

**KLASIFIKASI KOMENTAR PENGGUNA TWITTER/X TERHADAP  
KEGAGALAN TIMNAS INDONESIA LOLOS PIALA DUNIA  
MENGUNAKAN SUPPORT VECTOR MACHINE (SVM)**

**SKRIPSI SARJANA SISTEM INFORMASI**



Oleh

Syamsul Bakhri

227006516081

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNOLOGI KOMUNIKASI DAN  
INFORMATIKA  
UNIVERSITAS NASIONAL**

**2025**

**KLASIFIKASI KOMENTAR PENGGUNA TWITTER/X  
TERHADAP KEGAGALAN TIMNAS INDONESIA LOLOS  
PIALA DUNIA MENGGUNAKAN METODE SUPPORT  
VECTOR MACHINE (SVM)  
SKRIPSI SARJANA**

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem  
Informasi dari Fakultas Teknologi Komunikasi dan Informatika



Oleh  
Syamsul Bakhri  
227006516081

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNOLOGI KOMUNIKASI DAN  
INFORMATIKA  
UNIVERSITAS NASIONAL  
2025**

**HALAMAN PENGESAHAN**

**TUGAS AKHIR**

**KLASIFIKASI KOMENTAR PENGGUNA TWITTER/X  
TERHADAP KEGAGALAN TIMNAS INDONESIA LOLOS  
PIALA DUNIA MENGGUNAKAN METODE SUPPORT**



Dr. Aris Gunaryati, S.Si., MMSI.  
NIDN. 0313087705

## PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul:

### **KLASIFIKASI KOMENTAR PENGGUNA TWITTER/X TERHADAP KEGAGALAN TIMNAS INDONESIA LOLOS PIALA DUNIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)**

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 27 Februari 2026



NPM. 227006516081

## LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul:

**KLASIFIKASI KOMENTAR PENGGUNA TWITTER/X  
TERHADAP KEGAGALAN TIMNAS INDONESIA LOLOS  
PIALA DUNIA MENGGUNAKAN METODE SUPPORT  
VECTOR MACHINE (SVM)**

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2025-2026 pada tanggal 24 Februari Tahun 2026

Dosen Pembimbing 1



Dr. Aris Gunaryati, S.Si., MMSI.  
NIDN: 0313087705

Ketua Program Studi



Dr. Ir. Andrianingsih, S.Kom., MMSI.  
NIDN: 0303097902

## LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Syamsul Bakhri  
 NPM : 227006516081  
 Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika  
 Program Studi : Sistem Informasi  
 Tanggal Sidang : 24 Februari 2026

JUDUL DALAM BAHASA INDONESIA:

Klasifikasi Komentar Pengguna Twitter/X terhadap Kegagalan  
 Timnas Indonesia Lolos Piala Dunia Menggunakan Metode  
 Support Vector Machine (SVM)

JUDUL DALAM BAHASA INGGRIS:

Classification of Twitter/X User Comments on the Failure of the  
 Indonesian National Football Team to Qualify for the World Cup  
 Using Support Vector Machine (SVM) Method

### TANDA TANGAN DAN TANGGAL

| Pembimbing 1                                                                        | Ka. Prodi                                                                           | Mahasiswa                                                                             |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| TGL : 2 Maret 2026                                                                  | TGL : 2 Maret 2026                                                                  | TGL : 2 Maret 2026                                                                    |
|  |  |  |
| Dr. Aris Gunaryati, S.Si.,<br>MMSI.<br>NIDN: 0313087705                             | Dr. Ir. Andrianingsih,<br>S.Kom., MMSI.<br>NIDN: 0303097902                         | Syamsul Bakhri<br>NPM: 227006516081                                                   |

## KATA PENGANTAR

Puji dan syukur Penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia sehingga Penulis dapat menyelesaikan skripsi dengan judul **“Klasifikasi Komentar Pengguna Twitter/X Terhadap Kegagalan Timnas Indonesia Lolos Piala Dunia Menggunakan Metode Support Vector Machine (SVM)”** sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Program Studi Sistem Informasi.

Penelitian skripsi ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak terima kasih terutama kepada dosen pembimbing Tugas Akhir, Dr. Aris Gunaryati, S.Si., MMSI. Yang telah meluangkan banyak waktu, tenaga, pikiran, arahan, motivasi serta memaklumi segala kekurangan penulis selama penelitian tugas akhir dan penyusunan skripsi. Penulis juga mengucapkan banyak terima kasih kepada:

1. Kepada Ibu Suhermi dan Ibu Leny Daisy Astri selaku orang tua penulis atas doa, kasih sayang, serta dukungan moral dan motivasi yang tiada henti.
2. Bapak Dr. Agung Triayudi, S.Kom., M.Kom., selaku Dekan Fakultas Teknologi Komunikasi dan Informatika (FTKI) Universitas Nasional.
3. Ibu Dr. Ir. Andrianingsih, S.Kom., MMSI. selaku Ketua Program Studi Sistem Informasi Universitas Nasional.
4. Seluruh dosen pengajar di Program Studi Sistem Informasi FTKI maupun dosen di Program Studi lain yang memberikan banyak ilmu.
5. Teman-teman Segardu yang telah membantu, mendukung, dan memotivasi penulis.

Akhir kata, semoga Tuhan Yang Maha Esa membalas kebaikan dan bantuan yang telah diberikan dengan hal yang lebih baik. Penulis mengharapkan kritik dan saran dan semoga skripsi ini dapat memberikan manfaat di bidang Teknologi Informatika.

Jakarta, 17 Desember 2025.



Penulis

## ABSTRAK

Penelitian ini bertujuan memahami perasaan masyarakat di media sosial Twitter/X tentang kegagalan Timnas Indonesia lolos ke Piala Dunia, yang memunculkan beragam opini publik dalam bentuk teks tidak terstruktur. Banyaknya komentar dengan variasi bahasa, singkatan, dan noise membuat analisis manual kurang efisien, sehingga diperlukan pendekatan text mining untuk mengklasifikasikan sentimen secara otomatis. Data komentar dikumpulkan melalui web scraping menggunakan Harvest. Tahap praproses mencakup pembersihan tautan, tanda baca, dan emotikon, case folding, tokenisasi, stopwords removal, serta stemming. Pelabelan sentimen dilakukan secara otomatis menggunakan TextBlob berdasarkan nilai *polarity* untuk membentuk tiga kelas sentimen, yaitu positif, netral, dan negatif. Representasi fitur teks dibangun dengan TF-IDF, kemudian proses klasifikasi dilakukan menggunakan Support Vector Machine (SVM) kernel linear. Dataset dibagi dengan rasio 80:20 menggunakan *stratified split* agar proporsi kelas tetap seimbang pada data latih dan uji. Evaluasi model dilakukan menggunakan confusion matrix serta metrik Accuracy, macro-Precision, macro-Recall, macro-F1 score, dan ROC-AUC (OvR macro). Hasil pengujian menunjukkan SVM mencapai Accuracy 0,7588; macro-Precision 0,7647; macro-Recall 0,7557; macro-F1 score 0,7584; dan ROC-AUC 0,9599, dengan 475 prediksi benar dari 626 data uji. Dibandingkan Naive Bayes dan KNN, SVM memberikan performa terbaik pada metrik utama. Dengan demikian, kombinasi TF-IDF dan SVM efektif untuk klasifikasi sentimen multikelas pada data Twitter/X yang berdimensi tinggi dan bervariasi.

**Kata kunci:** Twitter/X, analisis sentimen, TF-IDF, Support Vector Machine, Timnas Indonesia.

## ABSTRACT

This study aims to understand public feelings on Twitter/X regarding Indonesia's national team failure to qualify for the World Cup, which generated a large amount of public opinion in the form of unstructured text. The volume of comments, language variations, abbreviations, and noise makes manual analysis inefficient; therefore, a text mining approach is required to automatically classify sentiment. Twitter/X comments were collected through web scraping using Harvest. Text preprocessing included removing URLs, punctuation, and emojis, followed by case folding, tokenization, stopword removal, and stemming. Sentiment labels were generated automatically using TextBlob based on polarity scores, producing three sentiment classes: positive, neutral, and negative. Text features were represented using TF-IDF, and classification was performed using a linear-kernel Support Vector Machine (SVM). The dataset was split into training and testing sets with an 80:20 stratified split to preserve class proportions. Model performance was evaluated using a confusion matrix and the following metrics: Accuracy, macro-Precision, macro-Recall, macro-F1 score, and macro ROC-AUC (one-vs-rest). The experimental results show that the SVM model achieved an Accuracy of 0.7588, macro-Precision of 0.7647, macro-Recall of 0.7557, macro-F1 score of 0.7584, and a macro ROC-AUC of 0.9599, with 475 correct predictions out of 626 test samples. Compared to Naive Bayes and KNN, SVM produced the best performance on the main evaluation metrics. Therefore, the combination of TF-IDF and SVM is effective for multi-class sentiment classification on high-dimensional and diverse Twitter/X text data.

**Keywords:** *Twitter/X, sentiment analysis, TF-IDF, Support Vector Machine, Timnas Indonesia.*

## DAFTAR ISI

|                                                           |      |
|-----------------------------------------------------------|------|
| HALAMAN PENGESAHAN.....                                   | iii  |
| PERNYATAAN KEASLIAN TUGAS AKHIR .....                     | iv   |
| LEMBAR PERSETUJUAN REVIEW AKHIR.....                      | v    |
| LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI .. | vi   |
| KATA PENGANTAR.....                                       | vii  |
| ABSTRAK.....                                              | viii |
| ABSTRACT.....                                             | ix   |
| DAFTAR ISI.....                                           | x    |
| DAFTAR GAMBAR .....                                       | xiv  |
| DAFTAR TABEL.....                                         | xv   |
| BAB I PENDAHULUAN.....                                    | 1    |
| 1.1 Latar Belakang.....                                   | 1    |
| 1.2 Identifikasi Masalah.....                             | 4    |
| 1.3 Batasan Masalah.....                                  | 4    |
| 1.4 Tujuan Penelitian.....                                | 5    |
| 1.5 Kontribusi.....                                       | 5    |
| BAB II TINJAUAN PUSTAKA.....                              | 6    |
| 2.1 Landasan Teori .....                                  | 6    |
| 2.1.1 Media Sosial.....                                   | 6    |
| 2.1.2 Twitter/X sebagai Sumber Data Opini Publik.....     | 7    |
| 2.1.3 Klasifikasi Sentimen .....                          | 8    |
| 2.1.4 Text Mining.....                                    | 11   |
| 2.1.5 Web Scraping .....                                  | 11   |

|                                            |                                                          |           |
|--------------------------------------------|----------------------------------------------------------|-----------|
| 2.1.6                                      | Preprocessing Teks .....                                 | 12        |
| 2.1.7                                      | Term Frequency–Inverse Document Frequency (TF-IDF) ..... | 14        |
| 2.1.8                                      | Support Vector Machine (SVM) .....                       | 14        |
| 2.1.9                                      | Evaluasi Model Klasifikasi .....                         | 16        |
| 2.2                                        | Penelitian Terdahulu .....                               | 18        |
| <b>BAB III METODOLOGI PENELITIAN .....</b> |                                                          | <b>23</b> |
| 3.1                                        | Lokasi Penelitian .....                                  | 23        |
| 3.2                                        | Waktu Penelitian .....                                   | 23        |
| 3.3                                        | Alur Penelitian .....                                    | 24        |
| 3.3.1                                      | Penentuan Subjek Penelitian .....                        | 24        |
| 3.3.2                                      | Fokus Penelitian .....                                   | 25        |
| 3.3.3                                      | Sumber Data .....                                        | 25        |
| 3.3.4                                      | Teknik Pengumpulan Data .....                            | 26        |
| 3.3.5                                      | Tahap Pemodelan SVM Baseline .....                       | 26        |
| 3.3.6                                      | Data Scraping .....                                      | 27        |
| 3.3.7                                      | Data Selection .....                                     | 29        |
| 3.3.8                                      | Preprocessing .....                                      | 29        |
| 3.3.9                                      | Modelling .....                                          | 29        |
| 3.3.10                                     | Formulasi Algoritma SUPPORT VECTOR MACHINE (SVM) ..      | 30        |
| 3.3.11                                     | TF-IDF Weighting .....                                   | 34        |
| 3.4                                        | Evaluation .....                                         | 35        |
| 3.5                                        | Aspek Etika dan Legalitas Data .....                     | 36        |
| <b>BAB IV HASIL DAN DISKUSI .....</b>      |                                                          | <b>37</b> |
| 4.1                                        | Hasil Pengumpulan Data Scraping Twitter/X .....          | 37        |
| 4.2                                        | Data Selection .....                                     | 38        |

|                     |                                                            |    |
|---------------------|------------------------------------------------------------|----|
| 4.3                 | Preprocessing Teks Menggunakan TextBlob .....              | 39 |
| 4.3.1               | Cleaning Text .....                                        | 39 |
| 4.3.2               | Translasi Bahasa Indonesia ke Bahasa Inggris .....         | 39 |
| 4.3.3               | Tokenisasi dan Normalisasi dengan TextBlob .....           | 40 |
| 4.4                 | Pelabelan Otomatis Sentimen menggunakan TextBlob .....     | 41 |
| 4.5                 | Transformasi Fitur Menggunakan TF-IDF .....                | 41 |
| 4.6                 | Pembagian Data Latih dan Data Uji .....                    | 41 |
| 4.7                 | Hasil Klasifikasi Menggunakan Support Vector Machine ..... | 42 |
| 4.8                 | Analisis Confusion Matrix .....                            | 42 |
| 4.9                 | Visualisasi dan Interpretasi Hasil .....                   | 44 |
| 4.9.1               | Grafik Metrik Utama .....                                  | 44 |
| 4.9.2               | ROC Curve .....                                            | 44 |
| 4.9.3               | WordCloud .....                                            | 45 |
| 4.10                | Perbandingan Algoritma SVM, Naive Bayes, Dan KNN .....     | 48 |
| 4.10.1              | Tujuan Perbandingan .....                                  | 48 |
| 4.10.2              | Skenario Pengujian dan Parameter Model .....               | 48 |
| 4.10.3              | Hasil Perbandingan Metrik Evaluasi .....                   | 49 |
| 4.10.4              | Analisis Hasil per Kelas (Classification Report) .....     | 49 |
| 4.10.5              | Pembahasan Perbedaan Performa Algoritma .....              | 55 |
| 4.10.6              | Perhitungan Manual Metrik Evaluasi .....                   | 56 |
| 4.11                | Pembahasan .....                                           | 61 |
| 4.12                | Keterkaitan Hasil dengan Tujuan Penelitian .....           | 62 |
| BAB V PENUTUP ..... |                                                            | 63 |
| 5.1                 | Kesimpulan .....                                           | 63 |
| 5.2                 | Saran .....                                                | 64 |

|                      |    |
|----------------------|----|
| DAFTAR PUSTAKA ..... | 65 |
| LAMPIRAN .....       | 68 |



## DAFTAR GAMBAR

|                                                          |    |
|----------------------------------------------------------|----|
| Gambar 3. 1 Alur Permodelan SVM.....                     | 27 |
| Gambar 3. 2 Diagram Pengumpulan Data .....               | 28 |
| Gambar 3. 3 Transformasi TF-IDF dan Klasifikasi SVM..... | 34 |
| Gambar 4. 1 Hasil Scraping Twitter/X.....                | 37 |
| Gambar 4. 2 Hasil Tahap Seleksi Data.....                | 38 |
| Gambar 4. 3 Hasil Data setelah Cleaning.....             | 39 |
| Gambar 4. 4 Hasil Convert ID-EN.....                     | 40 |
| Gambar 4. 5 Hasil Preprocessing TextBlob .....           | 40 |
| Gambar 4. 6 Distribusi Tabel dengan TextBlob .....       | 41 |
| Gambar 4. 7 Hasil Split Data .....                       | 42 |
| Gambar 4. 8 Confusion Matrix .....                       | 44 |
| Gambar 4. 9 ROC-AUC .....                                | 45 |
| Gambar 4. 10 Wordcloud Semua Komentar .....              | 46 |
| Gambar 4. 11 Wordcloud Komentar Negatif .....            | 47 |
| Gambar 4. 12 WordCloud Komentar Positif.....             | 47 |
| Gambar 4. 13 WordCloud Komentar Netral .....             | 48 |
| Gambar 4. 14 Classification Report SVM .....             | 50 |
| Gambar 4. 15 Classification Report naïve Bayes.....      | 51 |
| Gambar 4. 16 Classification Report KNN .....             | 52 |
| Gambar 4. 17 Confusion Matrix SVM.....                   | 52 |
| Gambar 4. 18 Confusion Matrix Naïve Bayes .....          | 53 |
| Gambar 4. 19 Confusion Matrix KNN.....                   | 54 |
| Gambar 4. 20 ROC Curve SVM .....                         | 59 |
| Gambar 4. 21 ROC Curve Naïve Bayes .....                 | 60 |
| Gambar 4. 22 ROC Curve KNN .....                         | 60 |

## DAFTAR TABEL

|                                                              |    |
|--------------------------------------------------------------|----|
| Tabel 2. 1 Penelitian Terdahulu.....                         | 19 |
| Tabel 2. 2 Perbandingan SVM, KNN dan Naive Bayes .....       | 22 |
| Tabel 3. 1 Waktu Penelitian.....                             | 23 |
| Tabel 3. 2 Tabel Confusion Matrix.....                       | 32 |
| Tabel 4. 1 Hasil Evaluasi Model.....                         | 42 |
| Tabel 4. 2 Perbandingan Performa Model (Macro Average) ..... | 49 |
| Tabel 4. 3 Tabel Perbandingan Confusion Matrix.....          | 54 |
| Tabel 4. 4 Tabel Perbandingan ROC-AUC .....                  | 61 |

