

**PENERAPAN METODE *LOGISTIC REGRESSION* DAN  
TF-IDF DALAM ANALISIS SENTIMEN TERHADAP  
APLIKASI SHOPEE**

**SKRIPSI**

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



disusun oleh

**DEDEN DARMAWANSYAH**

**22.11.5190**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2026**

**PENERAPAN METODE *LOGISTIC REGRESSION* DAN  
TF-IDF DALAM ANALISIS SENTIMEN TERHADAP  
APLIKASI SHOPEE**

**SKRIPSI**

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



disusun oleh

**DEDEN DARMAWANSYAH**

**22.11.5190**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2026**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PENERAPAN METODE *LOGISTIC REGRESSION* DAN TF-IDF DALAM  
ANALISIS SENTIMEN TERHADAP APLIKASI SHOPEE**

yang disusun dan diajukan oleh

**Deden Darmawansyah**

**22.11.5190**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 12 Mei 2026

Dosen Pembimbing,



**Bayu Setiaji M. Kom**

**NIK. 190302216**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**PENERAPAN METODE *LOGISTIC REGRESSION* DAN TF-IDF DALAM ANALISIS SENTIMEN TERHADAP APLIKASI SHOPEE**

yang disusun dan diajukan oleh

**Deden Darmawansyah**

**22.11.5190**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 12 Mei 2026

**Susunan Dewan Penguji**

**Nama Penguji**

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

**Bambang Pilu Hartato, S.Kom., M.Eng.**  
**NIK. 190302707**

**Bayu Setiaji, M.Kom.**  
**NIK. 190302216**

**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 12 Mei 2026

**DEKAN FAKULTAS ILMU KOMPUTER**



**Prof. Dr. Kusrini, M.Kom.**  
**NIK. 190302106**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Deden Darmawansyah  
NIM : 22.11.5190

Menyatakan bahwa Skripsi dengan judul berikut:

**Penerapan Metode *Logistic Regression* dan TF-IDF dalam Analisis Sentimen Terhadap Aplikasi Shopee**

Dosen Pembimbing : Bayu Setiaji 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 Mei 2026

Yang Menyatakan,



Deden Darmawansyah

## HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan kepada:

1. Allah SWT, atas segala rahmat, nikmat kesehatan, kemudahan, serta karunia-Nya yang tidak pernah putus sehingga penulis mampu menyelesaikan skripsi ini dengan penuh rasa syukur.
2. Kedua orang tua tercinta, Bapak Aksah selaku ayah penulis yang telah menjadi sosok pembentuk karakter, panutan, dan teladan dalam kehidupan penulis. Terima kasih atas kerja keras, pengorbanan, nasihat, ketegasan, serta nilai-nilai kehidupan yang selalu diajarkan sehingga penulis dapat tumbuh menjadi pribadi yang lebih kuat, disiplin, dan bertanggung jawab. Serta Ibu Sukrah selaku ibu penulis yang selalu memberikan kasih sayang, doa tanpa henti, perhatian, dukungan, dan semangat dalam setiap langkah penulis. Terima kasih atas kesabaran, kelembutan, serta cinta tulus yang selalu menjadi kekuatan terbesar bagi penulis dalam menyelesaikan pendidikan dan skripsi ini. Tiada kata yang mampu menggambarkan rasa terima kasih penulis atas segala perjuangan, pengorbanan, dan cinta yang telah diberikan oleh kedua orang tua penulis.
3. Keluarga yang senantiasa memberikan motivasi dan semangat dalam setiap proses yang dijalani.
4. Dosen pembimbing dan seluruh civitas akademika yang telah memberikan bimbingan, ilmu, dan arahan.
5. Teman-teman serta semua pihak yang telah membantu dan memberikan dukungan dalam penyusunan skripsi ini.

Semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan.

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada program studi yang penulis tempuh.

Dalam penyusunan skripsi ini, penulis menyadari bahwa tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak, penyelesaian skripsi ini tidak dapat berjalan dengan baik. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor.
2. Ibu Prof. Dr. Kusri, M.Kom. selaku Dekan.
3. Ibu Eli Pujastuti, M.Kom. selaku Ketua Program Studi.
4. Bapak Bayu Setiaji, M.Kom. selaku Dosen Pembimbing yang telah memberikan bimbingan, arahan, serta motivasi selama penyusunan skripsi ini.
5. Kedua orang tua tercinta, Ayah Aksah dan Ibu Sukrah, atas doa, dukungan, serta pengorbanan yang tiada henti.
6. Seluruh dosen dan civitas akademika yang telah memberikan ilmu dan bantuan selama masa perkuliahan.
7. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu dalam penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan.

Yogyakarta, 28 April 2026

Penulis

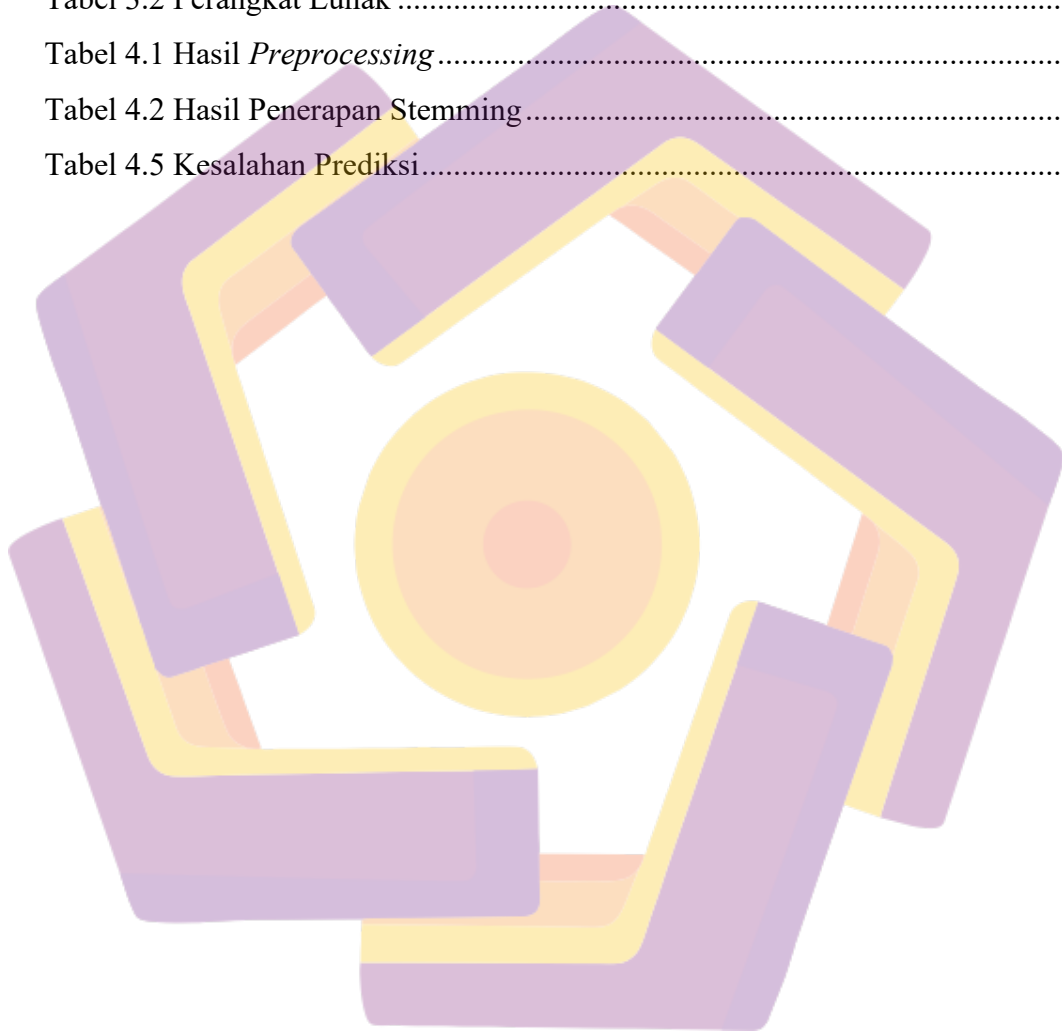
## DAFTAR ISI

|                                           |      |
|-------------------------------------------|------|
| 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 LAMBANG DAN SINGKATAN .....        | xi   |
| INTISARI .....                            | xii  |
| <i>ABSTRACT</i> .....                     | xiii |
| BAB I PENDAHULUAN.....                    | 1    |
| 1.1 Latar Belakang .....                  | 1    |
| 1.2 Rumusan Masalah.....                  | 2    |
| 1.3 Batasan Masalah .....                 | 3    |
| 1.4 Tujuan Penelitian .....               | 4    |
| 1.5 Manfaat Penelitian .....              | 4    |
| 1.6 Sistematika Penulisan .....           | 5    |
| BAB II TINJAUAN PUSTAKA .....             | 7    |
| 2.1 Studi Literatur .....                 | 7    |
| 2.2 Dasar Teori.....                      | 14   |
| BAB III METODE PENELITIAN .....           | 28   |
| 3.1 Objek Penelitian.....                 | 28   |

|                                   |                                                          |    |
|-----------------------------------|----------------------------------------------------------|----|
| 3.2                               | Alur Penelitian .....                                    | 30 |
| 3.3                               | Alat dan Bahan.....                                      | 33 |
| BAB IV HASIL DAN PEMBAHASAN ..... |                                                          | 35 |
| 4.1                               | Deskripsi Dataset dan Karakteristik Data .....           | 35 |
| 4.2                               | Jumlah Data yang Digunakan .....                         | 36 |
| 4.3                               | Distribusi <i>Rating</i> dan Karakteristik Teks.....     | 36 |
| 4.4                               | Pelabelan Sentimen Otomatis .....                        | 37 |
| 4.5                               | Pemrosesan Teks (Teks Preprocessing).....                | 38 |
| 4.6                               | Validasi dan Pembersihan Data Lanjutan .....             | 42 |
| 4.7                               | Ekstraksi Fitur Menggunakan TF-IDF.....                  | 42 |
| 4.8                               | Penyeimbangan Kelas dengan SMOTE.....                    | 43 |
| 4.9                               | Pembagian Data .....                                     | 43 |
| 4.10                              | Pembangunan dan Optimasi Model Logistic Regression ..... | 44 |
| 4.11                              | Evaluasi Kinerja Model .....                             | 45 |
| 4.12                              | Frekuensi Kemunculan kata.....                           | 46 |
| 4.13                              | Analisis Kesalahan (Error Analysis).....                 | 51 |
| BAB V PENUTUP .....               |                                                          | 53 |
| 5.1                               | Kesimpulan .....                                         | 53 |
| 5.2                               | Saran .....                                              | 54 |
| REFERENSI .....                   |                                                          | 55 |

## DAFTAR TABEL

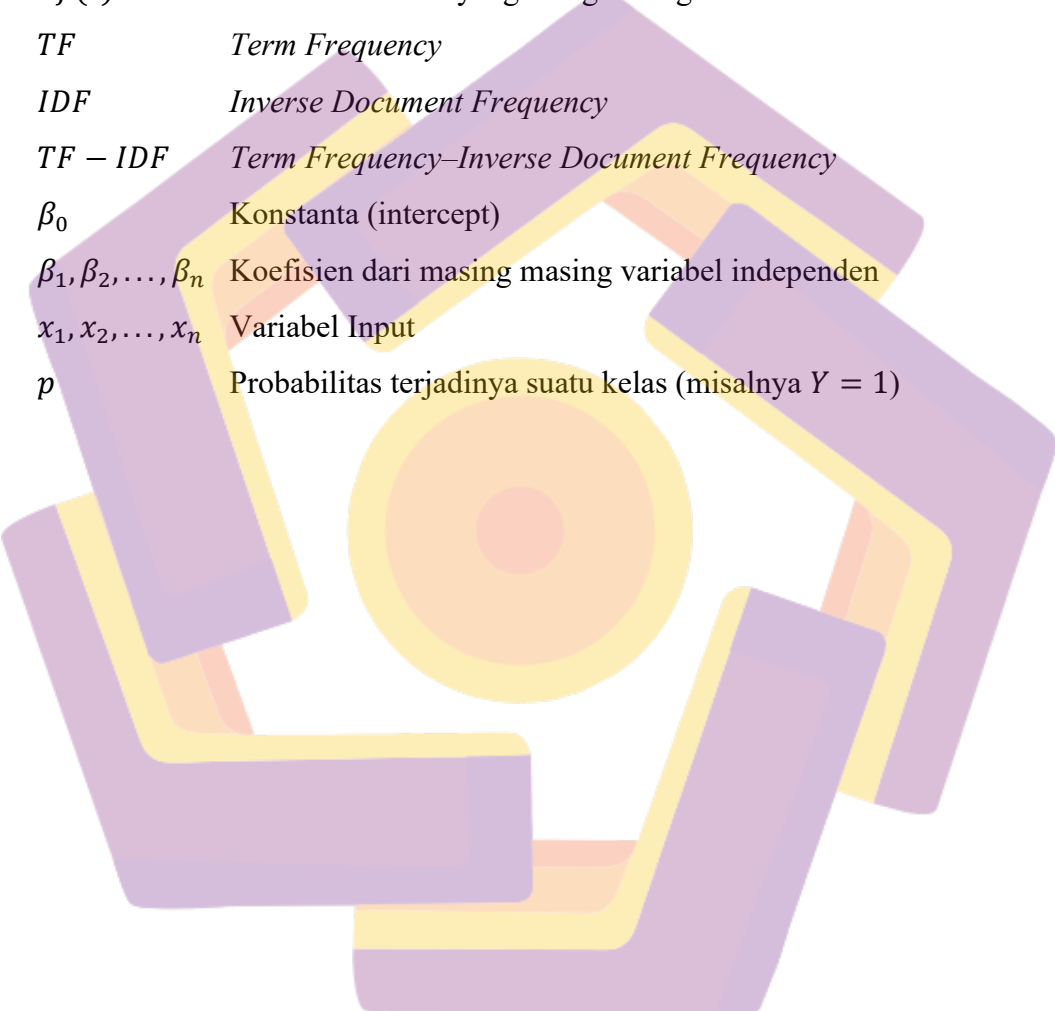
|                                            |    |
|--------------------------------------------|----|
| Tabel 2.1 Tabel Keaslian .....             | 12 |
| Tabel 2.2 <i>Confusion Matrix</i> .....    | 23 |
| Tabel 3.1 Perangkat Keras .....            | 33 |
| Tabel 3.2 Perangkat Lunak .....            | 34 |
| Tabel 4.1 Hasil <i>Preprocessing</i> ..... | 41 |
| Tabel 4.2 Hasil Penerapan Stemming .....   | 41 |
| Tabel 4.5 Kesalahan <i>Prediksi</i> .....  | 51 |



## DAFTAR GAMBAR

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| Gambar 3.1 Alur Penelitian .....                                          | 30 |
| Gambar 4.1 Dataset.....                                                   | 35 |
| Gambar 4.2 Distribusi <i>Rating</i> Ulasan Pengguna Aplikasi Shopee ..... | 36 |
| Gambar 4.3 Distribusi Panjang Teks Ulasan .....                           | 37 |
| Gambar 4.4 Distribusi Kelas Sentimen Positif, Negatif dan Netral .....    | 38 |
| Gambar 4.5 Perbandingan Distribusi Kelas Sebelum dan Sesudah SMOTE .....  | 43 |
| Gambar 4.6 Kombinasi Parameter Terbaik .....                              | 44 |
| Gambar 4.7 Hasil Evaluasi Kinerja Model .....                             | 45 |
| Gambar 4.8 <i>Confusion Matrix</i> Model Logistic Regression .....        | 46 |
| Gambar 4.9 Wordcloud kemunculan kata pada ulasan positif.....             | 47 |
| Gambar 4.10 Wordcloud kemunculan kata pada ulasan negatif.....            | 48 |
| Gambar 4.11 Wordcloud kemunculan kata pada ulasan netral.....             | 49 |
| Gambar 4.12 Wordcloud kemunculan kata ulasan. ....                        | 50 |

## DAFTAR LAMBANG DAN SINGKATAN



|                                    |                                                         |
|------------------------------------|---------------------------------------------------------|
| $t$                                | term (kata)                                             |
| $d$                                | Dokumen                                                 |
| $N$                                | Jumlah dokumen dalam korpus                             |
| $df(t)$                            | Jumlah dokumen yang mengandung kata $t$                 |
| $TF$                               | <i>Term Frequency</i>                                   |
| $IDF$                              | <i>Inverse Document Frequency</i>                       |
| $TF - IDF$                         | <i>Term Frequency-Inverse Document Frequency</i>        |
| $\beta_0$                          | Konstanta (intercept)                                   |
| $\beta_1, \beta_2, \dots, \beta_n$ | Koefisien dari masing masing variabel independen        |
| $x_1, x_2, \dots, x_n$             | Variabel Input                                          |
| $p$                                | Probabilitas terjadinya suatu kelas (misalnya $Y = 1$ ) |

## INTISARI

Perkembangan *e-commerce* yang pesat, khususnya pada aplikasi Shopee, mendorong banyak pengguna untuk memberikan ulasan terhadap sistem dan fitur layanan yang tersedia pada aplikasi tersebut. Ulasan tersebut mengandung opini yang dapat dimanfaatkan untuk mengetahui tingkat kepuasan pelanggan. Namun, jumlah ulasan yang sangat banyak menyebabkan sulitnya melakukan analisis secara manual. Oleh karena itu, diperlukan suatu metode otomatis untuk mengklasifikasikan sentimen ulasan pengguna.

Penelitian ini dilakukan untuk membantu proses analisis ulasan pengguna secara otomatis sehingga informasi mengenai kepuasan maupun keluhan pengguna terhadap aplikasi Shopee dapat diketahui dengan lebih cepat dan efisien. Pendekatan yang digunakan adalah *Natural Language Processing* (NLP) melalui tahapan *preprocessing* teks. Selanjutnya dilakukan ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengubah data teks menjadi representasi numerik. Proses klasifikasi dilakukan menggunakan algoritma *Logistic Regression* untuk mengelompokkan sentimen ke dalam tiga kategori, yaitu positif, negatif dan netral.

Hasil penelitian menunjukkan bahwa model yang dibangun mampu mengklasifikasikan sentimen dengan baik. Berdasarkan evaluasi menggunakan *confusion matrix*, diperoleh nilai akurasi sebesar 85%, dengan nilai *precision*, *recall*, dan *f1-score* pada kelas sentimen negatif masing-masing sebesar 0,86, 0,73, dan 0,79, pada kelas netral sebesar 0,84, 0,98, dan 0,90, serta pada kelas positif sebesar 0,85, 0,84, dan 0,84. Hasil tersebut menunjukkan bahwa metode yang digunakan cukup efektif dalam melakukan analisis sentimen terhadap ulasan pengguna Shopee.

Dengan demikian, penelitian ini dapat membantu dalam memahami opini pelanggan secara otomatis serta dapat dijadikan sebagai bahan pertimbangan bagi pelaku bisnis dalam meningkatkan kualitas produk dan layanan.

**Kata kunci:** Analisis Sentimen, Natural Language Processing, TF-IDF, Logistic Regression, Shopee

## **ABSTRACT**

*The rapid growth of e-commerce, especially in the Shopee application, has encouraged many users to provide reviews regarding the system and service features available on the platform. These reviews contain opinions that can be utilized to determine customer satisfaction levels. However, the large number of reviews makes manual analysis difficult. Therefore, an automatic method is needed to classify user review sentiments.*

*This study was conducted to assist the process of analyzing user reviews automatically so that information regarding user satisfaction and complaints about the Shopee application can be identified more quickly and efficiently. The approach used in this research is Natural Language Processing (NLP) through text preprocessing stages. Furthermore, feature extraction was carried out using the Term Frequency-Inverse Document Frequency (TF-IDF) method to transform text data into numerical representations. The classification process was performed using the Logistic Regression algorithm to categorize sentiments into three classes: positive, negative, and neutral.*

*The results showed that the developed model was able to classify sentiments effectively. Based on the evaluation using a confusion matrix, the model achieved an accuracy of 85%, with precision, recall, and f1-score values for the negative class of 0.86, 0.73, and 0.79, respectively, for the neutral class of 0.84, 0.98, and 0.90, and for the positive class of 0.85, 0.84, and 0.84. These results indicate that the proposed method is sufficiently effective in performing sentiment analysis on Shopee user reviews.*

*Therefore, this research can help in understanding customer opinions automatically and can be used as a consideration for businesses in improving product and service quality.*

**Keywords:** *Sentiment Analysis, Natural Language Processing, TF-IDF, Logistic Regression, Shopee*