

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KENAIKAN  
PPN MENJADI 12% DI INDONESIA PADA MEDIA SOSIAL X  
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE***

**SKRIPSI**

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



disusun oleh

**OCTAFIANA HANANI FITYATI SYIFA**

**21.12.2158**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2025**

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KENAIKAN  
PPN MENJADI 12% DI INDONESIA PADA MEDIA SOSIAL X  
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE***

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi S1 Sistem Informasi



disusun oleh

**OCTAFIANA HANANI FITYATI SYIFA**

**21.12.2158**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2025**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KENAIKAN  
PPN MENJADI 12% DI INDONESIA PADA MEDIA SOSIAL X  
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE***

yang disusun dan diajukan oleh

**Octafiana Hanani Fityati Syifa**

**21.12.2158**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 18 Februari 2025

Dosen Pembimbing,

  
**Hanif Al Fatta, S.Kom., M.Kom., Ph.D**

**NIK. 190302096**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**ANALISIS SENTIMEN MASYARAKAT TERHADAP KENAIKAN**  
**PPN MENJADI 12% DI INDONESIA PADA MEDIA SOSIAL X**  
**MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE***

yang disusun dan diajukan oleh

**Octafiana Hanani Fityati Syifa**

**21.12.2158**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 18 Februari 2025

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

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



**Atik Nurmasani, S.Kom., M.Kom**  
**NIK. 190302354**



**Hanif Al Fatta, S.Kom., M.Kom., Ph.D**  
**NIK. 190302096**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 18 Februari 2025

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom., Ph.D**  
**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Octafiana Hanani Fityati Syifa  
NIM : 21.12.2158

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Masyarakat Terhadap Kenaikan Ppn Menjadi 12% Di Indonesia Pada Media Sosial X Menggunakan Algoritma *Support Vector Machine***

Dosen Pembimbing : Hanif Al Fatta, S.Kom., 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, 18 Februari 2025

Yang Menyatakan,



Octafiana Hanani Fityati Syifa

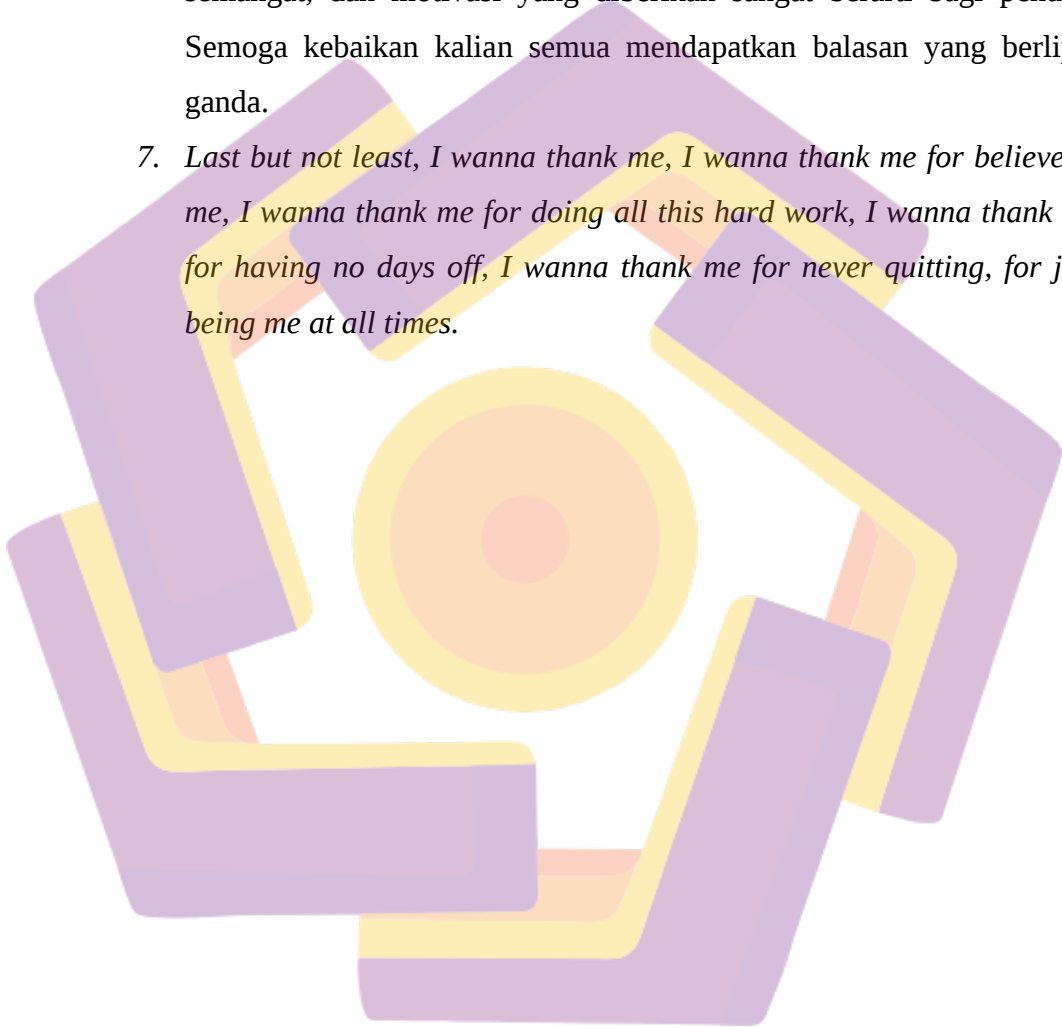
## HALAMAN PERSEMBAHAN

Puji syukur dihaturkan kehadiran Allah SWT atas segala limpahan karunia-Nya, sehingga penulisan ini dapat diselesaikan dengan baik. Dengan rasa bangga, karya ini penulis persembahkan kepada:

1. Pintu surgaku, Ibu Sulaifah, Beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan, namun beliau bekerja keras serta mendidik, memberi motivasi, memberi dukungan sehingga penulis mampu menyelesaikan studi hingga sarjana.
2. Cinta pertama dan panutanku, bapak Slamet Hanan Subkhi. Beliau berperan sangat penting dalam proses menyelesaikan program studi penulis, beliau juga tidak sempat merasakan pendidikan pada bangku perkuliahan, namun beliau tidak henti memberi semangat serta doa yang selalu mengiringi langkah penulis sehingga penulis bisa menyelesaikan program studi sampai selesai.
3. Sosok ayah yang penuh kasih, Bapak Imam Suyuti. Beliau selalu memberikan dukungan, motivasi, serta doa yang tak pernah putus dalam setiap langkah perjalanan studi ini. Kehadiran dan bimbingannya menjadi salah satu kekuatan besar yang mengantarkan penulis hingga menyelesaikan pendidikan ini.
4. Untuk adik tercinta dan keluarga besar yang selalu hadir dengan doa, dukungan, serta kehangatan. Terima kasih atas kasih sayang dan semangat yang tak pernah putus. Kehadiran kalian menjadi sumber kebahagiaan dan motivasi dalam menyelesaikan perjalanan ini. Semoga kebersamaan dan keberkahan selalu menyertai kita semua.
5. Sahabat-sahabat terbaik yang selalu ada di setiap langkah perjuangan ini, Echa Amelya, Triantaka Jagad, Desy Ikasari, dan Kurnia Amanatul. Terima kasih atas segala dukungan, kebersamaan, dan tawa yang mewarnai perjalanan kuliah ini. Kalian bukan hanya teman, tetapi keluarga yang selalu memberikan semangat, menemani di saat sulit, dan

menghadirkan kebahagiaan di setiap momen. Perjalanan ini tak akan sama tanpa kalian.

6. Terima kasih kepada semua pihak yang tidak bisa penulis sebutkan satu per satu atas segala doa, dukungan, dan bantuan yang telah diberikan dalam proses penyelesaian skripsi ini. Setiap bentuk kebaikan, semangat, dan motivasi yang diberikan sangat berarti bagi penulis. Semoga kebaikan kalian semua mendapatkan balasan yang berlipat ganda.
7. *Last but not least, I wanna thank me, I wanna thank me for believe in me, I wanna thank me for doing all this hard work, I wanna thank me for having no days off, I wanna thank me for never quitting, for just being me at all times.*



## KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir ini dengan lancar. Tugas akhir ini disusun untuk memenuhi salah satu syarat kelulusan Program Sarjana pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selain itu penulis dengan segala kerendahan hati ingin menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa memberikan dukungan dan bantuan untuk menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D selaku Dekan Program Fakultas Ilmu Komputer serta dosen pembimbing yang memberikan arahan, saran, dan motivasi terhadap penulis
3. Bapak Anggit Dwi Hartanto, S.Kom., M.Kom selaku Ketua Program Studi Sistem Informasi
4. Orang tua, adik, keluarga besar, dan teman-teman tercinta yang memberikan semangat dan doa kepada penulis.

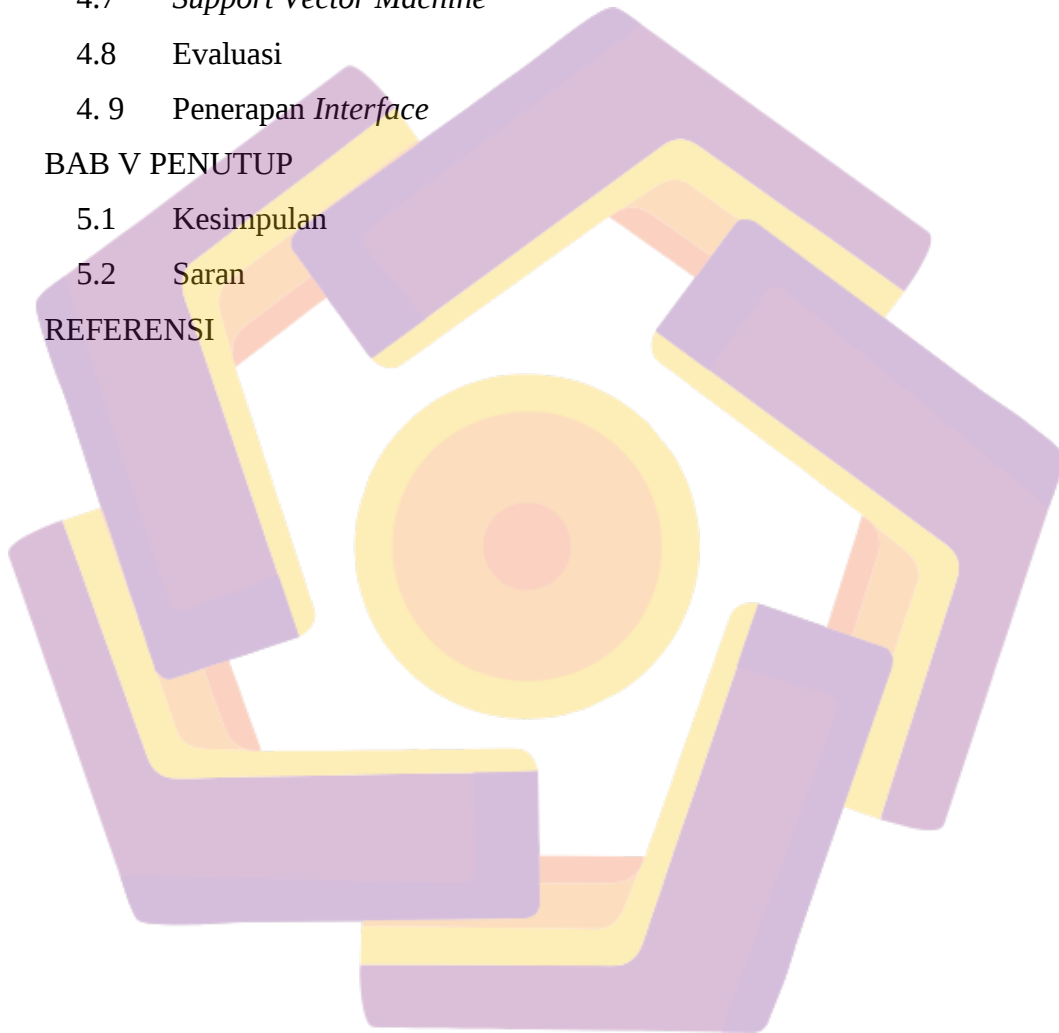
Yogyakarta, 18 Februari 2025

Penulis

## DAFTAR ISI

|                                     |      |
|-------------------------------------|------|
| HALAMAN JUDUL                       | i    |
| HALAMAN PERSETUJUAN                 | ii   |
| HALAMAN PENGESAHAN                  | iii  |
| HALAMAN PERNYATAAN KEASLIAN SKRIPSI | iv   |
| HALAMAN PERSEMBAHAN                 | v    |
| KATA PENGANTAR                      | vii  |
| DAFTAR ISI                          | viii |
| DAFTAR TABEL                        | x    |
| DAFTAR GAMBAR                       | xi   |
| DAFTAR LAMBANG DAN SINGKATAN        | xiii |
| DAFTAR ISTILAH                      | xiv  |
| INTISARI                            | xv   |
| ABSTRACT                            | xvi  |
| BAB I PENDAHULUAN                   | 1    |
| 1.1 Latar Belakang                  | 1    |
| 1.2 Rumusan Masalah                 | 2    |
| 1.3 Batasan Masalah                 | 2    |
| 1.4 Tujuan Penelitian               | 3    |
| 1.5 Manfaat Penelitian              | 3    |
| 1.6 Sistematika Penulisan           | 3    |
| BAB II TINJAUAN PUSTAKA             | 5    |
| 2.1 Studi Literatur                 | 5    |
| 2.2 Dasar Teori                     | 15   |
| BAB III METODE PENELITIAN           | 23   |
| 3.1 Objek Penelitian                | 23   |
| 3.2 Alur Penelitian                 | 23   |
| 3.3 Alat dan Bahan                  | 28   |
| BAB IV HASIL DAN PEMBAHASAN         | 29   |
| 4.1 <i>Crawling</i> Data            | 29   |

|               |                                                           |    |
|---------------|-----------------------------------------------------------|----|
| 4.2           | <i>Labeling</i>                                           | 30 |
| 4.3           | <i>Preprocessing</i>                                      | 32 |
| 4.4           | Pembobotan TF-IDF                                         | 42 |
| 4.5           | Pembagian Data Latih dan Data Uji                         | 43 |
| 4.6           | <i>Synthetic Minority Over-sampling Technique (SMOTE)</i> | 44 |
| 4.7           | <i>Support Vector Machine</i>                             | 45 |
| 4.8           | Evaluasi                                                  | 46 |
| 4.9           | Penerapan <i>Interface</i>                                | 56 |
| BAB V PENUTUP |                                                           | 64 |
| 5.1           | Kesimpulan                                                | 64 |
| 5.2           | Saran                                                     | 64 |
| REFERENSI     |                                                           | 66 |



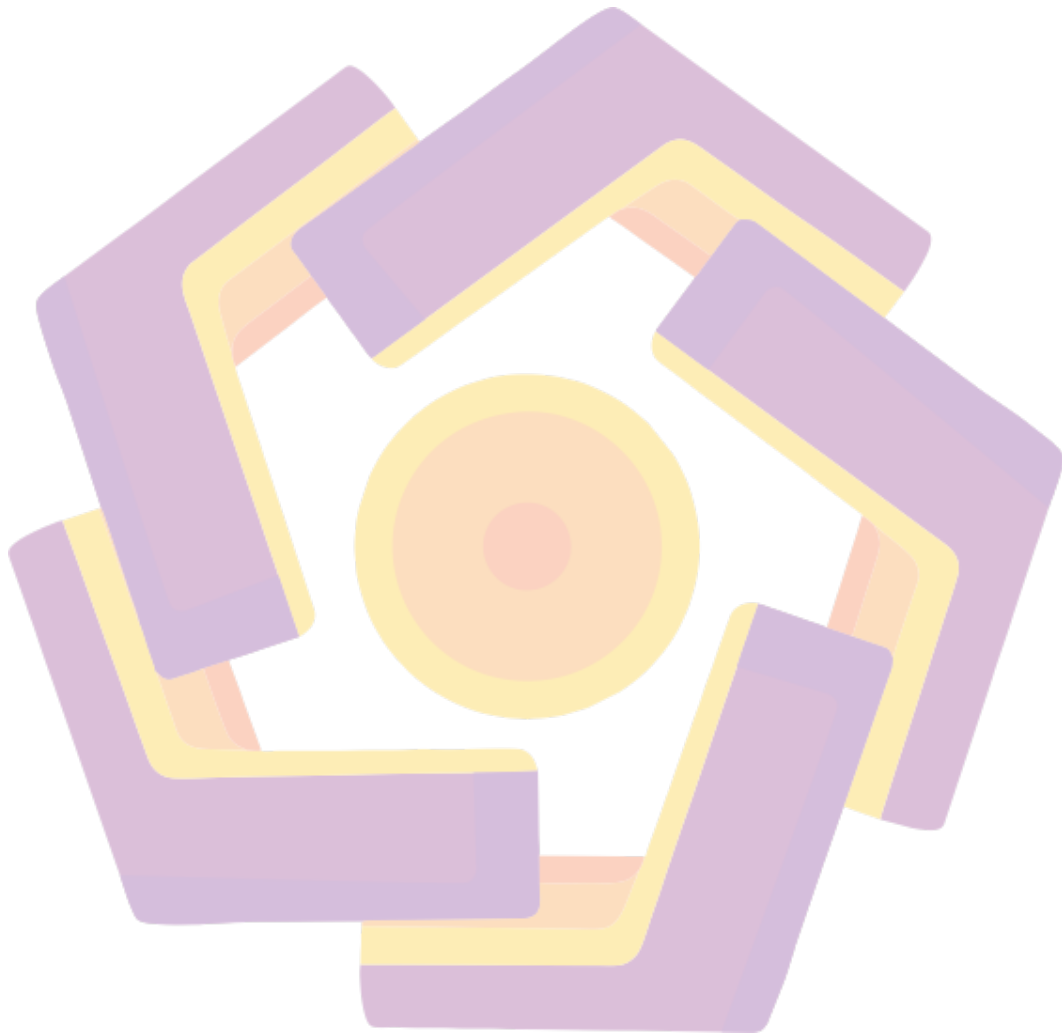
## DAFTAR TABEL

|                                                           |    |
|-----------------------------------------------------------|----|
| Tabel 2. 1 Keaslian Penelitian                            | 9  |
| Tabel 2. 2 Tabel <i>Confusion Matrix</i>                  | 20 |
| Tabel 4.1 Proses <i>Cleaning</i>                          | 34 |
| Tabel 4.2 Proses <i>Case Folding</i>                      | 35 |
| Tabel 4.3 Proses Normalisasi Kata                         | 36 |
| Tabel 4.4 Proses <i>Tokenizing</i>                        | 37 |
| Tabel 4. 5 Proses <i>Stopword Removal</i>                 | 39 |
| Tabel 4. 6 Proses <i>Stemming</i>                         | 40 |
| Tabel 4.7 Pembobotan TF-IDF                               | 42 |
| Tabel 4.8 Pembagian Data Latih dan Data Uji               | 44 |
| Tabel 4.9 SMOTE                                           | 45 |
| Tabel 4.10 Tingkat Performa <i>Support Vector Machine</i> | 55 |

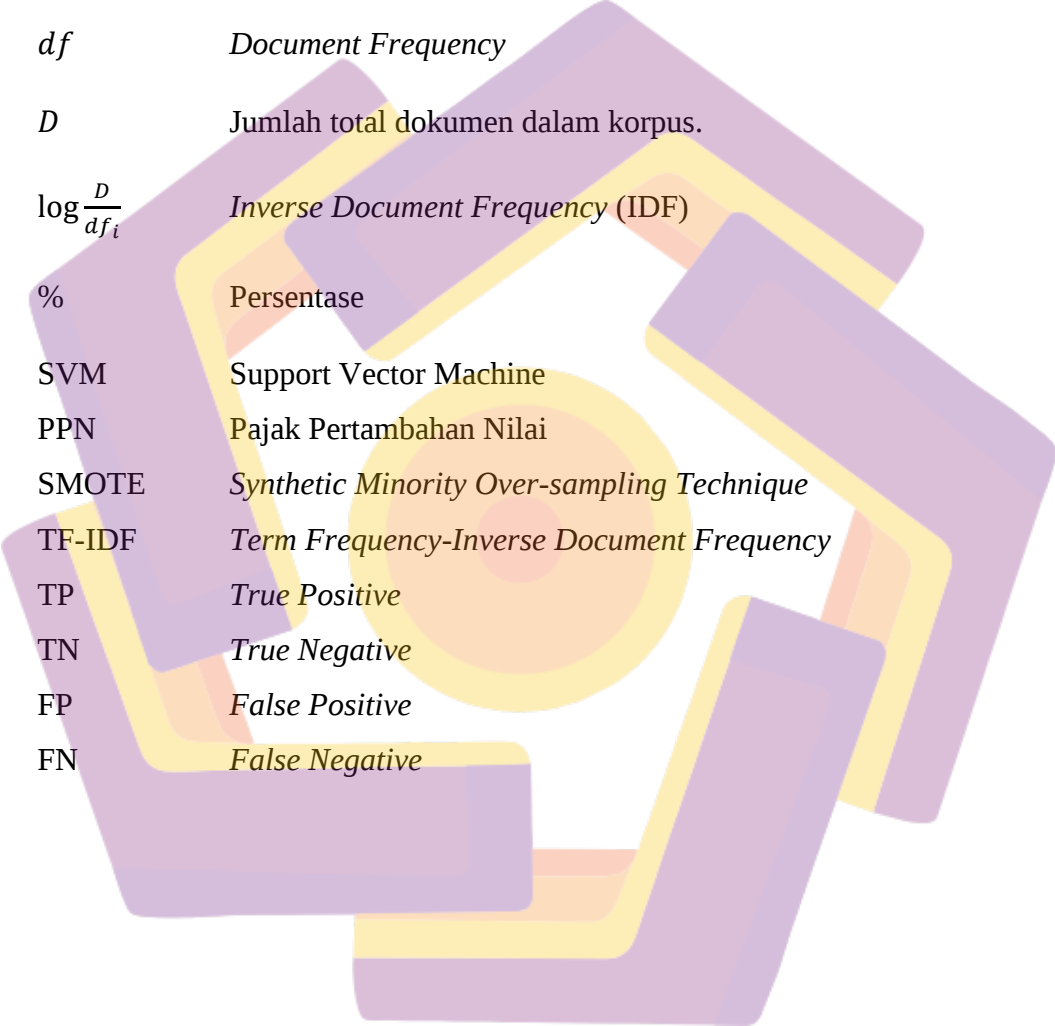
## DAFTAR GAMBAR

|                                                            |    |
|------------------------------------------------------------|----|
| Gambar 3.1 Alur Penelitian Sumber referensi[8].            | 24 |
| Gambar 4.1 Skrip <i>Crawling</i>                           | 29 |
| Gambar 4.2 Hasil <i>Crawling</i> Data                      | 30 |
| Gambar 4.3 Hasil <i>Labeling</i>                           | 31 |
| Gambar 4.4 Diagram Pie                                     | 32 |
| Gambar 4.5 Skrip <i>Preprocessing</i>                      | 33 |
| Gambar 4. 6 Skrip <i>Cleaning</i>                          | 33 |
| Gambar 4.7 Skrip <i>Case Folding</i>                       | 35 |
| Gambar 4.8 Skrip Normalisasi Kata                          | 36 |
| Gambar 4.9 Skrip <i>Tokenizing</i>                         | 37 |
| Gambar 4.10 Skrip <i>Stopword Removal</i>                  | 38 |
| Gambar 4.11 Skrip <i>Stemming</i>                          | 39 |
| Gambar 4.12 <i>Wordcloud</i> Sentimen Positif              | 41 |
| Gambar 4.13 <i>Wordcloud</i> Sentimen Negatif              | 41 |
| Gambar 4.14 Skrip TF-IDF                                   | 42 |
| Gambar 4. 15 Skrip Pembagian data                          | 43 |
| Gambar 4.16 Skrip SMOTE                                    | 44 |
| Gambar 4.17 Skrip SVM                                      | 45 |
| Gambar 4. 18 Skrip Evaluasi Model                          | 46 |
| Gambar 4.19 <i>Heatmap Confusion Matrix</i> Skenario Satu  | 47 |
| Gambar 4.20 <i>Heatmap Confusion Matrix</i> Skenario Dua   | 49 |
| Gambar 4.21 <i>Heatmap Confusion Matrix</i> Skenario Tiga  | 51 |
| Gambar 4.22 <i>Heatmap Confusion Matrix</i> Skenario Empat | 53 |
| Gambar 4.23 Hasil Evaluasi Pembagian Data 80:20            | 55 |
| Gambar 4.24 <i>Interface Dashboard</i>                     | 56 |
| Gambar 4. 25 Skrip <i>Import utils.py</i>                  | 57 |
| Gambar 4. 26 Skrip <i>Import app.py</i>                    | 58 |
| Gambar 4. 27 Skrip Menampilkan Dashboard app.py            | 59 |
| Gambar 4.28 <i>Interface Model Evaluasi</i>                | 59 |

|                                                                        |    |
|------------------------------------------------------------------------|----|
| Gambar 4. 29 Skrip app.py Memanggil Fungsi <i>Train Model</i> utils.py | 60 |
| Gambar 4. 30 Skrip Train Model utils.py                                | 61 |
| Gambar 4.31 <i>Interface</i> Dataset                                   | 62 |
| Gambar 4. 32 Skrip Dataset app.py                                      | 62 |
| Gambar 4. 33 Skrip Dataset utils.py                                    | 63 |

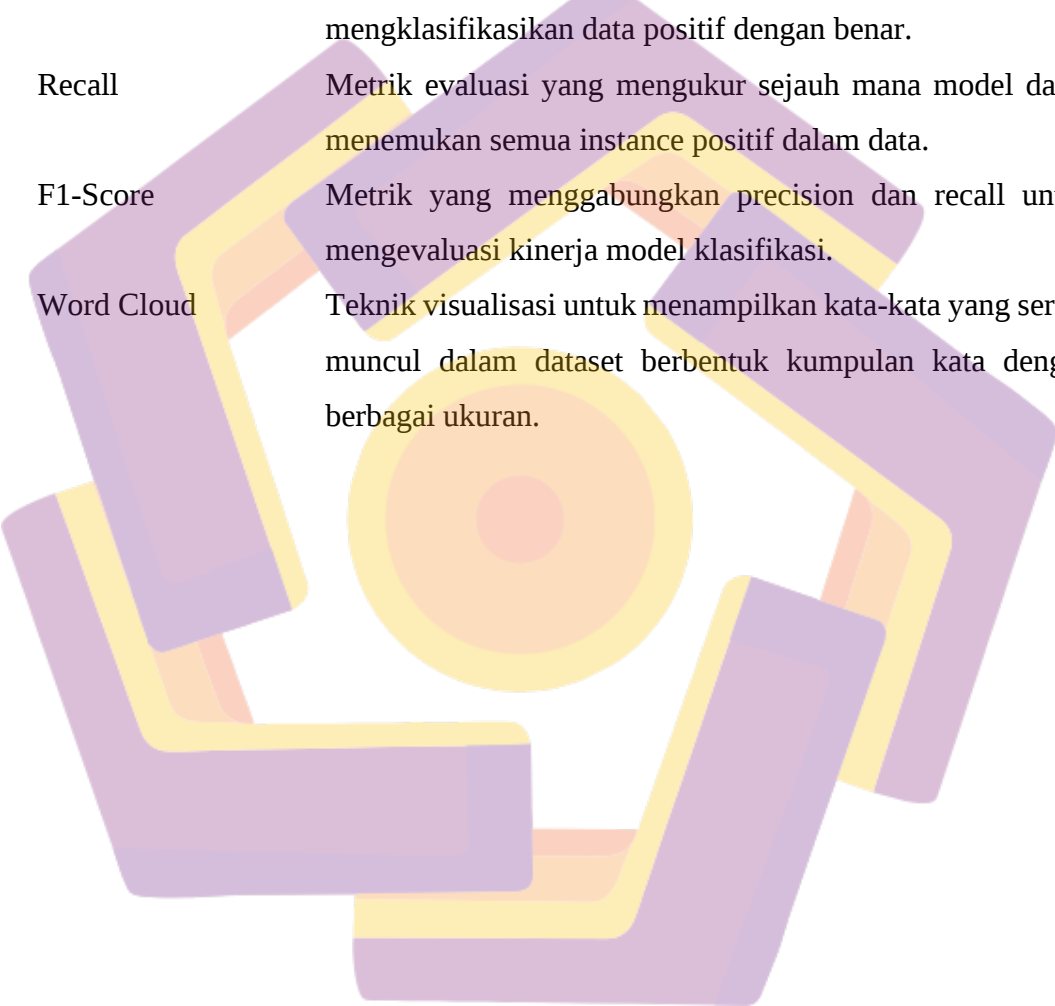


## DAFTAR LAMBANG DAN SINGKATAN



|                       |                                                         |
|-----------------------|---------------------------------------------------------|
| $\Omega$              | Tahanan Listrik                                         |
| $W_{ij}$              | Bobot TF-IDF untuk kata ke- $i$ dalam dokumen ke- $j$ . |
| $tf$                  | <i>Term Frequency</i>                                   |
| $df$                  | <i>Document Frequency</i>                               |
| $D$                   | Jumlah total dokumen dalam korpus.                      |
| $\log \frac{D}{df_i}$ | <i>Inverse Document Frequency (IDF)</i>                 |
| %                     | Persentase                                              |
| SVM                   | Support Vector Machine                                  |
| PPN                   | Pajak Pertambahan Nilai                                 |
| SMOTE                 | <i>Synthetic Minority Over-sampling Technique</i>       |
| TF-IDF                | <i>Term Frequency-Inverse Document Frequency</i>        |
| TP                    | <i>True Positive</i>                                    |
| TN                    | <i>True Negative</i>                                    |
| FP                    | <i>False Positive</i>                                   |
| FN                    | <i>False Negative</i>                                   |

## DAFTAR ISTILAH



|            |                                                                                                                                                                 |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Accuracy   | Metrik evaluasi dalam machine learning yang digunakan untuk mengukur seberapa banyak prediksi model yang benar dibandingkan dengan keseluruhan jumlah prediksi. |
| Precision  | Metrik evaluasi yang mengukur ketepatan model dalam mengklasifikasikan data positif dengan benar.                                                               |
| Recall     | Metrik evaluasi yang mengukur sejauh mana model dapat menemukan semua instance positif dalam data.                                                              |
| F1-Score   | Metrik yang menggabungkan precision dan recall untuk mengevaluasi kinerja model klasifikasi.                                                                    |
| Word Cloud | Teknik visualisasi untuk menampilkan kata-kata yang sering muncul dalam dataset berbentuk kumpulan kata dengan berbagai ukuran.                                 |

## INTISARI

Kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12% di Indonesia pada tahun 2025 menimbulkan beragam reaksi di masyarakat. Media sosial X (Twitter) menjadi platform utama bagi masyarakat untuk menyampaikan pendapat mereka, dengan mayoritas opini bersifat negatif. Analisis sentimen diperlukan untuk memahami persepsi publik guna memberikan wawasan bagi pengambil kebijakan dalam merancang strategi komunikasi yang lebih efektif.

Penelitian ini menggunakan metode analisis sentimen dengan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan sentimen masyarakat menjadi dua kategori, yaitu positif dan negatif. Data dikumpulkan melalui teknik *crawling* dari media sosial X dalam rentang waktu 1 Oktober 2024 hingga 20 Januari 2025. Tahapan *preprocessing* mencakup *cleaning*, *case folding*, normalisasi kata, *tokenizing*, *stopword removal*, dan *stemming*. Data diberi bobot menggunakan TF-IDF dan dilakukan *balancing* dengan *Synthetic Minority Over-sampling Technique* (SMOTE). Model SVM digunakan untuk klasifikasi, yang kemudian dievaluasi menggunakan akurasi, presisi, *recall*, dan F1-score.

Hasil penelitian menunjukkan bahwa algoritma SVM dengan skenario terbaik (pembagian data 80:20) mencapai akurasi 89,78%, dengan presisi 74% untuk sentimen positif dan 92% untuk sentimen negatif, *recall* 57% untuk sentimen positif dan 96% untuk sentimen negatif, serta F1-score 64% untuk sentimen positif dan 94% untuk sentimen negatif. Meskipun akurasi tinggi, model lebih efektif dalam memprediksi sentimen negatif. Penelitian ini diharapkan dapat memberikan manfaat bagi pemerintah dalam memahami persepsi masyarakat serta menjadi referensi bagi penelitian lebih lanjut di bidang analisis sentimen.

**Kata kunci:** Analisis Sentimen, Pajak Pertambahan Nilai, Media Sosial X, Support Vector Machine, SMOTE.

## ABSTRACT

*The increase in Value Added Tax (VAT) to 12% in Indonesia in 2025 has sparked various reactions among the public. Social media platform X (Twitter) has become the primary platform for people to express their opinions, with the majority of sentiments being negative. Sentiment analysis is necessary to understand public perception and provide insights for policymakers in designing more effective communication strategies.*

*This research employs sentiment analysis using the Support Vector Machine (SVM) algorithm to classify public sentiment into two categories: positive and negative. Data was collected through crawling techniques from social media platform X within the timeframe of October 1, 2024, to January 20, 2025. The preprocessing stages include cleaning, case folding, word normalization, tokenizing, stopword removal, and stemming. The data was weighted using TF-IDF and balanced using the Synthetic Minority Over-sampling Technique (SMOTE). The SVM model was used for classification, which was then evaluated using accuracy, precision, recall, and F1-score.*

*The results of the study show that the SVM algorithm with the best scenario (80:20 data split) achieved an accuracy of 89.78%, with a precision of 74% for positive sentiment and 92% for negative sentiment, recall of 57% for positive sentiment and 96% for negative sentiment, and an F1-score of 64% for positive sentiment and 94% for negative sentiment. Although the accuracy is high, the model is more effective in predicting negative sentiment. This research is expected to provide benefits for the government in understanding public perception and serve as a reference for further research in the field of sentiment analysis.*

**Keyword:** Sentiment Analysis, Value Added Tax, Social Media X, Support Vector Machine, SMOTE.