

**ANALISIS KLASIFIKASI SENTIMEN ULASAN TOKOPEDIA
MENGUNAKAN TF-IDF, SMOTE, DAN MODEL
MACHINE LEARNING TRADISIONAL**

JALUR SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

HERIANTA BARUS

22.12.2343

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS KLASIFIKASI SENTIMEN ULASAN TOKOPEDIA
MENGUNAKAN TF-IDF, SMOTE, DAN MODEL
MACHINE LEARNING TRADISIONAL**

JALUR NON REGULER - SCIENTIST

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

HERIANTA BARUS

22.12.2343

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

JALUR NON REGULER - SCIENTIST

**ANALISIS KLASIFIKASI SENTIMEN ULASAN TOKOPEDIA
MENGUNAKAN TF-IDF, SMOTE, DAN MODEL
MACHINE LEARNING TRADISIONAL**

yang disusun dan diajukan oleh

HERIANTA BARUS

22.12.2343

telah disetujui oleh Dosen Pembimbing
pada tanggal 27 Januari 2026

Dosen Pembimbing,



Ika Nur Fairi, S.Kom., M.Kom
NIK. 190302268

HALAMAN PENGESAHAN

JALUR SCIENTIST

**ANALISIS KLASIFIKASI SENTIMEN ULASAN TOKOPEDIA
MENGUNAKAN TF-IDF, SMOTE, DAN MODEL
MACHINE LEARNING TRADISIONAL**

yang disusun dan diajukan oleh

HERIANTA BARUS

22.12.2343

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Atik Nurmasani, S.Kom., M.Kom.
NIK. 190302354

Irma Rofni Wulandari, S.Pd., M.Eng.
NIK. 190302329

Ika Nur Fajri, S.Kom., M.Kom.
NIK. 190302268



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KONTRIBUSI KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : **HERIANTA BARUS**
NIM : **22.12.2343**

Menyatakan bahwa karya dengan judul berikut:

Analisis Klasifikasi Sentimen Ulasan Tokopedia Menggunakan TF-IDF, SMOTE, Dan Model Machine Learning Tradisional

Dosen Pembimbing : **Ika Nur Fajri, S.kom., M.kom**

1. Karya adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya merupakan gagasan, rumusan maupun penelitian yang orisinal dan **SAYA** memiliki **KONTRIBUSI** terhadap karya tersebut.
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 atau Referensi 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, 7 Januari 2026

Yang Menyatakan,



HERIANTA BARUS

HALAMAN PERSEMBAHAN

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat, kesehatan, serta kekuatan yang telah diberikan, sehingga penulis dapat menyelesaikan laporan tugas akhir ini dengan baik. Karya ini merupakan hasil dari proses panjang yang penuh dengan pembelajaran, kesabaran, dan tanggung jawab, serta tidak terlepas dari dukungan dan peran berbagai pihak yang berarti dalam perjalanan akademik penulis.

Dengan penuh rasa hormat dan syukur, laporan tugas akhir ini penulis persembahkan kepada:

1. Bapak Mimpin Barus dan Ibu Murni Br. Sembiring, sebagai orang tua tercinta. Terima kasih atas doa yang tidak pernah terputus, kasih sayang, perhatian, serta dukungan moril dan materiil yang senantiasa diberikan kepada penulis. Segala pengorbanan dan kepercayaan yang diberikan menjadi sumber kekuatan utama bagi penulis untuk terus berjuang dan menyelesaikan pendidikan hingga tahap ini.
2. Bapak Ika Nur Fajri, S.Kom., M.Kom., selaku Dosen Pembimbing. Terima kasih atas bimbingan, arahan, masukan, serta kesabaran yang telah diberikan selama proses penyusunan laporan tugas akhir ini. Dukungan dan pengawasan yang diberikan sangat membantu penulis dalam menyusun penelitian secara sistematis dan bertanggung jawab.
3. Bapak Yoga Pristyanto, S.Kom., M.Eng., yang telah memberikan saran dan masukan penting, khususnya terkait pengembangan penelitian dengan membandingkan beberapa model yang digunakan. Arahan tersebut memberikan sudut pandang yang lebih luas dan memperkuat kualitas penelitian yang dilakukan penulis.
4. Teman-teman Program Studi Sistem Informasi, terima kasih atas kebersamaan, dukungan, diskusi, serta semangat yang terjalin selama masa

perkuliahan. Proses belajar dan perjuangan bersama menjadi pengalaman berharga yang turut membentuk penulis hingga sampai pada tahap ini.

5. Terakhir, untuk diri penulis sendiri, terima kasih telah bertahan, berusaha, dan tidak menyerah dalam menghadapi berbagai tantangan selama proses perkuliahan dan penyusunan tugas akhir ini. Penyelesaian laporan ini merupakan bukti bahwa dengan ketekunan, disiplin, dan keyakinan, setiap proses yang dijalani dengan sungguh-sungguh akan membuahkan hasil.

Penulis berharap laporan tugas akhir ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan maupun sebagai referensi bagi penelitian selanjutnya. Kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang.



KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir ini dengan lancar. Tugas akhir ini disusun untuk memenuhi salah satu syarat kelulusan Program Sarjana pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selain itu penulis dengan segala kerendahan hati ingin menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa memberikan dukungan dan bantuan untuk menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta
2. Ibu Prof. Dr. Kusriani, S.Kom., M.Kom., selaku Dekan Program Fakultas Ilmu Komputer
3. Bapak Anggit Dwi Hartanto, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi
4. Bapak Ika Nur Fajri, S.Kom., M.Kom., selaku dosen pembimbing yang memberikan arahan, saran, dan motivasi terhadap penulis
5. Kedua orang tua, keluarga besar, dan teman-teman tercinta yang memberikan semangat dan doa kepada penulis.
6. Seluruh pihak yang telah membantu dan berkontribusi dalam proses penyusunan serta penyelesaian laporan tugas akhir ini.

Yogyakarta, 7 Januari 2026

HERIANTA BARUS

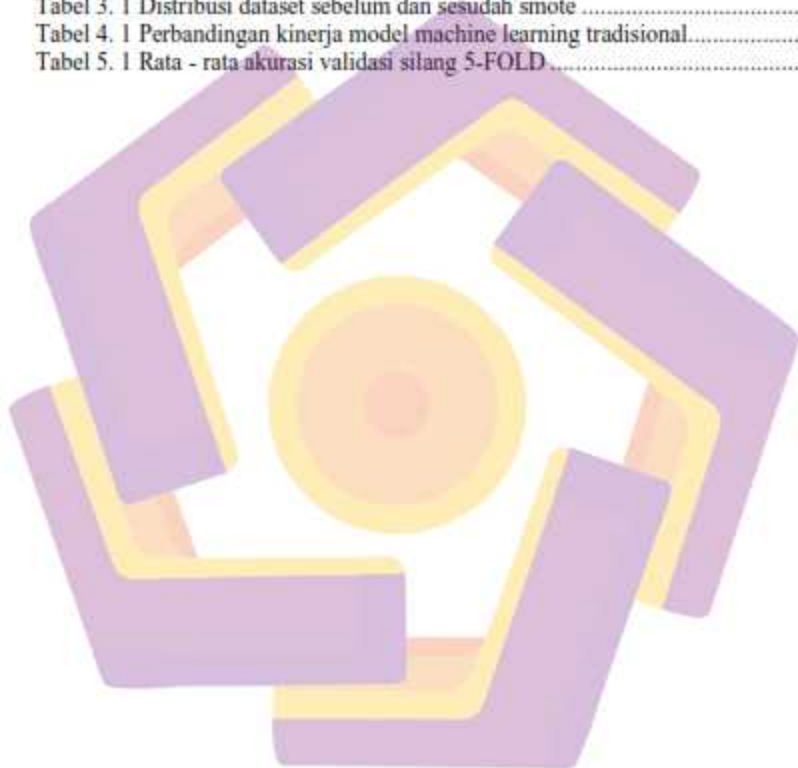
DAFTAR ISI

Contents

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KONTRIBUSI KARYA	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
BAB I IDENTITAS PUBLIKASI	1
BAB II PROSES SUBMIT	2
2.1 Lembar Review	2
2.2.1 Review Round 1	4
2.2.2 Review Round 2	27
2.2 Lembar Persetujuan (LoA)	27
2.3 Hasil Diskusi Ujian (Opsional)	28
BAB III ISI KARYA ILMIAH	35
3.1 Intisari	35
3.2 Pendahuluan	36
3.3 Metode	37
3.4 Hasil dan Pembahasan	49
3.5 Kesimpulan	58
3.6 Referensi	60
LAMPIRAN	63

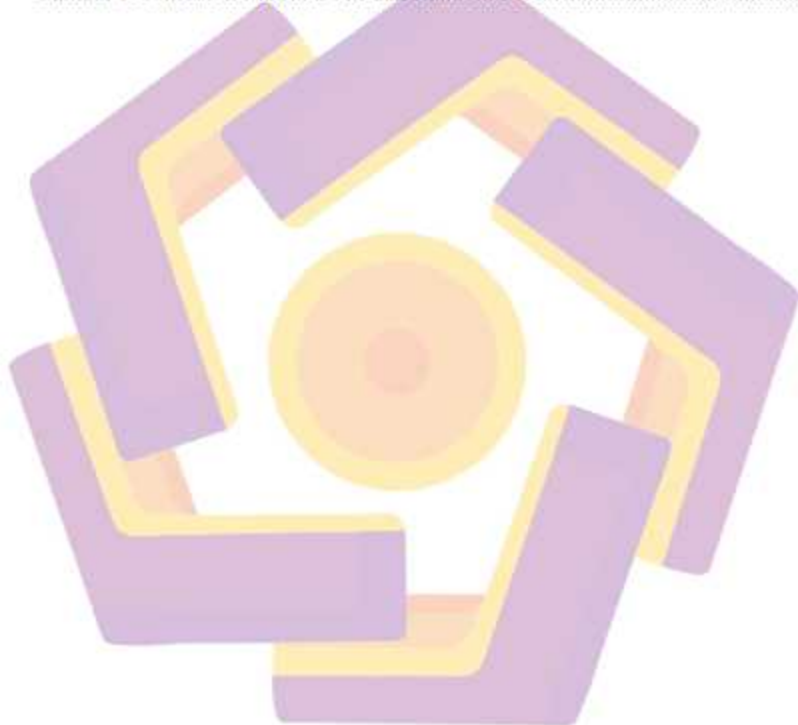
DAFTAR TABEL

Tabel 1. 1 Identitas Publikasi	1
Tabel 2. 1 Bukti Lembar Evaluasi Reviewer A	2
Tabel 2. 2 Bukti Lembar Evaluasi Reviewer B	3
Tabel 2. 3 Bukti Komentar Reviewer A dan B	3
Tabel 2. 4 Lembar Evaluasi Reviewer	4
Tabel 2. 5 Ringkasan Hasil Review A	5
Tabel 2. 6 Ringkasan Hasil Review B	18
Tabel 3. 1 Distribusi dataset sebelum dan sesudah smote	50
Tabel 4. 1 Perbandingan kinerja model machine learning tradisional	53
Tabel 5. 1 Rata - rata akurasi validasi silang 5-FOLD	54



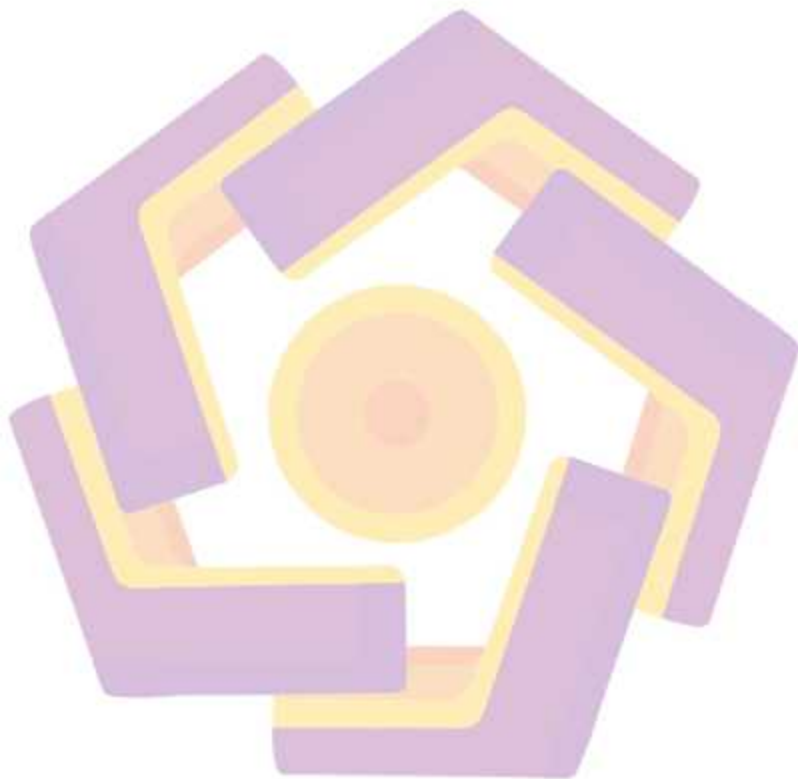
DAFTAR GAMBAR

Gambar 1 Letter Of Accept (Loa)	28
Gambar 2 gambar tahapan penelitian	38
Gambar 3 gambar Data sebelum terapkan SMOTE	50
Gambar 4 gambar Data setelah terapkan SMOTE	51
Gambar 5 Wordcloud Sentimen Positif dan Negatif	52
Gambar 6 Top 10 Kata pada Ulasan Positif dan Negatif	52
Gambar 7 Top 10 Kata pada Ulasan Positif dan Negatif	53
Gambar 8 Visualisasi Fitur Penting (Koefisien Logistic Regression)	55
Gambar 9 Confusion Matrix SVM dan Random Forest	57



DAFTAR LAMPIRAN

Lampiran 1 Naskah publikasi original yang telah terbit / published (dalam format PDF).....	63
Lampiran 2 Dataset (Opsional)	63
Lampiran 3 Hasil seluruh eksperimen yang telah dilakukan Link Google collabs :	63



INTISARI

Penelitian ini mengkaji klasifikasi sentimen pada ulasan pengguna Tokopedia dengan menggunakan TF-IDF sebagai metode ekstraksi fitur dan SMOTE untuk menangani ketidakseimbangan kelas. Dari hampir satu juta ulasan mentah yang bersumber dari Kaggle ("E-Commerce Ratings and Reviews in Bahasa Indonesia"), diperoleh sebanyak 6.477 data relevan setelah melalui tahapan prapemrosesan yang ketat, meliputi case folding, penghapusan noise (emoji, URL, dan angka), normalisasi sesuai standar KBBI, tokenisasi, penghapusan *stopword*, serta proses stemming menggunakan Sastrawi. Dataset akhir terdiri atas 5.213 ulasan positif dan 1.264 ulasan negatif (80,4% positif). Untuk mengatasi ketidakseimbangan kelas, diterapkan metode SMOTE sehingga data pelatihan menjadi seimbang dengan total 10.426 ulasan dan rasio kelas 1:1. Lima model machine learning tradisional dievaluasi, yaitu Naive Bayes, Logistic Regression, Support Vector Machine (SVM), Decision Tree, dan Random Forest. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, F1-score, ROC-AUC, serta waktu komputasi, dengan skema pembagian data stratified 80:20 dan validasi silang 5-fold. Hasil penelitian menunjukkan bahwa Random Forest memberikan performa keseluruhan terbaik dengan akurasi sebesar 0,9163, F1-score 0,9133, dan ROC-AUC 0,9784. Sementara itu, model SVM yang telah dituning (C=10, kernel RBF) mencapai akurasi tertinggi sebesar 0,9473 dan F1-score 0,9321. Validasi silang pada model Naive Bayes menunjukkan hasil yang konsisten dengan rata-rata akurasi sebesar 88,09%. Analisis lanjutan menggunakan koefisien Logistic Regression mengidentifikasi fitur kata yang berpengaruh terhadap sentimen, di mana sentimen positif berkorelasi dengan kata-kata seperti "mantap", "mudah", dan "sukses", sedangkan sentimen negatif berkaitan dengan kata-kata seperti "kecewa", "parah", dan "lemot". Temuan ini memberikan nilai praktis bagi Tokopedia, khususnya sebagai dasar peningkatan pengalaman pengguna, seperti optimalisasi kecepatan aplikasi dan penanganan keluhan pelanggan. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa metode machine learning tradisional efektif dan efisien untuk analisis sentimen dalam konteks Bahasa Indonesia.

Keyword: Analisis Sentimen, Tokopedia, E-commerce, Naive Bayes, Random Forest, Support Vector Machine (SVM), Logistic Regression, Decision Tree, TF-IDF, SMOTE

ABSTRACT

This study explores sentiment classification on Tokopedia user reviews using TF-IDF for feature extraction and SMOTE to handle class imbalance. From nearly one million raw reviews sourced from Kaggle ("E-Commerce Ratings and Reviews in Bahasa Indonesia"), a final set of 6,477 relevant entries was obtained after rigorous preprocessing, including case folding, noise removal (emojis, URLs, numbers), normalization to KBBI standards, tokenization, stopword removal, and stemming with Sastrawi. The dataset consisted of 5,213 positive and 1,264 negative reviews (80.4% positive). SMOTE balanced the classes to 10,426 reviews with a 1:1 ratio for training. Five traditional machine learning models were evaluated: Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), Decision Tree, and Random Forest. Assessments were based on accuracy, precision, recall, F1-score, ROC-AUC, and computational time, using an 80:20 stratified split and 5-fold cross-validation. Random Forest achieved the best overall performance (accuracy: 0.9163, F1-score: 0.9133, ROC-AUC: 0.9784), while tuned SVM (C=10, RBF kernel) attained the highest accuracy of 0.9473 and F1-score of 0.9321. Cross-validation on Naïve Bayes showed consistent results with an average accuracy of 88.09%. Further analysis using Logistic Regression coefficients identified influential features: positive sentiment associated with words like "mantap", "mudah", and "sukses", while negative sentiment correlated with "kecewa", "parah", and "lemot". These insights provide practical value for Tokopedia's teams to enhance user experience, such as improving app speed and addressing complaints. The findings demonstrate the effectiveness and efficiency of traditional machine learning techniques for sentiment analysis in Bahasa Indonesia contexts.

Keyword: Sentiment Analysis, Tokopedia, E-commerce, Naïve Bayes, Random Forest, Support Vector Machine (SVM), Logistic Regression, Decision Tree, TF-IDF, SMOTE