

**PERBANDINGAN METODE NAIVE BAYES, SUPPORT VECTOR MACHINE, DAN  
LOGISTIC REGRESSION DALAM ANALISIS SENTIMEN KOMENTAR TIKTOK  
TERHADAP PENGELOLAAN SAMPAH DI INDONESIA DAN BELANDA**

**TUGAS AKHIR**



**Nofita Munir**

**122201019**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN ILMU KOMPUTER  
UNIVERSITAS BAKRIE  
JAKARTA  
2026**

**PERBANDINGAN METODE NAIVE BAYES, SUPPORT VECTOR MACHINE, DAN  
LOGISTIC REGRESSION DALAM ANALISIS SENTIMEN KOMENTAR TIKTOK  
TERHADAP PENGELOLAAN SAMPAH DI INDONESIA DAN BELANDA**

**TUGAS AKHIR**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom)**



**Nofita Munir**

**122201019**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN ILMU KOMPUTER  
UNIVERSITAS BAKRIE  
JAKARTA  
2026**

## HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Nofita Munir

NIM : 1222001019

Tanda Tangan : 

Tanggal : 15 September 2026

## HALAMAN PENGESAHAN

Proposal Tugas Akhir ini diajukan oleh :

Nama : Nofita Munir  
NIM : 1222001019  
Program Studi : Informatika  
Fakultas : Teknik dan Ilmu Komputer  
Judul Penelitian : Perbandingan Metode Naive Bayes, Support Vector Machine, dan Logistic Regression dalam Analisis Sentimen Komentar TikTok Terhadap Pengelolaan Sampah di Indonesia dan Belanda

Telah berhasil menyelesaikan tugas akhir dan diterima sebagai bagian persyaratan yang diperlukan dalam mengikuti sidang tugas akhir untuk memperoleh gelar Sarjana (S.Kom) pada Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

### DEWAN PENGUJI

Pembimbing 1 : Iwan Adhicandra, S.T., M.Sc, MIEEEE, MACM.



Pembimbing 2 : Dimas Aryo Anggoro, S.Kom., M.Sc.



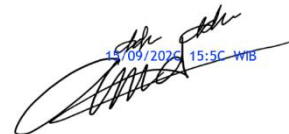
15/09/2026 9:05 WIB

Signed with  privy

Pembahas 1 : Berkah Iman Santoso, S.T., M.T.I



Pembahas 2 : Dewi Fatmawati Surianto, S.Kom., M.Kom.



15/09/2026 15:50 WIB

Ditetapkan di : Jakarta

Tanggal : 15 September 2026

## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama : Nofita Munir  
NIM : 1222001019  
Program Studi : Informatika  
Fakultas : Teknik dan Ilmu Komputer  
Jenis Tugas Akhir : Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul: **Perbandingan Metode Naive Bayes, Support Vector Machine, dan Logistic Regression dalam Analisis Sentimen Komentar TikTok Terhadap Pengelolaan Sampah di Indonesia dan Belanda**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data, merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 15 September 2026

Yang Menyatakan



Nofita Muni

## KATA PENGANTAR

Saya mengucapkan syukur kepada Tuhan yang maha Esa atas segala rahmat dan anugerah-nya yang telah memungkinkan saya untuk menyusun Proposal Tugas Akhir ini yang berjudul "Perbandingan Metode naive Bayes, Support Vector Machine, dan Logistic Regression dalam Analisis Sentimen Komentar TikTok Terhadap Pengelolaan Sampah di Indonesia dan Belanda" Proposal Tugas Akhir ini disusun sebagai salah satu persyaratan untuk mendapatkan gelar Sarjana Komputer dalam Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer di Universitas Bakrie. Selama proses penulisan proposal ini, saya menerima banyak bantuan, arahan, dan dukungan dari berbagai pihak. Oleh karena itu, saya ingin menyampaikan terima kasih kepada:

1. Kedua orang tua saya yang saya cintai, Mama Biba dan Papa Munir, berkat doa, cinta, bantuan, serta pengorbanan yang senantiasa mereka berikan kepada saya. Bantuan mereka menjadi pendorong bagi saya untuk menyelesaikan proposal ini.
2. Ibu Prof. Ir. Sofia W. Alisjahbana, M.Sc., Ph.D., IPU., ASEAN Eng. Selaku rektor Universitas Bakrie.
3. Bapak Dr. Mohammad Ihsan, S.T., M.Sc. selaku Dekan Fakultas Teknik dan Ilmu Komputer.
4. Bapak Iwan Adhicandra, S.T., M.Sc, MIEEE, MACM. Selaku Ketua Program Studi Teknik Informatika Universitas Bakrie. Sekaligus dosen pembimbing pertama saya yang telah memberikan bimbingan, arahan, dan dukungan selama proses penelitian dan penyusunan skripsi ini.
5. Bapak Dimas Aryo Anggoro, S.Kom., M.Sc. Selaku dosen pembimbing kedua saya yang telah membantu saya dalam proses penyusunan skripsi ini.
6. Seluruh dosen Universitas Bakrie, khususnya di lingkungan Program Studi Informatika, atas ilmu, bimbingan, serta dedikasi yang telah diberikan selama masa perkuliahan.
7. Saudara perempuan saya, yang senantiasa memberikan doa dan dukungan kepada saya dalam setiap langkah yang saya jalani.
8. Fenesia Samosir, yang adalah teman dekat selama kuliah, merupakan tempat untuk saling bertukar cerita, serta selalu menjadi sumber dukungan dan dorongan dalam aktivitas sehari-hari, termasuk dalam penyelesaian tugas akhir ini.
9. Teman-teman baik saya Talembor Squad: yaitu Theresia, Mifta, Killy, Pajar, Ilham, dan

Andy, yang senantiasa memberikan dukungan, kebersamaan, serta semangat dalam penyelesaian tugas akhir ini.

Jakarta, 15 September 2026

A handwritten signature in blue ink, appearing to read 'Nofita'.

Nofita Munir

1222001019

## ABSTRAK

### PERBANDINGAN METODE NAIVE BAYES, SUPPORT VECTOR MACHINE, DAN LOGISTIC REGRESSION DALAM ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP PENGELOLAAN SAMPAH DI INDONESIA DAN BELANDA

Nofita Munir  
1222001019@student.bakrie.ac.id

---

Permasalahan sampah masih menjadi isu lingkungan yang mendapat perhatian di banyak negara, termasuk Indonesia dan Belanda, yang memiliki sistem pengelolaan sampah berbeda dan kerap dibandingkan di media sosial, khususnya TikTok, meski opininya belum banyak diukur secara sistematis. Penelitian ini menganalisis sentimen komentar pengguna TikTok terhadap pengelolaan dan daur ulang sampah di kedua negara, sekaligus. Membandingkan kinerja tiga algoritma klasifikasi, yaitu Naive Bayes, Support Vector Machine, dan Logistic Regression. Data dikumpulkan dari tujuh unggahan TikTok menggunakan Apify, dengan komentar berbahasa Belanda diterjemahkan ke Bahasa Indonesia, kemudian melewati tahap preprocessing, pelabelan sentimen berbasis InSet Lexicon yang dilengkapi aturan negasi dan penanganan sarkasme, serta pembobotan kata menggunakan TF-IDF. Hasil analisis menunjukkan masyarakat Indonesia cenderung memberikan opini positif, sedangkan masyarakat Belanda cenderung lebih berimbang antara positif dan negatif. Pada perbandingan kinerja model, Logistic Regression memperoleh hasil terbaik dengan akurasi 75,91%, diikuti Support Vector Machine 74,62% dan Naive Bayes 71,83%, sehingga Logistic Regression menjadi algoritma paling optimal dalam penelitian ini. Belum tercapainya akurasi di atas 80% diduga dipengaruhi oleh keterbatasan data kelas netral dan karakteristik bahasa informal pada komentar media sosial.

**Kata kunci :**Analisi Sentimen, TikTok, Naive Bayes, Support Vector MACHine, Logistic Regression, TF-IDF, Pengelolaan Sampah

## **ABSTRACT**

### **A COMPARISON OF NAIVE BAYES, SUPPORT VECTOR MACHINE, AND LOGISTIC REGRESSION METHODS IN SENTIMENT ANALYSIS OF TIKTOK COMMENTS ON WASTE MANAGEMENT IN INDONESIA AND THE NETHERLANDS**

*Nofita Munir*  
*1222001019@student.bakrie.ac.id*

---

*Waste management remains a pressing environmental issue in many countries, including Indonesia and the Netherlands, which have different waste management systems that are frequently compared on social media, particularly TikTok, although public opinion in the comment sections has not been widely measured systematically. This study analyzes the sentiment of TikTok users' comments regarding waste management and recycling in both countries, while comparing the performance of three classification algorithms: Naive Bayes, Support Vector Machine, and Logistic Regression. Data were collected from seven TikTok uploads using Apify, with Dutch-language comments translated into Indonesian, then processed through preprocessing, rule-based sentiment labeling using the InSet Lexicon complemented by negation handling and sarcasm detection, and then weighting using TF-IDF. The results show that Indonesian users tend to express positive opinions, while Dutch users show a more balanced sentiment between positive and negative. In the model comparison, Logistic Regression achieved the best result with an accuracy of 75.91%, followed by Support Vector Machine at 74.62% and Naive Bayes at 71.83%, making Logistic Regression the most optimal algorithm in this study. The failure to reach accuracy above 80% is presumably influenced by the limited data in the neutral class and the informal language characteristics of social media comments.*

**Keywords :***Sentiment Analysis, TikTok, Naive Bayes, Support Vector Machine, Logistic Regression, TF-IDF, Waste Management*

## DAFTAR ISI

|                                                                                         |     |
|-----------------------------------------------------------------------------------------|-----|
| KATA PENGANTAR.....                                                                     | iv  |
| ABSTRAK .....                                                                           | vi  |
| ABSTRACT .....                                                                          | vii |
| DAFTAR GAMBAR.....                                                                      | xi  |
| DAFTAR TABEL.....                                                                       | xii |
| BAB I PENDAHULUAN .....                                                                 | 1   |
| 1.1 Latar Belakang .....                                                                | 1   |
| 1.2 Rumusan Masalah .....                                                               | 3   |
| 1.3 Batasan Masalah.....                                                                | 3   |
| 1.4 Tujuan Penelitian .....                                                             | 4   |
| 1.5 Manfaat Penelitian .....                                                            | 5   |
| 1.6 Sistematika Penulisan.....                                                          | 6   |
| BAB II TINJAUAN PUSTAKA.....                                                            | 7   |
| 2.1 Tinjauan Literatur .....                                                            | 7   |
| 2.2 Landasan teori .....                                                                | 14  |
| 2.2.1 Analisis Sentimen .....                                                           | 14  |
| 2.2.2 <i>Natural Language Processing</i> .....                                          | 15  |
| 2.2.3 <i>Preprocessing Teks</i> .....                                                   | 16  |
| 2.2.4 Transformasi Data <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> ..... | 17  |
| 2.2.5 <i>Rule - Based</i> .....                                                         | 19  |
| 2.2.6 Deteksi Sarkasme .....                                                            | 21  |
| 2.2.7 Naive Bayes.....                                                                  | 22  |
| 2.2.8 Support Vector Machine .....                                                      | 23  |
| 2.2.10 <i>Confusion Matrix</i> dan Evaluasi Model.....                                  | 26  |
| 2.2.11 Media Sosial TikTok.....                                                         | 29  |
| 2.2.12 <i>Web Scraping</i> dan Apify .....                                              | 30  |
| 2.2.13 Pengelolaan dan Daur Ulang Sampah .....                                          | 32  |
| 2.2.14 Kerangka Pemikiran .....                                                         | 33  |
| BAB III METODOLOGI PENELITIAN .....                                                     | 36  |
| 3.1 Jenis Penelitian .....                                                              | 36  |

|                                                              |    |
|--------------------------------------------------------------|----|
| 3.2 Alur Penelitian .....                                    | 36 |
| 3.3 Sumber Data .....                                        | 37 |
| 3.4 Tahapan Pengolahan Data.....                             | 41 |
| 3.4.1 <i>Data Selection</i> .....                            | 41 |
| 3.4.2 Penerjemahan Data .....                                | 42 |
| 3.4.3 <i>Text Preprocessing</i> .....                        | 42 |
| 3.4.4 <i>Labeling Data Menggunakan Rule-Based</i> .....      | 43 |
| 3.4.5 Transformasi Data Mengguna kan TF-IDF .....            | 45 |
| 3.5 Metode Klasifikasi .....                                 | 46 |
| 3.5.1 Pembagian Data Latih dan Data Uji.....                 | 47 |
| 3.5.2 Naive Bayes.....                                       | 48 |
| 3.5.3 Support Vector Machine .....                           | 48 |
| 3.6 Evaluasi Model .....                                     | 49 |
| 3.7 Alat dan Bahan Penelitian .....                          | 50 |
| 3.7.1 Alat Penelitian .....                                  | 50 |
| 3.7.2 Bahan Penelitian .....                                 | 51 |
| BAB IV HASIL DAN PEMBAHASAN.....                             | 53 |
| 4.1 Hasil Pengumpulan Data .....                             | 53 |
| 4.2 Hasil Seleksi dan Penerjemahan Data .....                | 54 |
| 4.2.1 <i>Data Selection</i> .....                            | 54 |
| 4.2.2 Penerjemahan Data .....                                | 55 |
| 4.3 Hasil <i>Text Preprocessing</i> .....                    | 56 |
| 4.3.1 <i>Cleaning dan Case Folding</i> .....                 | 56 |
| 4.3.3.1 <i>Cleaning</i> .....                                | 56 |
| 4.3.3.2 <i>Case Folding</i> .....                            | 57 |
| 4.3.2 <i>Normalization</i> .....                             | 58 |
| 4.3.3 <i>Tokenizing dan Stopword Removal</i> .....           | 58 |
| 4.3.3.1 <i>Tokenizing</i> .....                              | 59 |
| 4.3.3.2 <i>Stopword Removal</i> .....                        | 59 |
| 4.3.4 <i>Stemming</i> .....                                  | 60 |
| 4.4 Hasil Pelabelan Data Menggunakan <i>Rule-Based</i> ..... | 61 |
| 4.4.1 Pelabelan dengan <i>InSet Lexicon</i> .....            | 61 |

|                                                      |    |
|------------------------------------------------------|----|
| 4.4.2 Penerapan Aturan Negasi .....                  | 62 |
| 4.4.3 Penanganan Sarkasme.....                       | 63 |
| 4.5 Hasil Kecenderungan Opini Masyarakat.....        | 65 |
| 4.6 Hasil Pembagian Data dan Transformasi Data ..... | 66 |
| 4.6.1 Pembagian Data Latih dan Data Uji.....         | 66 |
| 4.6.2 Transformasi Data Menggunakan TF-IDF .....     | 67 |
| 4.6.3 Penanganan Ketidakseimbangan Data .....        | 68 |
| 4.7 Hasil Klasifikasi dan Evaluasi Model .....       | 68 |
| 4.7.1 Naive Bayes.....                               | 69 |
| 4.7.2 Support Vector Machine .....                   | 69 |
| 4.7.3 Logistic Regression .....                      | 70 |
| 4.8 Perbandingan Kinerja Model.....                  | 71 |
| 4.9 Pembahasan .....                                 | 72 |
| BAB V PENUTUP .....                                  | 74 |
| 5.1 Kesimpulan.....                                  | 74 |
| DAFTAR PUSTAKA .....                                 | 76 |
| LAMPIRAN .....                                       | 83 |

## DAFTAR GAMBAR

|                                                            |    |
|------------------------------------------------------------|----|
| Gambar 2. 1 Alur Pengambilan Data.....                     | 32 |
| Gambar 2. 2 Kerangka Pemikiran .....                       | 33 |
| Gambar 3. 1 Alur Penelitian .....                          | 37 |
| Gambar 4. 1 Hasil Sentimen Berdasarkan Negara .....        | 67 |
| Gambar 4. 2 Confusion Matrix Ketiga Model.....             | 72 |
| Gambar 4. 3 Grafik Perbandingan Kinerja Ketiga Model ..... | 73 |

## DAFTAR TABEL

|                                                                      |    |
|----------------------------------------------------------------------|----|
| Tabel 2. 1 Tabel Penelitian Terdahulu .....                          | 10 |
| Tabel 3. 1 Sumber Data Penelitian .....                              | 38 |
| Tabel 3. 2 Jumlah Data Penelitian .....                              | 40 |
| Tabel 3. 3 Cuplikan data Komentar Indonesia .....                    | 41 |
| Tabel 3. 4 Cuplikan Data Komentar Belanda .....                      | 41 |
| Tabel 3. 5 Alat Penelitian.....                                      | 51 |
| Tabel 3. 6 Bahan Penelitian .....                                    | 52 |
| Tabel 4. 1 Data Hasil Pengumpulan .....                              | 54 |
| Tabel 4. 2 Hasil Data Selection.....                                 | 55 |
| Tabel 4. 3 Data Sebelum dan Sesudah Penerjemahan .....               | 56 |
| Tabel 4. 4 Hasil Cleaning.....                                       | 57 |
| Tabel 4. 5 Hasil Case Folding.....                                   | 58 |
| Tabel 4. 6 Hasil Normalization.....                                  | 59 |
| Tabel 4. 7 Hasil Tokenizing .....                                    | 60 |
| Tabel 4. 8 Hasil Stopword Removal .....                              | 61 |
| Tabel 4. 9 Hasil Stemming.....                                       | 61 |
| Tabel 4. 10 Hasil Pelabelan dengan InsetLexicon .....                | 63 |
| Tabel 4. 11 Penerapan Aturan Negasi .....                            | 64 |
| Tabel 4. 12 Hasil Koreksi Sarkasme.....                              | 65 |
| Tabel 4. 13 Hasil Sentimen Berdasarkan Negara .....                  | 66 |
| Tabel 4. 14 Pembagian Data Latih dan Data Uji .....                  | 68 |
| Tabel 4. 15 Hasil Pembobotan TF-IDF .....                            | 68 |
| Tabel 4. 16 Jumlah Data Latih Sebelum dan Sesudah Oversampling ..... | 69 |
| Tabel 4. 17 Confusion Matrix Naive Bayes.....                        | 70 |
| Tabel 4. 18 Confusion Matrix SVM.....                                | 70 |
| Tabel 4. 19 Confusion Matrix Logistic Regression .....               | 71 |
| Tabel 4. 20 Perbandingan Kinerja Ketiga Model .....                  | 72 |