 5 Juni 2025

Pembimbing



Rudi Sutomo, S.Kom., M.Si., M.Kom.
0222057501

Ketua Sistem Informasi



Ririn Ikana Desanti S.Kom., M.Kom

HALAMAN PENGESAHAN

Skripsi dengan judul

Optimasi Analisis Sentimen Terhadap Aplikasi Pencari Kerja Menggunakan Perbandingan Algoritma Machine Learning

Oleh

Nama : Zaki Prastiansyah
NIM : 00000048831
Program Studi : Sistem Informasi
Fakultas : Teknik & Infotmatika

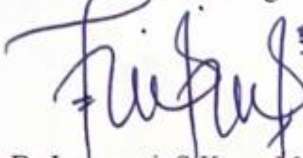
Telah diujikan pada hari Senin 16 Juni 2025

Pukul 10.00 s.d 12.00 dan dinyatakan

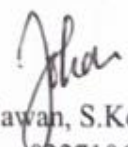
LULUS

Dengan susunan penguji sebagai berikut.

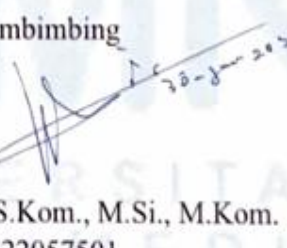
Ketua Sidang


Dr. Irmawati, S.Kom., M.M.S.I.
0805097703

Penguji


Johan Setiawan, S.Kom., M.M., MBA
0327106402

Pembimbing


Rudi Sutomo, S.Kom., M.Si., M.Kom.
0222057501

Ketua Prodi Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.
0313058001

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Zaki Prastiansyah

NIM : 00000048831

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : Optimasi Analisis Sentimen Terhadap Aplikasi Pencari Kerja Menggunakan Perbandingan Algoritma Machine Learning

Menyatakan dengan sesungguhnya bahwa saya bersedia* (pilih salah satu):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tangerang, 23 Juni 2025



(Zaki Prastiansyah)

KATA PENGANTAR

(Kata Pengantar dapat dikembangkan dan harus meliputi ucapan rasa syukur, tujuan pembuatan tugas akhir, ucapan terima kasih, dan harapan pada hasil Tugas Akhir ini.)

Mengucapkan terima kasih

1. Bapak Dr. Andrey Andoko, M.Sc, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng Niki Prastomo, S.T., M.Sc, selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Rudi Sutomo, S.Kom., M.Si., M.Kom, sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga saya dapat menyelesaikan tugas akhir ini.
6. Pacar dan Sahabat saya yang telah memberikan dukungan secara mental dan moral, sehingga tugas akhir ini dapat diselesaikan

Semoga karya ilmiah ini dapat bermanfaat untuk yang membaca dan hasil karya ini dapat selesai dengan hasil akhir dan nilai yang memuaskan.

Tangerang, 5 Juni 2025



Zaki Prastiansyah

Optimasi Analisis Sentimen Terhadap Aplikasi Pencari Kerja Menggunakan Perbandingan Algoritma Machine Learning

Zaki Prastiansyah

ABSTRAK

Pertumbuhan aplikasi pencari kerja seperti KitaLulus, JobStreet, dan LinkedIn di Indonesia memberikan kemudahan bagi masyarakat dalam mencari pekerjaan yang sesuai. Namun, keberhasilan aplikasi tersebut sangat bergantung pada kepuasan dan pengalaman pengguna. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna dari ketiga aplikasi tersebut yang diambil dari Google Play Store.

Data yang digunakan berjumlah masing-masing 1000 ulasan dan dianalisis menggunakan metode Knowledge Discovery in Databases (KDD) dengan tahapan mulai dari pengumpulan data, preprocessing, ekstraksi fitur menggunakan TF-IDF, hingga klasifikasi sentimen menggunakan lima algoritma machine learning: Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, dan Random Forest. Evaluasi dilakukan dengan menggunakan metrik akurasi, precision, recall, dan F1-score.

Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dengan akurasi tertinggi mencapai 91,50% dalam menganalisis sentimen aplikasi KitaLulus. Optimasi menggunakan Grid Search mampu meningkatkan akurasi secara signifikan, sementara teknik SMOTE menunjukkan hasil yang kurang optimal. Penelitian ini memberikan kontribusi dalam pemilihan model machine learning terbaik untuk analisis sentimen serta memberikan gambaran kepada pengembang aplikasi untuk meningkatkan kualitas layanan berdasarkan persepsi pengguna.

Kata kunci: Analisis Sentimen, Machine Learning, Random Forest, TF-IDF, KDD

Optimization of Sentiment Analysis on Job Search Applications Using a Comparison of Machine Learning Algorithms

Zaki Prastiansyah

ABSTRACT (English)

The growth of job search applications such as KitaLulus, JobStreet, and LinkedIn in Indonesia has made it easier for people to find jobs that match their skills. However, the success of these applications largely depends on user satisfaction and experience. This study aims to conduct sentiment analysis on user reviews of the three applications, collected from the Google Play Store.

A total of 1000 reviews for each application were analyzed using the Knowledge Discovery in Databases (KDD) framework, which includes data collection, preprocessing, feature extraction using TF-IDF, and sentiment classification using five machine learning algorithms: Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, and Random Forest. The models were evaluated using accuracy, precision, recall, and F1-score.

The results show that the Random Forest algorithm achieved the best performance, with the highest accuracy of 91.50% on KitaLulus reviews. Optimization using Grid Search significantly improved accuracy, while the SMOTE technique showed less optimal results. This study contributes to identifying the most effective machine learning model for sentiment analysis and provides insights for application developers to improve service quality based on user perceptions.

Keywords: KDD, Machine Learning, Random Forest, Sentiment Analysis, TF-IDF

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	iv
KATA PENGANTAR.....	v
ABSTRAK.....	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	i
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	6
1.3 Batasan Masalah.....	6
1.4 Tujuan dan Manfaat Penelitian.....	7
1.4.1 Tujuan Penelitian.....	7
1.4.2 Manfaat Penelitian.....	7
1.5 Sistematika Penulisan.....	8
BAB II LANDASAN TEORI.....	9
2.1 Penelitian Terdahulu.....	9
2.2 Teori Penelitian.....	13
2.2.1 Analisis Sentimen.....	13
2.2.2 Teori Bayes.....	14
2.2.3 Text Mining.....	15
2.2.4 Confusion Matrix.....	16
2.2.5 Web Scraping.....	18
2.2.6 TF-IDF.....	18
2.3 Framework dan Algoritma Penelitian.....	19

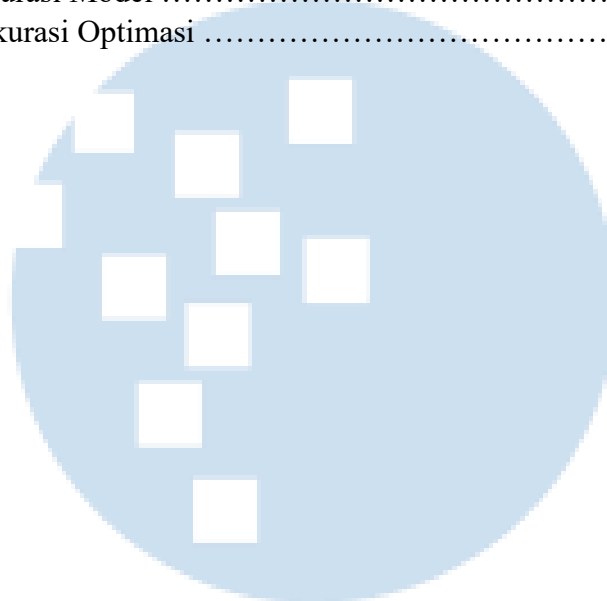
2.3.1 Knowledge Discovery in Databases (KDD)	19
2.3.2 Naïve Bayes	20
2.3.3 Support Vector Machine (SVM).....	21
2.3.4 K-Nearest Neighbors (KNN)	21
2.3.5 Decision Tree	22
2.3.6 Random Forest	23
2.4 Tools dan Software Penelitian.....	23
2.4.1 Microsoft Excel.....	24
2.4.2 Python	25
2.4.3 Google Colabpratory	26
BAB III METODELOGI PENELITIAN	27
3.1 Gambaran Umum Objek Penelitian.....	27
3.2 Metode Penelitian	27
3.3 Variabel Penelitian	31
3.3.1 Variabel Independen.....	31
3.3.2 Variabel Dependen	31
3.4 Teknik Pengumpulan Data.....	31
3.5 Teknik Analisis Data	33
BAB IV ANALISIS DAN HASIL PENELITIAN	35
4.1 Data Selection	35
4.2 Prapemrosesan Data (<i>Preprocessing</i>)	39
4.2.1 Pemeriksaan Missing Value	40
4.2.2 Pemeriksaan data duplikat.....	41
4.2.3 Pembersihan Simbol dan Elemen	43
4.2.4 Case Folding	45
4.2.5 Normalisasi Teks	46
4.2.6 Stopword Removal.....	47
4.2.7 Tokenization	48
4.2.8 Stemming	50
4.2.9 Translaction.....	53
4.2.10 Analisis Sentimen	55

Sentimen Kitalulus	55
Sentimen Jobstreet	60
Sentimen LinkedIn	64
4.3 Data Transformation	68
4.4 Data Mining	68
4.4.1 Naïve Bayes	69
4.4.2 KNN	78
4.4.3 Decision Tree	87
4.4.4 SVM (Support Vector Machine)	96
4.4.5 Random Forest	105
4.5 Evaluasi	114
4.5.1 Perbandingan akurasi model	114
4.5.2 Perbandingan Optimasi	114
4.6 Pembahasan Hasil Penelitian	116
BAB V SIMPULAN DAN SARAN	117
5.1 Simpulan	117
5.2 Saran	117
DAFTAR PUSTAKA	118
LAMPIRAN	122



DAFTAR TABEL

Tabel 1 1 Perbandingan Aplikasi	2
Tabel 2 1 Penelitian terdahulu	Error! Bookmark not defined.
Tabel 3 1Perbandingan KDD dan CRISP-DM. Sumber:[42]	Error! Bookmark not defined.
Tabel 3 2 Perbandingan Google Colab dengan Selenium[28]	Error! Bookmark not defined.
Tabel 3 3 Model Algoritma	Error! Bookmark not defined.
Tabel 4 51Perbandingan Akurasi Model	117
Tabel 4 5 2Perbandingan Akurasi Optimasi	117



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 2 1 Analisis Sentimen[23]	14
Gambar 2 2 Confusion Matrix.....	16
Gambar 2 3 Naive Bayes[32]	20
Gambar 2 4 Support Vector Machine[34]	21
Gambar 2 5 K-Nearest Neighbour[36]	22
Gambar 2 6 Microsoft Excel	24
Gambar 2 7 Python.....	25
Gambar 2 8 Google Colaboratory	26
Gambar 4.1 1 Install library google play scraper	35
Gambar 4.1 2 Import google play scraper pandas dan numpys	36
Gambar 4.1 3 Input URL kitalulus	36
Gambar 4.1 4 Tabel ulasan kita lulus	37
Gambar 4.1 5 Kode untuk menunjukan hasil data yang didapat.....	37
Gambar 4.1 6 pratinjau data ulasan kitalulus(username,score,at,content)	37
Gambar 4.1 7Data ulasan kitalulus yang diurutkan dari terbaru	38
Gambar 4.1 8 Pemilihan kolom yang bernilai sentiment	38
Gambar 4.1 9 Input URL Jobstreet untuk mengscraping data ulasan	39
Gambar 4.1 10 Input URL aplikasi LinkedIn	39
Gambar 4.2.1 1 Pemeriksaan missing value data aplikasi Kitalulus	40
Gambar 4.2.1 2 Pemeriksaan missing value pada aplikasi Jobstreet	40
Gambar 4.2.1 3 Pemeriksaan missing value LinkedIn	41
Gambar 4.2.2 1 Pengecekan Duplikat Kitalulus	45
Gambar 4.2.2 2 Pengecekan Duplikat Jobstreet	46
Gambar 4.2.2 3 Pengecekan Duplikat LinkedIn.....	46
Gambar 4.2.3 1 Pembersihan Simbol dan Elemen Dataset KitaLulus.....	46
Gambar 4.2.3 2 Pembersihan Simbol dan Elemen Dataset Jobstreet	47
Gambar 4.2.3 3 Pembersihan Simbol dan Elemen Dataset LinkedIn	47
Gambar 4.2.4 1 Case Folding KitaLulus	48
Gambar 4.2.4 2 Case Folding Jobstreet	48
Gambar 4.2.4 3 Case Folding LinkedIn	49
Gambar 4.2.7 1 Coding tokenized KitaLulus	52
Gambar 4.2.7 2 Coding Tokenized Jobstreet.....	53
Gambar 4.2.7 3 Coding Tokenized LinkedIn	53
Gambar 4.2.8 1 Coding Stemming KitaLulus	54
Gambar 4.2.8 2 Coding Stemming Jobstreet	55
Gambar 4.2.8 3 Coding Stemming LinkedIn	55
Gambar 4.2.9 1 Code Translator KitaLulus.....	56
Gambar 4.2.9 2 Hasil Translate KitaLulus	57
Gambar 4.2.9 3 Code Translator Jobstreet	57
Gambar 4.2.9 4 Hasil Translate Jobstreet	58

Gambar 4.2.9 5 Code Translator LinkedIn	58
Gambar 4.2.9 6 Hasil Translate LinkedIn	59
Gambar 4.2.10 1 Code Sentimen Text Blob KitaLulus	60
Gambar 4.2.10 2 Code WordCloud KitaLulus	61
Gambar 4.2.10 3 Visualisasi WordCloud KitaLulus	62
Gambar 4.2.10 4 Code visualisasi distribusi sentimen KitaLulus	62
Gambar 4.2.10 5 Visualisasi distribusi sentimen KitaLulus	63
Gambar 4.2.10 6 Code Sentimen Text Blob Jobstreet.....	64
Gambar 4.2.10 7 Code WordCloud Jobstreet.....	65
Gambar 4.2.10 8 Wordcloud Jobstreet	66
Gambar 4.2.10 9 Code visualisasi distribusi sentimen Jobstreet	66
Gambar 4.2.10 10 Distribusi sentimen Jobstreet	67
Gambar 4.2.10 11 Code Sentimen Text Blob LinkedIn.....	68
Gambar 4.2.10 12 Code WordCloud Jobstreet.....	69
Gambar 4.2.10 13 Wordcloud LinkedIn	69
Gambar 4.2.10 14 Code visualisasi distribusi sentimen LinkedIn	70
Gambar 4.2.10 15 Distribusi sentimen LinkedIn	71
Gambar 4.3 1 TF-IDF	72
Gambar 4.4.1 1 Naive Bayes KitaLulus	74
Gambar 4.4.1 2 Naive Bayes KitaLulus Search Grid	75
Gambar 4.4.1 3 Naive Bayes KitaLulus SMOTE.....	76
Gambar 4.4.1 4 Naive Bayes Jobstreet.....	77
Gambar 4.4.1 5 Naive Bayes Jobstreet Search Grid.....	78
Gambar 4.4.1 6 Naive Bayes Jobstreet SMOTE	79
Gambar 4.4.1 7 Naive Bayes LinkedIn.....	80
Gambar 4.4.1 8 Naive Bayes LinkedIn Grid Search	81
Gambar 4.4.1 9 Naive Bayes LinkedIn SMOTE	82
Gambar 4.4.2 1 KNN KitaLulus	83
Gambar 4.4.2 2 KNN KitaLulus Grid Search.....	84
Gambar 4.4.2 3 KNN KitaLulus SMOTE	85
Gambar 4.4.2 4 KNN Jobstreet.....	86
Gambar 4.4.2 5 KNN Jobstreet Grid Search	87
Gambar 4.4.2 6 KNN Jobstreet SMOTE	88
Gambar 4.4.2 7 KNN LinkedIn.....	89
Gambar 4.4.2 8 KNN LinkedIn GridSearch	90
Gambar 4.4.2 9 KNN LinkedIn SMOTE.....	91
Gambar 4.4.3 1 Decision Tree KitaLulus	92
Gambar 4.4.3 2 Decision Tree KitaLulus Grid Search.....	93
Gambar 4.4.3 3 Decision Tree KitaLulus SMOTE.....	94
Gambar 4.4.3 4 Decision Tree Jobstreet.....	95

Gambar 4.4.3 5 Decision Tree Jobstreet GridSearch.....	96
Gambar 4.4.3 6 Decision Tree Jobstreet SMOTE	97
Gambar 4.4.3 7 Decision Tree LinkedIn.....	98
Gambar 4.4.3 9 Decision Tree LinkedIn SMOTE	100
Gambar 4.4.4 1 SVM KitaLulus.....	101
Gambar 4.4.4 2 SVM KitaLulus GridSearch.....	102
Gambar 4.4.4 3 SVM KitaLulus SMOTE	103
Gambar 4.4.4 4 SVM Jobstreet	104
Gambar 4.4.4 5 SVM Jobstreet GridSearch.....	105
Gambar 4.4.4 6 SVM Jobstreet SMOTE	106
Gambar 4.4.4 7 SVM LinkedIn.....	107
Gambar 4.4.4 8 SVM LinkedIn GridSearch	108
Gambar 4.4.4 9 SVM LinkedIn SMOTE	109



DAFTAR LAMPIRAN

Rumus 2 1 Precision	17
Rumus 2 2 Recall	17
Rumus 2 3 FI-Score	18
Rumus 2 4 Accuracy	18
Rumus 2 5 Rumus Decision Tree	23

