

**STUDI KOMPARATIF ALGORITMA MACHINE  
LEARNING UNTUK KLASIFIKASI SENTIMEN  
TOKOH JOKOWI**

**SKRIPSI**

Diajukan Untuk Memenuhi Salah Satu Syarat  
Mencapai Gelar Sarjana Komputer  
Di Program Studi informatika  
Universitas Sahid Surakarta



**Disusun Oleh:**

**ILHAM YUSUF**

**NIM. 2021061024**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS SAINS, TEKNOLOGI DAN KESEHATAN  
UNIVERSITAS SAHID SURAKARTA**

**2025**

## LEMBAR PERSETUJUAN

### STUDI KOMPARATIF ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI SENTIMEN TOKOH JOKOWI

Disusun oleh:

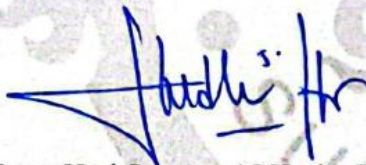
**ILHAM YUSUF**  
**NIM. 2021061024**

Skripsi ini telah disetujui untuk dipertahankan

Di hadapan dewan penguji

pada hari Jumat tanggal 18 Juli 2025

**Pembimbing I**



Firdhaus Hari Saputro Al Haris, S.T., M.Eng  
NIDN. 0614068201

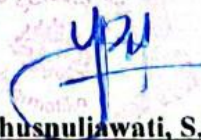
**Pembimbing II**



Hardika Khusnuliawati, S.Kom., M.kom  
NIDN. 0631089201

Mengetahui,

Ketua Program Studi



Hardika Khusnuliawati, S.Kom., M.Kom  
NIDN. 0631089201

## LEMBAR PENGESAHAN

### STUDI KOMPARATIF ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI SENTIMEN TOKOH JOKOWI

Disusun oleh:

ILHAM YUSUF

NIM. 2021061024

Skripsi ini telah diterima dan disahkan

Oleh dewan penguji Skripsi

Program Studi Informatika

Universitas Sahid Surakarta

Pada hari Jumat tanggal 18 Juli 2025

#### Dewan Penguji

1. Penguji 1 Firdhaus Hari Saputro Al Haris, S.T., M.Eng  
NIDN. 0614068201
2. Penguji 2 Hardika Khusnuliawati, S.Kom., M.kom  
NIDN. 0631089201
3. Penguji 3 Farid Fitriyadi, S.Kom., M.Kom  
NIDN. 0607058902



Mengetahui,

Ketua Program Studi,

Dekan,

  
Hardika Khusnuliawati, S.Kom., M.Kom.  
NIDN. 0631089201  
Apt. Ahwan, S.Farm., M.Sc.  
NIDN. 0626088401

**SURAT PERNYATAAN**  
**ORISINALITAS KARYA ILMIAH**

---

Saya mahasiswa Program Studi Informatika Fakultas Sains, Teknologi, Dan Kesehatan Universitas Sahid Surakarta yang bertanda tangan di bawah ini:

Nama : Ilham Yusuf

Nim : 2021061024

Menyatakan dengan sesungguhnya bahwa Tugas Akhir/ Skripsi:

Judul : STUDI KOMPARATIF ALGORITMA MACHINE LEARNING  
UNTUK KLASIFIKASI SENTIMEN TOKOH JOKOWI

Adalah benar-benar karya yang saya susun sendiri. Apabila dikemudian hari terbukti bahwa saya ternyata melakukan tindakan menyalin atau meniru tulisan orang lain seolah-olah hasil pemikiran saya sendiri, saya bersedia menerima sanksi sesuai peraturan yang telah berlaku di Universitas Sahid Surakarta termasuk pencabutan gelar sarjana yang telah saya peroleh.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya dan apabila dikemudian hari terbukti melakukan kebohongan maka saya akan menanggung konsekuensinya.

Surakarta, Juli 2025

Yang Menyatakan,



(Ilham Yusuf)

**PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH KEPENTINGAN AKADEMIS**

---

---

Sebagai sivitas akademika Universitas Sahid Surakarta, saya yang bertanda tangan di bawah ini:

Nama : Ilham Yusuf  
Nim : 2021061024  
Program Studi : Informatika  
Fakultas : SAINS, TEKNOLOGI, DAN KESEHATAN  
Email : Ilhamyusuf647@gmail.com  
Jenis Karya : SKRIPSI

Demikian perkembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Sahid Surakarta Hak Bebas Royalti Non Eksklusif (*Non-Exclusive Royalty Free Right*) atas skripsi saya yang berjudul:

**STUDI KOMPARATIF ALGORITMA MACHINE LEARNING UNTUK  
KLASIFIKASI SENTIMEN TOKOH JOKOWI**

Badan instrumen/desain/perangkat (jika ada). Berhak menyimpan, mengalihkan bentuk, digitalisasi, mengolah dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan karya ilmiah saya selama tetap mencantumkan nama saya sebagai penulis (*author*) dan pembimbing sebagai *cauthor* atau pencipta dan sebagai hak milik.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa paksaan dari pihak manapun.

Dibuat di: Surakarta

Pada Tanggal: Juli 2025

Yang membuat pernyataan,



(Ilham Yusuf)

## MOTTO

*You were born alone, you'll die alone.*

*But your mother's love will remain.*



## KATA PENGANTAR

Puji syukur ke hadirat Allah SWT, Tuhan Yang Maha Esa, atas limpahan rahmat, hidayah, dan karunia-Nya, penulis dapat menyelesaikan penulisan skripsi ini. Skripsi ini disusun sebagai salah satu syarat kelulusan Program Studi Informatika Universitas Sahid Surakarta. Penyusunan skripsi ini tidak lepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Sri Huning Anwariningsih, S.T., M.Kom. selaku Rektor Universitas Sahid Surakarta
2. Bapak Apt. Ahwan, S.Farm., M.Sc. selaku Dekan Fakultas Sains, Teknologi dan Kesehatan Universitas Sahid Surakarta.
3. Ibu Hardika Khusnuliawati, S.Kom., M.Kom. selaku Ketua Prodi Studi Informatika Universitas Sahid Surakarta dan selaku Dosen Pembimbing II.
4. Bapak Firdhaus Hari Saputro Al Haris, S.T., M.T. selaku Dosen Pembimbing I yang telah memberi arahan dan dukungan selama proses penyelesaian tugas akhir ini.
5. Bapak dan Ibu Dosen Program Studi Informatika Fakultas Sains, Teknologi dan Kesehatan Universitas Sahid Surakarta atas bimbingan dan ilmu yang diberikan selama proses perkuliahan.
6. Kedua orang tua yang telah mengusahakan semuanya.
7. Semua pihak yang tidak dapat disebutkan satu per satu namun telah memberikan kontribusi dalam penyelesaian skripsi ini.

Skripsi ini masih memiliki kekurangan dan diharapkan dapat memperoleh kritik serta saran yang membangun,

Surakarta, Juni 2025

Penulis

## DAFTAR ISI

|                                         |      |
|-----------------------------------------|------|
| LEMBAR PERSETUJUAN.....                 | ii   |
| LEMBAR PENGESAHAN .....                 | iii  |
| SURAT PERNYATAAN .....                  | iv   |
| ORISINALITAS KARYA ILMIAH.....          | iv   |
| PERNYATAAN PERSETUJUAN PUBLIKASI .....  | v    |
| KARYA ILMIAH KEPENTINGAN AKADEMIS ..... | v    |
| MOTTO.....                              | vi   |
| KATA PENGANTAR.....                     | vii  |
| DAFTAR ISI .....                        | viii |
| DAFTAR GAMBAR .....                     | xi   |
| DAFTAR TABEL.....                       | xiii |
| ABSTRAK.....                            | xiv  |
| ABSTRACT.....                           | xv   |
| BAB I PENDAHULUAN .....                 | 1    |
| 1.1 Latar belakang.....                 | 1    |
| 1.2 Rumusan Masalah .....               | 2    |
| 1.3 Batasan Masalah.....                | 2    |
| 1.4 Tujuan Dan Manfaat Penelitian.....  | 3    |
| 1.4.1 Tujuan Penelitian.....            | 3    |
| 1.4.2 Manfaat Penelitian .....          | 3    |
| 1.5. Metodologi Penelitian .....        | 3    |
| 1.6. Sistematika Penulisan .....        | 5    |
| BAB II TINJAUAN PUSTAKA .....           | 7    |
| 2.1 Penelitian Terdahulu.....           | 7    |
| 2.2 Kerangka Pemikiran.....             | 9    |
| 2.3 Landasan Teori .....                | 10   |
| 2.3.1 Analisis Sentimen.....            | 10   |
| 2.3.2 Media Sosial X (twitter).....     | 11   |
| 2.3.3 Tokoh Jokowi .....                | 13   |
| 2.3.4 <i>Machine Learning</i> .....     | 13   |

|                                           |                                                        |           |
|-------------------------------------------|--------------------------------------------------------|-----------|
| 2.3.5                                     | Metode SEMMA.....                                      | 14        |
| 2.3.6                                     | <i>Text Preprocessing</i> .....                        | 14        |
| 2.3.7                                     | <i>Support Vector Machine</i> .....                    | 15        |
| 2.3.8                                     | <i>Logistic Regression</i> .....                       | 17        |
| 2.3.9                                     | <i>Confusion Matrix</i> .....                          | 19        |
| 2.3.10                                    | <i>Lexicon Based</i> .....                             | 20        |
| 2.3.11                                    | <i>Grid Search</i> .....                               | 20        |
| 2.3.12                                    | <i>Google Colab Research</i> .....                     | 21        |
| <b>BAB III METODOLOGI PENELITIAN.....</b> |                                                        | <b>22</b> |
| 3.1                                       | Sumber Data.....                                       | 22        |
| 3.2                                       | Metode Pengumpulan Data .....                          | 22        |
| 3.3                                       | Metode SEMMA.....                                      | 22        |
| 3.4                                       | Pengumpulan Data ( <i>Sample</i> ).....                | 23        |
| 3.5                                       | Eksplorasi Data ( <i>Explore</i> ).....                | 23        |
| 3.6                                       | <i>Preprocessing (Modify)</i> .....                    | 25        |
| 3.7                                       | Pembuatan Dan Pengembangan Model ( <i>Model</i> )..... | 28        |
| 3.7.1                                     | Algoritma Support Vector Machine (SVM) .....           | 30        |
| 3.7.2                                     | Algoritma <i>Logistic Regression</i> .....             | 31        |
| 3.8                                       | Evaluasi ( <i>Asses</i> ) .....                        | 31        |
| <b>BAB IV HASIL DAN PEMBAHASAN.....</b>   |                                                        | <b>33</b> |
| 4.1                                       | Pengumpulan Data .....                                 | 33        |
| 4.2                                       | Deskripsi Data .....                                   | 33        |
| 4.3                                       | Eksplorasi Data .....                                  | 34        |
| 4.3.1                                     | Frekuensi Kata Sebelum <i>Preprocessing</i> .....      | 35        |
| 4.3.2                                     | <i>Wordcloud</i> Sebelum <i>Preprocessing</i> .....    | 35        |
| 4.4                                       | Hasil <i>Preprocessing</i> .....                       | 36        |
| 4.4.1                                     | Frekuensi Kata Setelah <i>Preprocessing</i> .....      | 42        |
| 4.4.2                                     | <i>Wordcloud</i> Setelah <i>Preprocessing</i> .....    | 42        |
| 4.5                                       | Hasil Pelatihan Dan Evaluasi Model.....                | 43        |
| 4.5.1                                     | Support Vector Machine.....                            | 43        |
| 4.6.2                                     | Logistic Regression.....                               | 45        |
| 4.7                                       | Perbandingan Kinerja Model .....                       | 47        |

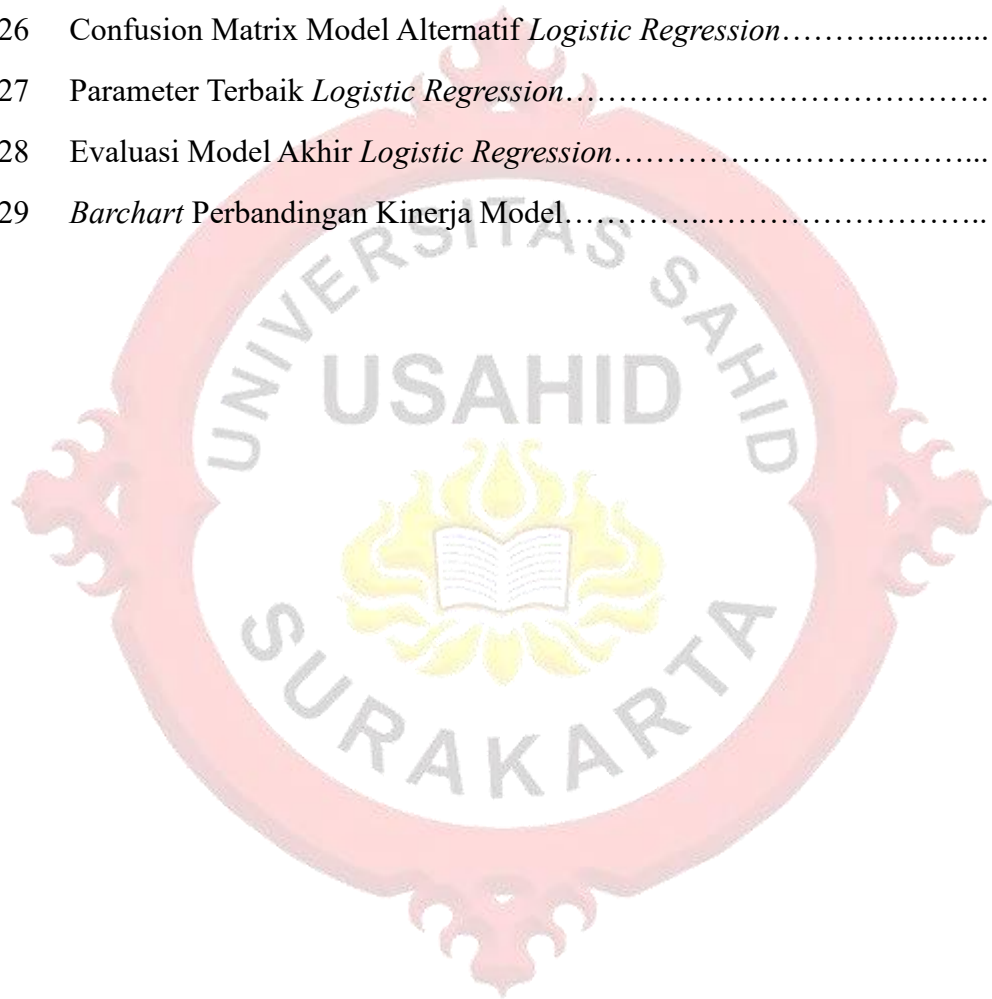
|                                 |                      |    |
|---------------------------------|----------------------|----|
| 4.8                             | Hasil Analisis ..... | 49 |
| BAB V KESIMPULAN DAN SARAN..... |                      | 51 |
| 5.1                             | Kesimpulan .....     | 51 |
| 5.2                             | Saran.....           | 52 |
| DAFTAR PUSTAKA .....            |                      | 53 |
| LAMPIRAN.....                   |                      | 57 |



## DAFTAR GAMBAR

|      |                                                      |    |
|------|------------------------------------------------------|----|
| 1.1  | <i>Flowchart</i> Metode Penelitian SEMMA.....        | 4  |
| 2.1  | Kerangka Pemikiran.....                              | 9  |
| 2.2  | Data Pengguna Media Sosial di Indonesia.....         | 12 |
| 2.3  | Ilustrasi Algoritma SVM.....                         | 17 |
| 2.4  | Ilustrasi Kurva Sigmoid Logistic Regression.....     | 19 |
| 3.1  | Kolom Dataset.....                                   | 24 |
| 3.2  | Pengembangan Model.....                              | 29 |
| 4.1  | <i>Crawling</i> .....                                | 33 |
| 4.2  | Struktur Dataset.....                                | 34 |
| 4.3  | Kolom yang Digunakan.....                            | 34 |
| 4.4  | Frekuensi Kata Sebelum <i>Preprocessing</i> .....    | 35 |
| 4.5  | <i>Worldcloud</i> Sebelum <i>Preprocessing</i> ..... | 36 |
| 4.6  | Sebelum <i>Cleaning</i> .....                        | 36 |
| 4.7  | Setelah <i>Cleaning</i> .....                        | 37 |
| 4.8  | Sebelum <i>Casefolding</i> .....                     | 37 |
| 4.9  | Setelah <i>Casefolding</i> .....                     | 37 |
| 4.10 | Sebelum Tokenisasi.....                              | 38 |
| 4.11 | Setelah Tokenisasi.....                              | 38 |
| 4.12 | Sebelum <i>Stopwords Removal</i> .....               | 39 |
| 4.13 | Setelah <i>Stopwords Removal</i> .....               | 39 |
| 4.14 | Sebelum <i>Stemming</i> .....                        | 39 |
| 4.15 | Setelah <i>Stemming</i> .....                        | 39 |
| 4.16 | Hasil Pelabelan.....                                 | 40 |
| 4.17 | <i>Barchart</i> Jumlah Sentimen.....                 | 41 |
| 4.18 | Fitur Vektorisasi.....                               | 41 |
| 4.19 | Frekuensi Kata Setelah <i>Preprocessing</i> .....    | 42 |

|      |                                                                           |    |
|------|---------------------------------------------------------------------------|----|
| 4.20 | Wordcloud Setelah <i>Preprocessing</i> .....                              | 43 |
| 4.21 | Evaluasi Model Alternatif SVM.....                                        | 44 |
| 4.22 | <i>Confusion Matrix</i> Model Alternatif SVM.....                         | 44 |
| 4.23 | Parameter Terbaik SVM.....                                                | 44 |
| 4.24 | Evaluasi Model Akhir SVM.....                                             | 45 |
| 4.25 | Evaluasi Model Alternatif <i>Logistic Regression</i> .....                | 46 |
| 4.26 | <i>Confusion Matrix</i> Model Alternatif <i>Logistic Regression</i> ..... | 46 |
| 4.27 | Parameter Terbaik <i>Logistic Regression</i> .....                        | 46 |
| 4.28 | Evaluasi Model Akhir <i>Logistic Regression</i> .....                     | 47 |
| 4.29 | <i>Barchart</i> Perbandingan Kinerja Model.....                           | 49 |



## DAFTAR TABEL

|     |                               |    |
|-----|-------------------------------|----|
| 2.1 | <i>Confusion Matrix</i> ..... | 20 |
| 4.1 | Ringkasan Evaluasi.....       | 48 |



## ABSTRAK

Media sosial telah menjadi *platform* penting dalam menyampaikan opini publik, termasuk terhadap tokoh politik seperti Joko Widodo (Jokowi). Opini tersebut dapat dianalisis melalui pendekatan analisis sentimen untuk memahami persepsi masyarakat secara luas. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma *machine learning*, yaitu *Support Vector Machine* (SVM) dan *Logistic Regression* dalam mengklasifikasikan sentimen publik terhadap Jokowi pada *platform* media sosial X (Twitter). Penelitian ini menggunakan metode SEMMA (*Sample, Explore, Modify, Model, Assess*) dalam proses analisis, dengan data yang dikumpulkan melalui teknik *crawling* sebanyak 1.500 tweet. Data diproses melalui tahapan pembersihan teks, pelabelan sentimen secara *lexicon-based*, vektorisasi TF-IDF, serta pelatihan dan evaluasi model menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil evaluasi menunjukkan bahwa algoritma *Logistic Regression* memberikan kinerja lebih baik dibandingkan SVM, dengan akurasi mencapai 96%, sedangkan SVM mencapai 84%. Dengan demikian, *Logistic Regression* terbukti lebih akurat dan stabil dalam mengklasifikasikan sentimen terhadap tokoh publik dalam konteks ini.

**Kata Kunci:** analisis sentimen, jokowi, klasifikasi, machine learning, sosial media.

## ABSTRACT

Social media has become a crucial platform for conveying public opinion, including among political figures such as Joko Widodo (Jokowi). This opinion can be analyzed through a sentiment analysis approach to gain insight into the public perception. This study aims to compare the performance of two machine learning algorithms, namely Support Vector Machine (SVM) and Logistic Regression in classifying public sentiment towards Jokowi on the social media platform X (Twitter). This study used the SEMMA (Sample, Explore, Modify, Model, Assess) method in the analysis process. Data were collected through a crawling technique of 1,500 tweets. The data was processed through the stages of text cleaning, lexicon-based sentiment labeling, TF-IDF vectorization, and training. Model evaluation used accuracy, precision, recall, and F1-score metrics. The evaluation results show that the Logistic Regression algorithm performed better than SVM, with an accuracy of 96%, while SVM achieved an accuracy of 84%. Thus, Logistic Regression proved to be more accurate and stable in classifying sentiment towards public figures in this context.

**Keywords:** Sentiment Analysis, Jokowi, Classification, Machine Learning, Social Media.

