

**PERBANDINGAN KINERJA RANDOM FOREST DAN
XGBOOST UNTUK ANALISIS SENTIMEN ULASAN
PENGGUNA APLIKASI SHOPEEFOOD DRIVER DI GOOGLE
PLAY STORE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi SI Informatika



disusun oleh

MUHNISA APRILLIA SARI

22.11.5255

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**PERBANDINGAN KINERJA RANDOM FOREST DAN
XGBOOST UNTUK ANALISIS SENTIMEN ULASAN
PENGGUNA APLIKASI SHOPEEFOOD DRIVER DI GOOGLE
PLAY STORE**

SKRIPSI

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



disusun oleh

MUHNISA APRILLIA SARI

22.11.5255

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

PERBANDINGAN KINERJA RANDOM FOREST DAN XGBOOST UNTUK ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SHOPEEFOOD DRIVER DI GOOGLE PLAY STORE

yang disusun dan diajukan oleh

Muhnisa Aprillia Sari

22.11.5255

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Mei 2026

Dosen Pembimbing,


Mulia Sulistivono, S.Kom., M.Kom.

NIK. 190302248

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN KINERJA RANDOM FOREST DAN XGBOOST UNTUK ANALISIS
SENTIMEN ULASAN PENGGUNA APLIKASI SHOPEEFOOD DRIVER DI GOOGLE
PLAY STORE

yang disusun dan diajukan oleh

Muhnisa Aprillia Sari

22.11.5255

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Mei 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya Dhuwita, S.Kom., M.Kom
NIK. 190302185

Rizqi Sukma Kharisma, S.Kom., M.Kom
NIK. 190302215

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

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Mei 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhnisa Aprillia Sari
NIM : 22.11.5255

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan kinerja Random Forest dan XGBoost untuk analisis sentimen ulasan pengguna aplikasi shopeefood driver di google play store

Dosen Pembimbing : Mulia Sulistiyono, S.Kom., 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, 20 Mei 2026

Yang Menyatakan,



Muhnisa Aprillia Sari

HALAMAN PERSEMBAHAN

Alhamdulillah rabbil 'ālamīn, segala puji dan syukur penulis panjatkan kepada Allah SWT atas limpahan rahmat, nikmat, dan karunia-Nya sehingga Tugas Akhir ini dapat diselesaikan dengan baik. Penyusunan karya ini merupakan salah satu bentuk tanggung jawab akademik sekaligus tahapan akhir dalam menyelesaikan Pendidikan sarjana, seluruh proses yang dilalui dalam penyusunan tugas akhir ini tentu tidak mudah, namun dapat terlewati berkat usaha, doa, serta dukungan dari berbagai pihak. Dengan penuh rasa Syukur, penulis mempersembahkan Tugas Akhir ini kepada:

1. Allah SWT yang senantiasa memberikan Kesehatan, kekuatan, kelancaran, serta kemudahan kepada penulis dalam menjalani proses penelitian dan penyusunan tugas akhir hingga dapat terselesaikan dengan baik.
2. Ibunda tersayang, Ibu Indi, Bapak Mustari selaku bapak kandung penulis, serta Bapak Karno selaku bapak sambung penulis, yang selalu menjadi sumber kekuatan, semangat, dan motivasi bagi penulis. Terima kasih atas segala doa, kasih sayang, perhatian, serta dukungan yang diberikan tanpa henti, baik secara moril maupun materiil, dalam menemani setiap proses perjalanan penulis. Segala pengorbanan dan usaha yang telah diberikan menjadi penyemangat terbesar bagi penulis dalam menyelesaikan tugas akhir ini.
3. Saudara kandung penulis, Talitha dan Nizam, yang selalu memberikan warna dalam kehidupan penulis, baik melalui kebersamaan maupun dukungan yang diberikan, sehingga menjadi salah satu penyemangat bagi penulis dalam menyelesaikan penelitian ini.
4. Mbah Kakung tercinta dan Almh. Mbah Putri tercinta yang senantiasa memberikan kasih sayang, doa, perhatian, serta dukungan dengan penuh ketulusan kepada penulis, baik melalui nasihat, perhatian, maupun dukungan materiil. Begitu pula keluarga dari pihak bapak, yaitu Alm. Tete

- dan Almh. Nenek tercinta, yang selalu menjadi penyemangat dan memberikan motivasi bagi penulis dalam menyelesaikan tugas akhir ini.
5. Kepada teman-teman seperjuangan penulis, Khoirunnisa Salsa Sabila, Alti Arindika, Nafisa Dian Ramadhani, Fiola Utri Aulya, dan Heni Septa Mawar, terima kasih atas kebersamaan, cerita, dukungan, serta semangat yang selalu diberikan kepada penulis. Kehadiran kalian menjadi bagian penting yang menemani dan menguatkan penulis selama menjalani masa perkuliahan hingga proses penyusunan skripsi ini.
 6. Teman-teman tercinta, Dian Nurtika Sari, Ella Avitasari, dan Firda Umi Khorih, terima kasih atas segala dukungan, perhatian, serta kebersamaan yang telah diberikan kepada penulis. Terima kasih telah menemani dalam setiap suka dan duka, menghadirkan kebahagiaan, canda, dan tawa, serta selalu menjadi tempat pulang yang hangat seperti sedia kala. Semangat, dukungan, perhatian, serta kasih sayang yang kalian berikan tanpa henti menjadi kekuatan bagi penulis dalam menyelesaikan perkuliahan dan skripsi ini.
 7. Kepada keluarga kecil Griceto terkhusus (Muhammad Djumroni, Aldhy Wijayanto, Ridwan Taufik Qurrahman, Qowiy Ardiansyah, Rahmah Nur aini, Siva Eka Fatihah, Cintya Devi Wulandari, dan Resa Dwi Anggraeni) terima kasih atas kebersamaan, canda tawa, dukungan, serta semangat yang selalu diberikan kepada penulis. Kehadiran kalian telah membawa banyak kebahagiaan dan menjadi penyemangat bagi penulis selama menjalani proses perkuliahan hingga penyusunan skripsi ini.
 8. Muhnisa Aprillia Sari, terima kasih karena telah bertahan sejauh ini, melewati berbagai keraguan, dan tetap melangkah tanpa menyerah. Meskipun banyak rintangan, tantangan, serta ujian yang datang silih berganti, semua itu akhirnya dapat dilewati dengan baik. Terima kasih karena telah memilih untuk tetap kuat, terus berjuang, dan berani menyelesaikan apa yang telah dimulai.

KATA PENGANTAR

Segala Puji dan Syukur penulis panjatkan kepada Allah SWT, Tuhan Yang Maha Esa, nikmat, serta karunia-Nya sehingga penulis dapat menyelesaikan laporan kelulusan ini dengan baik dan lancar. dalam proses penelitian hingga penyusunan laporan ini, penulis menyadari bahwa seluruh tahapan yang telah dilalui tidak terlepas dari bantuan, dukungan, arahan, serta doa dari berbagai pihak. Oleh sebab itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta
2. Ibu Prof. Dr. Kusri, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom. selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
4. Bapak Mulia Sulistiyono, S.kom, M.Kom. selaku Dosen Pembimbing penulis.
5. Kedua orang tua beserta keluarga penulis yang selalu memberikan doa, perhatian, semangat, serta dukungan baik secara moral maupun materiil selama proses penyusunan laporan ini hingga selesai.
6. Sahabat serta teman-teman seperjuangan yang selalu hadir memberikan motivasi, bantuan, dan semangat, serta menemani penulis dalam melewati berbagai proses selama penyusunan skripsi ini.

Penulis menyadari adanya kekurangan dalam skripsi ini dan mengharapkan kritik dan saran yang membangun demi perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan pihak terkait.

Yogyakarta, 09 Mei 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvii
ABSTRACT	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.5.1 Manfaat Teoritis	5
1.5.2 Manfaat Praktis	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	8
2.1 Studi Literatur	8

2.2 Dasar Teori	16
2.2.1 Analisis sentiment.....	16
2.2.2 Ulasan Pengguna Google Play Store Sebagai Sumber Data	16
2.2.3 Natural Language Preprocessing (NLP).....	16
2.2.4 Preprocessing Teks	17
2.2.5 Pelabelan Sentiment.....	17
2.2.6 <i>Exploratory Data Analysis</i> (EDA).....	18
2.2.7 Transformasi Klasifikasi (Multikelas ke Biner).....	19
2.2.8 <i>Label Encoding</i>	19
2.2.9 Pembagian Data (Training dan Testing).....	19
2.2.10 Ekstraksi Fitur Teks Menggunakan TF-IDF.....	20
2.2.11 Imbalanced Data dan SMOTE.....	21
2.2.12 Machine Learning dalam Klasifikasi Sentimen.....	22
2.2.13 Algoritma Random Forest	23
2.2.14 Algoritma XGBoost.....	24
2.2.15 Evaluasi Kinerja Model Klasifikasi.....	26
BAB III METODE PENELITIAN	29
3.1 Objek Penelitian	29
3.2 Alur Penelitian.....	29
3.2.1 Pengumpulan Data (Scraping Google Playstore)	31
3.2.2 Pemeriksaan Data Awal (Missing Value dan Duplikasi)	31
3.2.3 Analisis Awal Dataset.....	31
3.2.4 Pembersihan dan Pengolahan Data Teks (Preprocessing).....	32
3.2.5 Pelabelan Sentimen.....	32

3.2.6	<i>Exploratory Data Analysis (EDA) tiga kelas dan Wordcloud hasil Preprocessing</i>	33
3.2.7	Transformasi Klasifikasi (Tiga Kelas Menjadi Dua Kelas)	33
3.2.8	<i>Exploratory Data Analysis (EDA) dua kelas</i>	34
3.2.9	Label Encoding	34
3.2.10	Pembagian Data (Training dan Testing).....	34
3.2.11	Transformasi Fitur dengan TF-IDF	35
3.2.12	Penanganan Data Tidak Seimbang (SMOTE).....	35
3.2.13	Proses Pelatihan Model (Random Forest dan XGBoost)	35
3.2.14	Evaluasi Kinerja Model	36
3.2.15	Analisis Perbandingan Model.....	36
3.2.16	Tahap Akhir	36
3.3	Alat dan Bahan	36
3.3.1	Data Penelitian.....	37
3.3.2	Alat/instrumen	39
BAB IV HASIL DAN PEMBAHASAN		40
4.1	Pengumpulan Data.....	40
4.2	Pembersihan Data	42
4.3	Hasil Preprocessing Teks	43
4.3.1	Basic Cleaning	44
4.3.2	Tokenization	45
4.3.3	Stopword Removal	45
4.3.4	Stemming.....	45
4.3.5	Normalisasi	46
4.4	Hasil Pelabelan Sentimen	48

4.5 Exploratory Data Analysis (EDA) Tiga kelas	49
4.5.1 Distribusi Sentimen	49
4.5.2 Wordcloud	51
4.6 Pembentukan Dataset Binary	54
4.7 EDA Setelah Transformasi 2 kelas	55
4.8 label encoding	57
4.9 Pembagian Data dan Ekstraksi Fitur	57
4.10TF-IDF	58
4.11Penanganan Imbalance dengan SMOTE	59
4.12Pemodelan (Random Forest & XGBoost)	60
4.13Evaluasi	60
4.14Confusion Matrix	61
4.15Perbandingan kinerja model	63
BAB V PENUTUP	65
5.1 Kesimpulan	65
REFERENSI	67

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	11
Tabel 3.1 Hasil Preprocessing Data	38
Tabel 3.2 Spesifikasi Perangkat Keras dan Lunak.....	39
Tabel 4.1 Dataset Scraping	41
Tabel 4.2 Hasil Pembersihan Data.....	43
Tabel 4.3 Hasil preprocessing Teks	46
Tabel 4.4 Contoh Ulasan Berdasarkan Kelas Sentimen (Tiga Kelas)	50
Tabel 4.5 Contoh Ulasan Berdasarkan Kelas Sentimen (Dua Kelas)	56
Tabel 4.6 label encoding	57
Tabel 4.7 Pembagian Data Training dan Testing	57
Tabel 4.8 Pembagian Data Training dan Testing	58
Tabel 4.9 Evaluasi	61

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	30
Gambar 4.1 Tahapan Preprocessing	44
Gambar 4.2 Implementasi Basic Cleaning pada Teks	44
Gambar 4.3 Kode preprocessing	46
Gambar 4.4 Hasil Pelabelan Sentimen	49
Gambar 4.5 Hasil Exploratory Data Analysis (EDA) Tiga Kelas	50
Gambar 4.6 Worcloud Clean Sentimen Positif	52
Gambar 4.7 Worcloud Clean Sentimen Negatif	53
Gambar 4.8 Worcloud Clean Sentimen Neutral	54
Gambar 4.9 Hasil Pembentukan Dataset Binary	55
Gambar 4.10 EDA Setelah Transformasi Dua Kelas	56
Gambar 4.11 Penanganan Imbalance dengan SMOTE	59
Gambar 4.12 Confusion Matrix	62
Gambar 4.13 Perbandingan Kinerja Model	63

DAFTAR LAMBANG DAN SINGKATAN



<i>NLP</i>	<i>Natural Language Processing</i>
<i>TF-IDF</i>	<i>Term Frequency–Inverse Document Frequency</i>
<i>EDA</i>	<i>Exploratory Data Analysis</i>
<i>SMOTE</i>	<i>Synthetic Minority Over-sampling Technique</i>
<i>SVM</i>	<i>Support Vector Machine</i>
<i>RF</i>	<i>Random Forest</i>
<i>XGBoost</i>	<i>Extreme Gradient Boosting</i>
<i>CSV</i>	<i>Comma Separated Values</i>
<i>TP</i>	<i>True Positive</i>
<i>TN</i>	<i>True Negative</i>
<i>FP</i>	<i>False Positive</i>
<i>FN</i>	<i>False Negative</i>
<i>TF</i>	<i>Term Frequency</i>
<i>IDF</i>	<i>Inverse Document Frequency</i>
<i>F1-Score</i>	<i>harmonic mean antara precision dan recall</i>

DAFTAR ISTILAH

<i>Analisis Sentimen</i>	metode analisis teks yang digunakan untuk mengetahui kecenderungan opini atau penilaian pengguna terhadap suatu objek, baik bersifat positif, negative, maupun netral
<i>Machine Learning</i>	cabang kecerdasan buatan yang memungkinkan sistem komputer mempelajari pola dari data dan menghasilkan prediksi atau Keputusan secara otomatis
<i>Preprocessing</i>	tahapan awal dalam pengolahan data teks yang dilakukan untuk membersihkan dan mempersiapkan data agar lebih terstruktur sebelum diproses oleh model
<i>Tokenisasi</i>	proses memecah kalimat atau teks menjadi bagian-bagian kecil berupa kata atau token untuk mempermudah analisis
<i>Stemming</i>	proses mengubah kata berimbuhan menjadi bentuk kata dasar agar variasi kata dapat disederhanakan
<i>Stopword Removal</i>	proses menghapus kata-kata umum yang sering muncul tetapi tidak memiliki pengaruh besar terhadap makna teks
<i>Wordcloud</i>	bentuk visualisasi data teks yang menampilkan kata berdasarkan tingkat frekuensi kemunculannya dalam suatu Kumpulan data
<i>Dataset</i>	Kumpulan data yang digunakan sebagai sumber utama dalam proses pelatihan, pengujian, dan analisis penelitian
<i>Klasifikasi</i>	proses pengelompokan data ke dalam kategori tertentu berdasarkan karakteristik atau pola yang dimiliki
<i>Accuracy</i>	ukuran evaluasi yang menunjukkan tingkat ketepatan model dalam memprediksi data pada kelas positif
<i>Precision</i>	ukuran evaluasi yang menunjukkan kemampuan model dalam memprediksi data pada kelas positif
<i>Recall</i>	ukuran evaluasi yang menunjukkan kemampuan model dalam mengenali seluruh data positif yang sebenarnya



<i>Confusion Matrix</i>	tabel evaluasi hasil yang digunakan untuk membandingkan hasil prediksi model dengan kondisi data yang sebenarnya
<i>Overfitting</i>	kondisi ketika model terlalu menyesuaikan diri dengan data pelatihan sehingga kemampuan prediksi terhadap data baru menjadi kurang baik
<i>Label Encoding</i>	proses mengubah data kategorikal atau label teks menjadi bentuk numerik agar dapat diproses oleh algoritma machine learning
<i>Scraping</i>	Teknik pengambilan data secara otomatis dari halaman website menggunakan bantuan program atau library tertentu
<i>Training Data</i>	bagian dataset yang digunakan dalam proses pelatihan model untuk mempelajari pola data
<i>Testing Data</i>	bagian dataset yang digunakan untuk menguji dan mengevaluasi performa model setelah proses pelatihan selesai

INTISARI

Penelitian ini membahas analisis sentiment terhadap ulasan pengguna aplikasi ShopeeFood Driver bertujuan untuk mengetahui opini pengguna terhadap layanan aplikasi. Banyaknya ulasan yang diberikan pengguna mengandung informasi penting mengenai kualitas layanan, kendala aplikasi, maupun kepuasan pengguna. Namun jumlah data ulasan yang besar menyebabkan proses analisis secara manual menjadi kurang efektif dan membutuhkan waktu yang lama. Oleh karena itu, diperlukan metode otomatis untuk mengklasifikasikan sentiment ulasan agar informasi yang diperoleh dapat dimanfaatkan sebagai bahan evaluasi dan pengambilan Keputusan bagi pengembang aplikasi.

Tahapan penelitian diawali dengan proses pengambilan data ulasan melalui Teknik scraping, kemudian dilanjutkan dengan preprocessing data yang meliputi cleaning, case folding, tokenizing, stopword removal, dan stemming. Setelah itu dilakukan pelabelan sentiment, transformasi data menjadi klasifikasi biner, serta label encoding untuk mengubah label kategorikal ke bentuk numerik. Data yang telah diproses kemudian diberikan bobot menggunakan metode TF-IDF sebelum dilakukan klasifikasi menggunakan algoritma Random Forest dan XGBoost.

Berdasarkan hasil pengujian, algoritma Random Forest menunjukkan performa yang lebih baik dibandingkan XGBoost. Random Forest memperoleh nilai accuracy sebesar 83,84 %, precision 76,42%, recall 69,14%, dan f1-score 72,60%, sedangkan XGBoost memperoleh accuracy 79,66%, precision 65,05%, recall 74,10%, dan f1-score 69,28%. Hasil penelitian ini diharapkan dapat membantu pengembang ShopeeFood Driver dalam memahami kebutuhan dan kepuasan pengguna sehingga kualitas layanan aplikasi dapat terus ditingkatkan.

Kata kunci: analisis sentimen, Random Forest, XGBoost, TF-IDF, ShopeeFood Driver

ABSTRACT

This study discusses sentiment analysis on user reviews of the ShopeeFood Driver application to identify user opinions regarding the services provided by the application. User reviews contain important information related to service quality, application issues, and user satisfaction. However, the large amount of review data makes manual analysis inefficient and time-consuming. Therefore, an automatic method is needed to classify review sentiments so that the information obtained can be used as evaluation material and decision-making support for application developers.

The research process began with collecting review data through a scraping technique, followed by data preprocessing stages consisting of cleaning, case folding, tokenizing, stopword removal, and stemming. Afterward, sentiment labeling, binary classification transformation, and label encoding were carried out to convert categorical labels into numerical form. The processed data were then weighted using the TF-IDF method before classification was performed using the Random Forest and XGBoost algorithms.

Based on the test results, the Random Forest algorithm showed better performance compared to XGBoost. Random Forest achieved an accuracy of 83.84%, precision of 76.42%, recall of 69.14%, and F1-score of 72.60%, while XGBoost achieved an accuracy of 79.66%, precision of 65.05%, recall of 74.10%, and F1-score of 69.28%. The results of this study are expected to help ShopeeFood Driver developers understand user needs and satisfaction so that the quality of the application's services can continue to be improved.

Keyword: sentiment analysis, Random Forest, XGBoost, TF-IDF, ShopeeFood Driver