

**KOMPARASI MODEL KLASIFIKASI NAÏVE BAYES DAN  
SVM DALAM ANALISIS SENTIMEN RESPON PUBLIK PADA  
MOMEN 100 HARI KERJA PRABOWO DI PLATFORM X**

**SKRIPSI**

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



disusun oleh

**ALFONSUS FERDY MAHARDEKA PUTRA**

**22.12.2310**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2026**

**KOMPARASI MODEL KLASIFIKASI NAÏVE BAYES DAN  
SVM DALAM ANALISIS SENTIMEN RESPON PUBLIK PADA  
MOMEN 100 HARI KERJA PRABOWO DI PLATFORM X**

**SKRIPSI**

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



disusun oleh

**ALFONSUS FERDY MAHARDEKA PUTRA**

**22.12.2310**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2026**

HALAMAN PERSETUJUAN

SKRIPSI

KOMPARASI MODEL KLASIFIKASI NAÏVE BAYES DAN SVM DALAM  
ANALISIS SENTIMEN RESPON PUBLIK PADA MOMEN 100 HARI  
KERJA PRABOWO DI PLATFORM X

yang disusun dan diajukan oleh

**Alfonsus Ferdy Mahardeka Putra**

**22.12.2310**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 9 Februari 2026

Dosen Pembimbing,



**Emha Taufiq Luthfi, S.T., M.Kom., Ph.D.**  
**NIK. 190302125**

HALAMAN PENGESAHAN

SKRIPSI

KOMPARASI MODEL KLASIFIKASI NAÏVE BAYES DAN SVM DALAM  
ANALISIS SENTIMEN RESPON PUBLIK PADA MOMEN 100 HARI  
KERJA PRABOWO DI PLATFORM X

yang disusun dan diajukan oleh

Alfonsus Ferdy Mahardeka Putra

22.12.2310

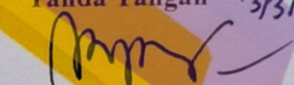
Telah dipertahankan di depan Dewan Penguji  
pada tanggal 19 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan 3/3/26

Krisnawati, S.Si., M.T.  
NIK. 190302038



Alva Hendi Muhammad, A.Md., S.T., M.Eng., Ph.D.  
NIK. 190302493



Emha Taufiq Luthfi, S.T., M.Kom., Ph.D.  
NIK. 190302125



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 19 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Alfonsus Ferdy Mahardeka Putra  
NIM : 22.12.2310

Menyatakan bahwa Skripsi dengan judul berikut:

**Komparasi Model Klasifikasi Naïve Bayes dan SVM dalam Analisis Sentimen Respon Publik pada Momen 100 Hari Kerja Prabowo di Platform X**

Dosen Pembimbing : Emha Taufiq Luthfi, S.T., M.Kom., Ph.D.

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, 19 Februari 2026

Yang Menyatakan,



Alfonsus Ferdy Mahardeka Putra

## HALAMAN PERSEMBAHAN

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, dan anugerah-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Penyusunan skripsi ini merupakan hasil dari proses Panjang yang tidak lepas dari berbagai tantangan, keterbatasan, serta pembelajaran berharga yang penulis lalui selama masa perkuliahan. Berkat bimbingan, dukungan, doa, dan motivasi dari berbagai pihak, Maka skripsi ini penulis persembahkan sebagai wujud dedikasi kepada :

1. Orang tua tercinta, yang telah memberikan doa, kasih sayang, dukungan moral maupun material, serta kesabaran dan pengorbanan yang tiada henti selama penulis menempuh Pendidikan hingga penyusunan skripsi ini selesai.
2. Kakak-kakak tercinta, Ignatius Setyo Pratama Putra dan Angela Rosy Pratiwi Putri, yang senantiasa memberikan motivasi, semangat, dan dukungan kepada penulis dalam menghadapi berbagai tantangan selama proses perkuliahan dan penyusunan skripsi.
3. Dosen Pembimbing, Bapak Emha Taufiq Luthfi, S.T., M.Kom., Ph.D. yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan serta masukan yang sangat berarti bagi penulis dalam penyelesaian skripsi ini.
4. Teman-teman, Zaki, Ricky, Zein, Devin, dan Biel serta seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan kebersamaan, dukungan, bantuan, dan semangat selama masa perkuliahan

## KATA PENGANTAR

Puji dan syukur penulis panjarkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan anugerah-Nya, sehingga skripsi yang berjudul “Komparasi Model Klasifikasi Naïve Bayes dan SVM dalam Analisis Sentimen Respon Publik pada Momen 100 Hari Kerja Prabowo di Platform X” dapat diselesaikan dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa penyelesaiannya tidak terlepas dari bantuan, bimbingan, arahan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan 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. Bapak Anggit Dwi Hartanto, M.Kom, selaku Ketua Prodi Sistem Informasi Universitas Amikom Yogyakarta
4. Bapak Emha Taufiq Luthfi, S.T., M.Kom., Ph.D. selaku dosen pembimbing, yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta masukan yang sangat berharga dalam penyusunan skripsi ini.

Yogyakarta, 1 Februari 2026

Penulis

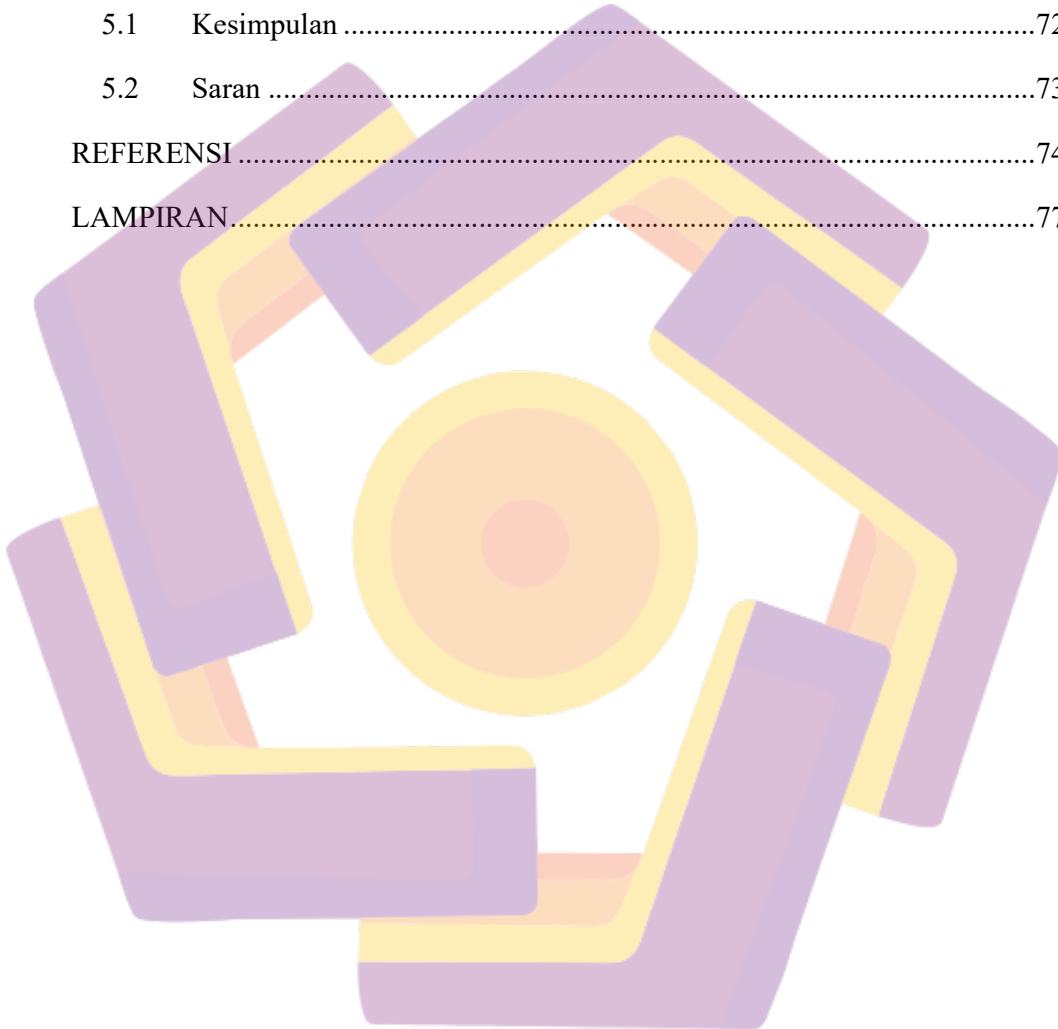
## DAFTAR ISI

|                                          |      |
|------------------------------------------|------|
| 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.....                        | xi   |
| DAFTAR GAMBAR .....                      | xii  |
| DAFTAR LAMPIRAN.....                     | xiii |
| DAFTAR LAMBANG DAN SINGKATAN .....       | xiv  |
| DAFTAR ISTILAH .....                     | xv   |
| INTISARI .....                           | xvi  |
| ABSTRACT.....                            | xvii |
| BAB I PENDAHULUAN.....                   | 1    |
| 1.1 Latar Belakang .....                 | 1    |
| 1.2 Rumusan Masalah .....                | 2    |
| 1.3 Batasan Masalah .....                | 3    |
| 1.4 Tujuan Penelitian .....              | 3    |
| 1.5 Manfaat Penelitian .....             | 4    |
| 1.6 Sistematika Penulisan .....          | 5    |
| BAB II TINJAUAN PUSTAKA .....            | 6    |
| 2.1 Studi Literatur .....                | 6    |

|                                   |                                                  |    |
|-----------------------------------|--------------------------------------------------|----|
| 2.2                               | Dasar Teori.....                                 | 17 |
| 2.2.1.                            | Analisis Sentimen .....                          | 17 |
| 2.2.2.                            | Pra-pemrosesan teks.....                         | 17 |
| 2.2.3.                            | Term Frequency-Inverse Document Frequency .....  | 17 |
| 2.2.4.                            | Data Imbalanced .....                            | 18 |
| 2.2.5.                            | Synthetic Minority Over-sampling Technique ..... | 19 |
| 2.2.6.                            | Naïve Bayes .....                                | 20 |
| 2.2.7.                            | Support Vector Machine .....                     | 21 |
| 2.2.8.                            | Metrik Evaluasi .....                            | 21 |
| 2.2.9.                            | Confusion Matrix .....                           | 22 |
| 2.2.10.                           | Learning Curve.....                              | 23 |
| BAB III METODE PENELITIAN .....   |                                                  | 24 |
| 3.1                               | Objek Penelitian.....                            | 24 |
| 3.2                               | Alur Penelitian .....                            | 24 |
| 3.2.1                             | Persiapan Data .....                             | 25 |
| 3.2.2                             | Pra-Pemrosesan Teks .....                        | 26 |
| 3.2.3                             | Pembentukan Dataset dan Transformasi Fitur ..... | 28 |
| 3.2.4                             | Pemodelan.....                                   | 28 |
| 3.2.5                             | Evaluasi dan Optimasi Model.....                 | 29 |
| 3.3                               | Alat dan Bahan.....                              | 29 |
| 3.3.1                             | Deskripsi Dataset .....                          | 29 |
| 3.3.2                             | Perangkat Lunak .....                            | 30 |
| 3.3.3                             | Perangkat Keras .....                            | 31 |
| BAB IV HASIL DAN PEMBAHASAN ..... |                                                  | 32 |
| 4.1                               | Persiapan Data .....                             | 32 |

|         |                                                          |    |
|---------|----------------------------------------------------------|----|
| 4.1.1   | Pengumpulan Data .....                                   | 32 |
| 4.1.2   | Pelabelan .....                                          | 39 |
| 4.2     | Pra-Pemrosesan Teks .....                                | 42 |
| 4.2.1   | Normalisasi Teks .....                                   | 43 |
| 4.2.2   | Tokenisasi .....                                         | 44 |
| 4.2.3   | Stopword Removal .....                                   | 44 |
| 4.2.4   | Stemming .....                                           | 45 |
| 4.3     | Pembentukan Dataset dan Transformasi Fitur .....         | 46 |
| 4.3.1   | Split Data .....                                         | 46 |
| 4.3.2   | TF-IDF .....                                             | 47 |
| 4.3.3   | SMOTE (Synthetic Minority Over-sampling Technique) ..... | 52 |
| 4.4     | Pemodelan .....                                          | 54 |
| 4.4.1   | Pemodelan Multinomial Naïve Bayes .....                  | 54 |
| 4.4.2   | Pemodelan Support Vector Machine Linear .....            | 55 |
| 4.4.3   | Ringkasan Tahap Pemodelan .....                          | 55 |
| 4.5     | Evaluasi dan Optimasi Model .....                        | 55 |
| 4.5.1   | Evaluasi Model Baseline .....                            | 56 |
| 4.5.2   | Analisis Confusion Matrix Model Baseline .....           | 56 |
| 4.5.3   | Optimasi .....                                           | 59 |
| 4.5.3.1 | Optimasi Multinomial Naïve Bayes .....                   | 59 |
| 4.5.3.2 | Optimasi Support Vector Machine Linear .....             | 60 |
| 4.5.4   | Evaluasi Akhir Model .....                               | 61 |
| 4.5.4.1 | Perbandingan Metrik Evaluasi Setelah Optimasi .....      | 61 |
| 4.5.4.2 | Analisis Confusion Matrix Setelah Optimasi .....         | 62 |
| 4.5.4.3 | Analisis Learning Curve .....                            | 64 |

|                     |                                                        |    |
|---------------------|--------------------------------------------------------|----|
| 4.5.4.3.1           | Metrik Accuracy.....                                   | 64 |
| 4.5.4.3.2           | Metrik F1-macro .....                                  | 67 |
| 4.6                 | Analisis Perbandingan dengan Penelitian Terdahulu..... | 70 |
| BAB V PENUTUP ..... |                                                        | 72 |
| 5.1                 | Kesimpulan .....                                       | 72 |
| 5.2                 | Saran .....                                            | 73 |
| REFERENSI .....     |                                                        | 74 |
| LAMPIRAN.....       |                                                        | 77 |



## DAFTAR TABEL

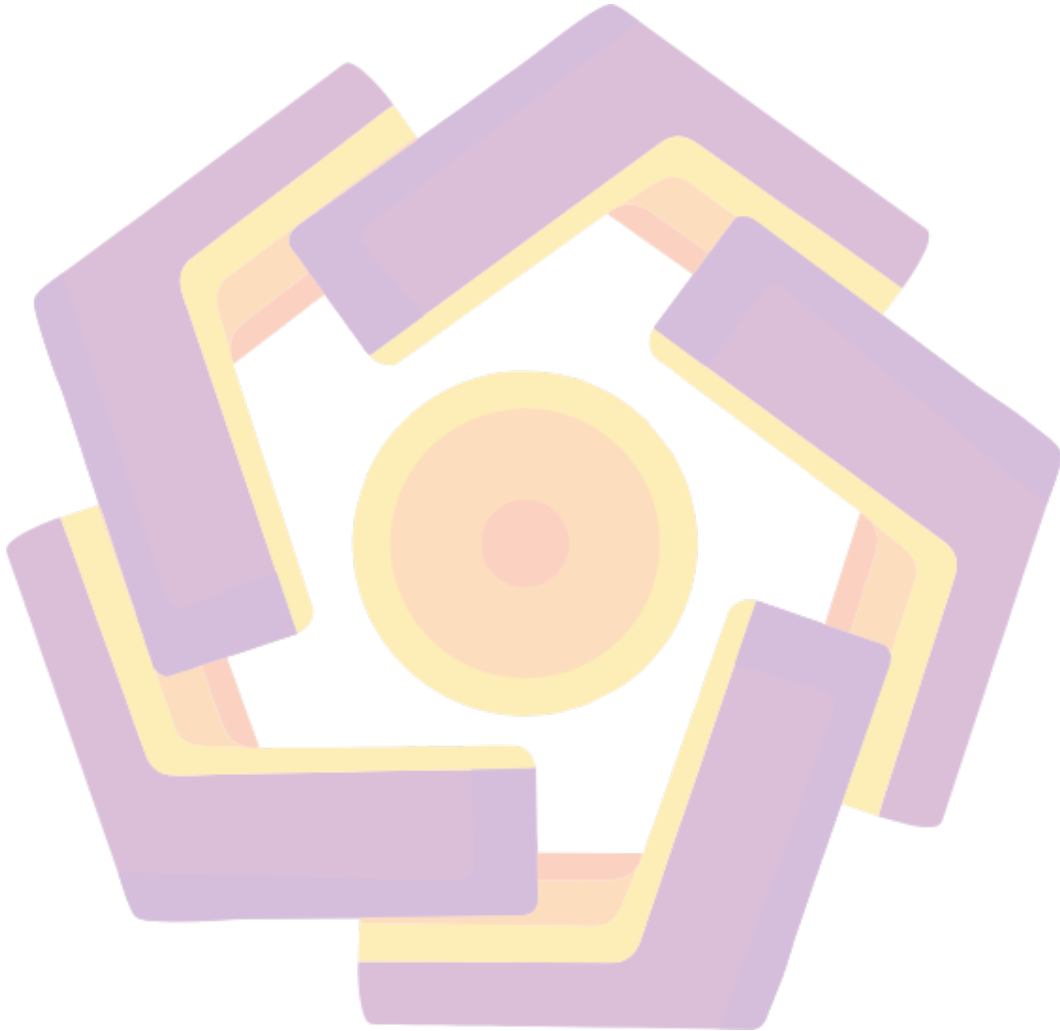
|                                                                                                       |    |
|-------------------------------------------------------------------------------------------------------|----|
| Tabel 2.1 Keaslian Penelitian .....                                                                   | 10 |
| Tabel 3.1 Perangkat Lunak .....                                                                       | 30 |
| Tabel 4.1 Cuplikan Tweet yang Dihapus karena Indikasi Buzzer .....                                    | 33 |
| Tabel 4.2 Contoh Konversi Tweet Berbahasa Inggris ke Bahasa Indonesia .....                           | 34 |
| Tabel 4.3 Cuplikan Dataset Hasil Pengumpulan Data Menggunakan Twitter Harvester setelah seleksi ..... | 35 |
| Tabel 4.4 Dataset Hasil Pengumpulan Data Menggunakan Grok AI .....                                    | 38 |
| Tabel 4.5 Aturan Klasifikasi Sentimen .....                                                           | 39 |
| Tabel 4.6 Cuplikan Dataset Hasil Pelabelan .....                                                      | 41 |
| Tabel 4.7 Hasil Normalisasi .....                                                                     | 43 |
| Tabel 4.8 Hasil Tokenisasi .....                                                                      | 44 |
| Tabel 4.9 Hasil Stopword Removal .....                                                                | 44 |
| Tabel 4.10 Hasil Stemming .....                                                                       | 45 |
| Tabel 4.11 Contoh Tweet Hasil Pra-Pemrosesan Teks .....                                               | 48 |
| Tabel 4.12 Pembobotan Term Frequency .....                                                            | 48 |
| Tabel 4.13 Hasil Perhitungan IDF .....                                                                | 50 |
| Tabel 4.14 Hasil Transformasi TF-IDF .....                                                            | 51 |
| Tabel 4.15 Baris Kode Penerapan SMOTE .....                                                           | 53 |
| Tabel 4.16 Baris Kode Pemodelan Multinomial Naive Bayes .....                                         | 54 |
| Tabel 4.17 Baris Kode Pemodelan SVM Linear .....                                                      | 55 |
| Tabel 4.18 Hasil Evaluasi Model Baseline .....                                                        | 56 |
| Tabel 4.19 Baris Kode Optimasi Multinomial Naive Bayes .....                                          | 59 |
| Tabel 4.20 Baris Kode Optimasi SVM Linear .....                                                       | 60 |
| Tabel 4.21 Metrik Evaluasi Setelah Optimasi .....                                                     | 61 |
| Tabel 4.22 Perbandingan Hasil Penelitian dengan Penelitian Terdahulu .....                            | 70 |

## DAFTAR GAMBAR

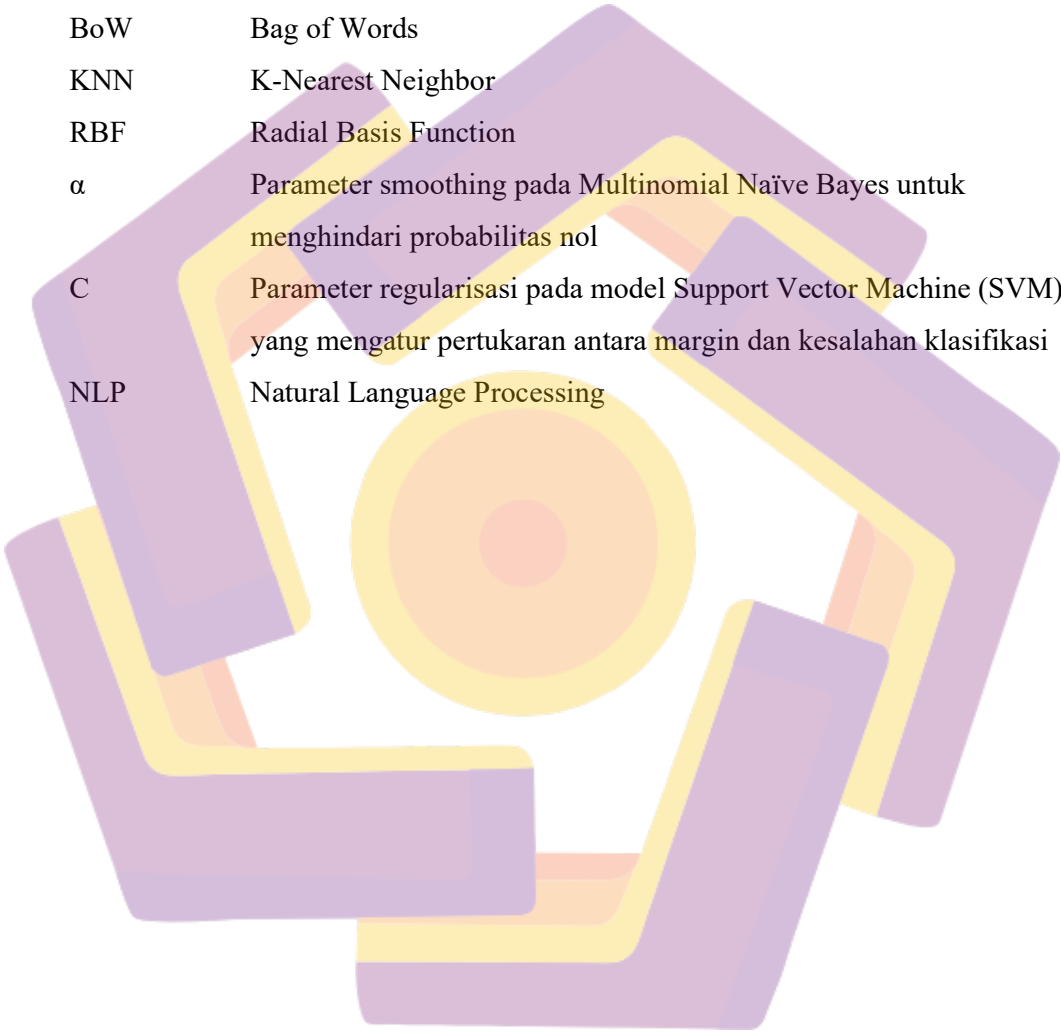
|                                                                        |    |
|------------------------------------------------------------------------|----|
| Gambar 2. 1. Distribusi Data Sebelum dan Sesudah Penerapan SMOTE ..... | 19 |
| Gambar 3.1. Alur Penelitian .....                                      | 25 |
| Gambar 4.1. Distribusi Kelas Sentimen Pada Dataset Awal .....          | 42 |
| Gambar 4.2. Pembagian Data Latih dan Uji .....                         | 47 |
| Gambar 4.3. Distribusi Sentimen Sebelum SMOTE pada Data Training ..... | 52 |
| Gambar 4.4. Distribusi Sentimen Setelah SMOTE pada Data Training ..... | 53 |
| Gambar 4.5. Confusion Matrix Naive Bayes Baseline .....                | 57 |
| Gambar 4.6. Confusion Matrix SVM Baseline .....                        | 58 |
| Gambar 4.7. Confusion Matrix Naive Bayes Tuned .....                   | 62 |
| Gambar 4.8. Confusion Matrix SVM Tuned .....                           | 63 |
| Gambar 4.9. Learning Curve - Accuracy Scoring Naive Bayes Tuned .....  | 65 |
| Gambar 4.10. Learning Curve - Accuracy Scoring SVM Tuned .....         | 66 |
| Gambar 4.11. Learning Curve - F1-macro Scoring NB Tuned .....          | 67 |
| Gambar 4.12. Learning Curve - F1-macro Scoring SVM Tuned .....         | 68 |

## DAFTAR LAMPIRAN

|                                                              |    |
|--------------------------------------------------------------|----|
| Lampiran. 1. Link Baris Kode Scraping Twitter Harvester..... | 77 |
| Lampiran. 2. Link Dataset Opini Publik Sosial Media X.....   | 77 |
| Lampiran. 3. Link Baris Kode Analisis Sentimen Publik.....   | 77 |

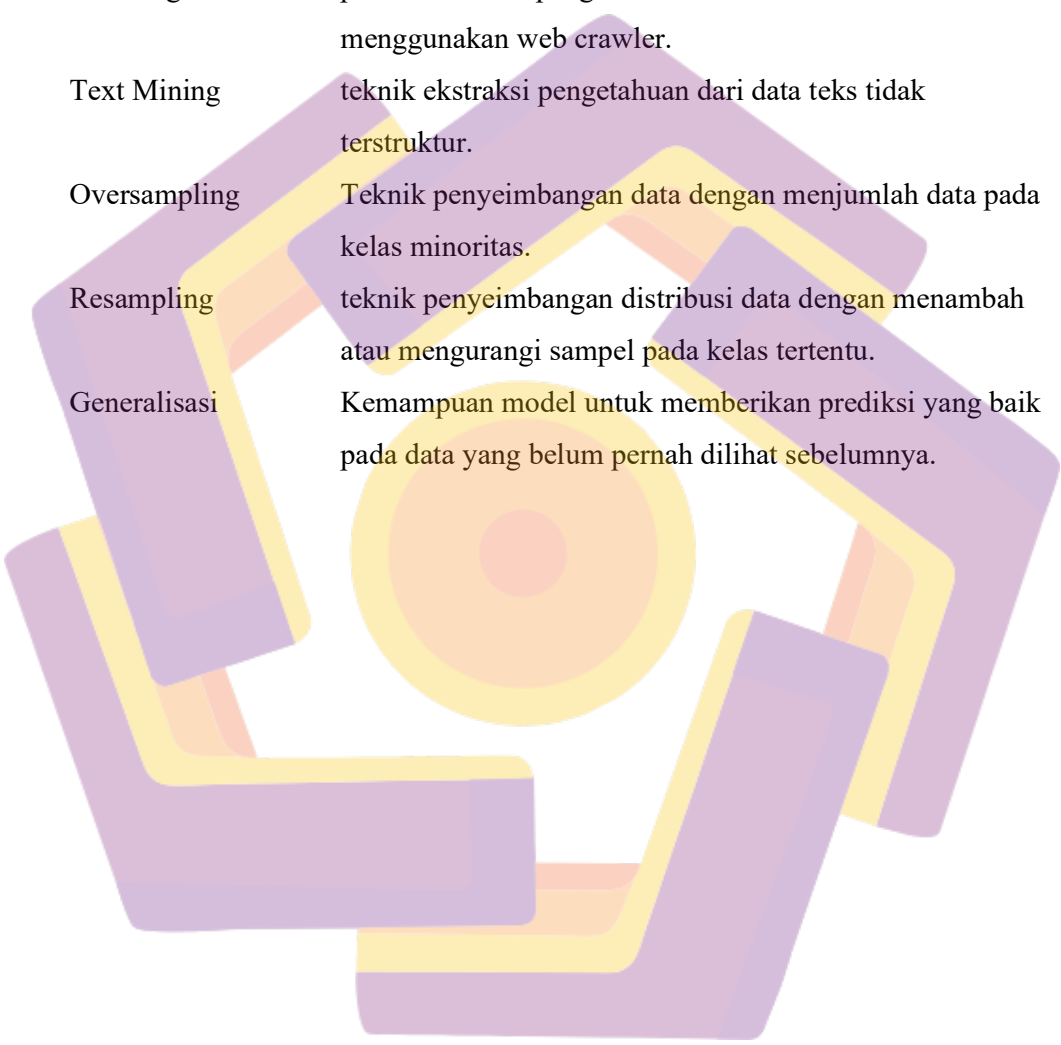


## DAFTAR LAMBANG DAN SINGKATAN



|          |                                                                                                                                 |
|----------|---------------------------------------------------------------------------------------------------------------------------------|
| SVM      | Support Vector Machines                                                                                                         |
| TF-IDF   | Term Frequency – Inverse Document Frequency                                                                                     |
| SMOTE    | Synthetic Minority Oversampling Technique                                                                                       |
| BoW      | Bag of Words                                                                                                                    |
| KNN      | K-Nearest Neighbor                                                                                                              |
| RBF      | Radial Basis Function                                                                                                           |
| $\alpha$ | Parameter smoothing pada Multinomial Naïve Bayes untuk menghindari probabilitas nol                                             |
| C        | Parameter regularisasi pada model Support Vector Machine (SVM) yang mengatur pertukaran antara margin dan kesalahan klasifikasi |
| NLP      | Natural Language Processing                                                                                                     |

## DAFTAR ISTILAH



|              |                                                                                                        |
|--------------|--------------------------------------------------------------------------------------------------------|
| Scraping     | proses pengambilan atau ekstraksi data dari situs web.<br>secara otomatis menggunakan script.          |
| Crawling     | proses otomatis pengambilan data dari situs web<br>menggunakan web crawler.                            |
| Text Mining  | teknik ekstraksi pengetahuan dari data teks tidak<br>terstruktur.                                      |
| Oversampling | Teknik penyeimbangan data dengan menjumlah data pada<br>kelas minoritas.                               |
| Resampling   | teknik penyeimbangan distribusi data dengan menambah<br>atau mengurangi sampel pada kelas tertentu.    |
| Generalisasi | Kemampuan model untuk memberikan prediksi yang baik<br>pada data yang belum pernah dilihat sebelumnya. |

## INTISARI

Perkembangan teknologi informasi telah mendorong media sosial menjadi ruang interaksi publik yang penting dalam menyampaikan opini dan tanggapan terhadap kebijakan pemerintah. Pada momen 100 hari kerja pemerintahan Prabowo Subianto, berbagai tanggapan masyarakat yang muncul di platform X (Twitter) membentuk kumpulan data tekstual yang mencerminkan kecenderungan sentimen publik. Kondisi tersebut menjadikan platform X sebagai sumber data yang relevan untuk dianalisis guna memahami persepsi masyarakat secara lebih luas dan objektif. Penelitian ini bertujuan untuk mengidentifikasi kecenderungan sentimen publik serta membandingkan kinerja dua algoritma klasifikasi teks, yaitu Naïve Bayes dan Support Vector Machine (SVM), dalam konteks data sosial-politik berbahasa Indonesia.

Data penelitian terdiri dari 1.369 tweet yang dikumpulkan melalui proses scraping. Data tersebut kemudian melalui tahapan pra-pemrosesan teks, pelabelan manual, pembobotan Term Frequency–Inverse Document Frequency (TF-IDF), serta penyeimbangan data menggunakan Synthetic Minority Oversampling Technique (SMOTE) yang diterapkan pada data latih. Model klasifikasi dilatih dan dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score, serta divisualisasikan melalui confusion matrix. Selain itu, analisis learning curve dilakukan untuk melihat pola pembelajaran masing-masing model.

Hasil penelitian menunjukkan bahwa sentimen publik terhadap 100 hari kerja pemerintahan Prabowo Subianto cenderung didominasi oleh sentimen positif dan netral. Dari sisi performa model, SVM Linear memperoleh accuracy sebesar 0,781 dan F1-score sebesar 0,78, lebih tinggi dibandingkan Naïve Bayes dengan accuracy sebesar 0,766 dan F1-score sebesar 0,76. Berdasarkan hasil tersebut, SVM Linear dapat disimpulkan sebagai model yang lebih optimal dalam analisis sentimen publik pada konteks sosial-politik berbahasa Indonesia. Temuan ini diharapkan dapat memberikan kontribusi teoretis dalam pengembangan analisis sentimen serta manfaat praktis bagi pemerintah, lembaga survei, dan pengembang sistem informasi.

**Kata kunci:** Analisis Sentimen, Naïve Bayes, Support Vector Machine, TF-IDF, SMOTE.

## **ABSTRACT**

*The advancement of information technology has positioned social media as a crucial space for public interaction, enabling the expression of opinions and responses to government policies. During the first 100 days of Prabowo Subianto's administration, various public reactions on platform X (Twitter) generated a textual dataset that reflects prevailing sentiment trends. This condition makes platform X a relevant source for analyzing public perceptions in a broader and more objective manner. This study aims to identify public sentiment tendencies and compare the performance of two text classification algorithms, namely Naïve Bayes and Support Vector Machine (SVM), within the context of Indonesian socio-political data.*

*The dataset consists of 1,369 tweets collected through scraping. The data underwent text preprocessing, manual labeling, Term Frequency-Inverse Document Frequency (TF-IDF) weighting, and balancing using the Synthetic Minority Oversampling Technique (SMOTE) applied to the training set. Classification models were trained and evaluated using accuracy, precision, recall, and F1-score metrics, with results visualized through confusion matrices. Additionally, learning curve analysis was conducted to observe the training patterns of each model.*

*The results indicate that public sentiment toward the first 100 days of Prabowo Subianto's administration was predominantly positive and neutral. In terms of model performance, Linear SVM achieved an accuracy of 0.781 and an F1-score of 0.78, outperforming Naïve Bayes, which obtained an accuracy of 0.766 and an F1-score of 0.76. Based on these results, Linear SVM can be considered the more optimal model for sentiment analysis in Indonesian socio-political contexts. These findings are expected to provide theoretical contributions to the development of sentiment analysis and practical benefits for government institutions, survey agencies, and information system developers.*

**Keyword:** Sentiment Analysis, Naïve Bayes, Support Vector Machine, TF-IDF, SMOTE.