

**PERBANDINGAN ALGORITMA NAIVE BAYES DAN SVM
DALAM ANALISIS SENTIMEN TWITTER TENTANG
KENAIKAN PPN 12%**

SKRIPSI

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



disusun oleh

ADITYA ARDY PRATAMA

21.11.3986

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**PERBANDINGAN ALGORITMA NAIVE BAYES DAN SVM
DALAM ANALISIS SENTIMEN TWITTER TENTANG
KENAIKAN PPN 12%**

SKRIPSI

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



disusun oleh

ADITYA ARDY PRATAMA

21.11.3986

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN ALGORITMA NAIVE BAYES DAN SVM
DALAM ANALISIS SENTIMEN TWITTER TENTANG
KENAIKAN PPN 12%**

yang disusun dan diajukan oleh

Aditya Ardy Pratama

21.11.3986

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Agustus 2025

Dosen Pembimbing,



Dwi Nurani, M. Kom

NIK. 190302236

HALAMAN PENGESAHAN

SKRIPSI

**PERBANDINGAN ALGORITMA NAIVE BAYES DAN SVM
DALAM ANALISIS SENTIMEN TWITTER TENTANG
KENAIKAN PPN 12%**

yang disusun dan diajukan oleh

Aditya Ardy Pratama

21.11.3986

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Agustus 2025

Susunan Dewan Penguji

Nama Penguji

Rizqi Sukma Kharisma, S.Kom., M.Kom.
NIK. 190302215

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Dwi Nurani, S.Kom., M.Kom.
NIK. 190302236

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Aditya Ardy Pratama
NIM : 21.11.3986

Menyatakan bahwa Skripsi dengan judul berikut:

**PERBANDINGAN ALGORITMA NAIVE BAYES DAN SVM DALAM
ANALISIS SENTIMEN TWITTER TENTANG KENAIKAN PPN 12%**

Dosen Pembimbing : Dwi Nurani, M.Kom.

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 Agustus 2025

Yang Menyatakan,



ADITYA ARDY PRATAMA

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa atas segala rahmat, karunia, dan kemudahan yang diberikan, karya sederhana ini saya persembahkan kepada:

1. Kedua orang tua tercinta, Bapak Kadimin dan Ibu Sutarmi yang selalu memberikan doa, dukungan, kasih sayang, serta pengorbanan yang tiada henti. Terima kasih atas segala perjuangan dan kepercayaan yang telah diberikan sehingga saya dapat menyelesaikan pendidikan ini.
2. Keluarga dan saudara-saudara saya, yang selalu memberikan semangat, motivasi, dan dukungan moral dalam setiap proses yang saya jalani.
3. Bapak/Ibu Dosen pembimbing dan seluruh dosen, yang telah membimbing, memberikan ilmu, arahan, serta masukan yang sangat berarti dalam penyusunan skripsi ini.
4. Sahabat dan teman-teman seperjuangan, yang selalu memberikan semangat, bantuan, dan kebersamaan selama masa perkuliahan hingga penyusunan skripsi ini.
5. Serta semua pihak yang tidak dapat saya sebutkan satu per satu, yang telah membantu dan mendukung saya dalam menyelesaikan karya ini.
6. Semoga karya ini dapat memberikan manfaat dan menjadi langkah awal menuju masa depan yang lebih baik.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada program studi yang penulis tempuh. Penulis menyadari bahwa dalam proses penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, dan bimbingan 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. Kusrini, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
4. Ibu Dwi Nurani, M.Kom., selaku dosen pembimbing yang telah memberikan waktu, arahan, saran, kritik, dan motivasi sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
5. Bapak Kadimin dan Ibu Sutarmi, selaku orang tua penulis yang selalu memberikan doa, dukungan, dan motivasi tanpa henti.

Yogyakarta, 19 Agustus 2025

Penulis

DAFTAR ISI

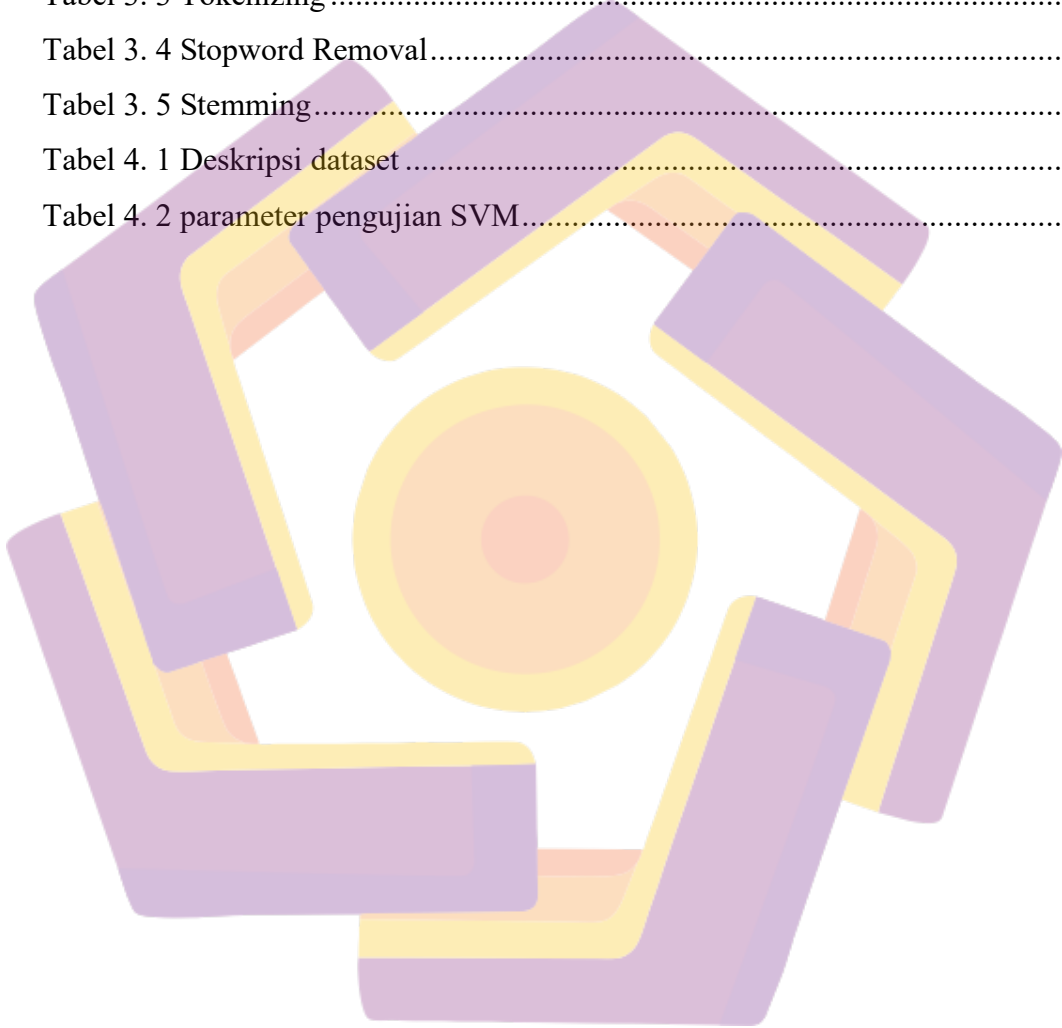
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xiv
DAFTAR LAMBANG DAN SINGKATAN	xv
DAFTAR ISTILAH	xvi
INTISARI	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5

2.1	Studi Literatur	5
2.2	Dasar Teori	12
2.2.1	Pajak Pertambahan Nilai (PPN) 12%	12
2.2.2	Python	12
2.2.3	Analisis Sentimen (Sentiment Analysis)	12
2.2.4	Crawling Data	13
2.2.5	Preprocessing Data	13
2.2.6	Labeling	13
2.2.7	GridSearchCV	14
2.2.8	Confusion Matrix.....	14
2.2.9	TF-IDF	15
2.2.10	Teknik Penyeimbangan Data (Data Balancing Techniques)	15
2.2.11	Naive Bayes	17
2.2.12	Support Vector Machine (SVM)	18
BAB III METODE PENELITIAN		21
3.1	Objek Penelitian	21
3.2	Alur Penelitian	21
3.2.1	Pengumpulan data	22
3.2.2	Duplicate Handling.....	22
3.2.3	Labeling	22
3.2.4	Pra-pemrosesan Data (Preprocessing)	23
3.2.5	TF-IDF	26
3.2.6	Model Naïve Bayes	26
3.2.7	Model SVM (Support Vector Machine)	27
3.3	Alat dan Bahan.....	27

3.3.1	Alat Penelitian	27
3.3.2	Bahan Penelitian.....	28
BAB IV		29
HASIL DAN PEMBAHASAN		29
4.1	Deskripsi Dataset	29
4.2	Pengujian Model Naïve Bayes.....	30
4.2.1	Penambahan SMOTE Pada Naïve Bayes	37
4.3	Pengujian model Support Vector Machine (SVM).....	44
4.3.1	SVM Kernel Linear	44
4.3.2	SVM Kernel RBF	53
4.3.3	SVM Kernel Polynomial.....	62
4.3.4	SVM Kernel Sigmoid	70
4.3.5	SVM Hasil Terbaik + SMOTE.....	78
4.4	Hasil.....	88
BAB V		89
5.1	Kesimpulan	89
5.2	Saran.....	90
REFERENSI.....		91
LAMPIRAN		96

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	7
Tabel 3. 1 Casefolding	23
Tabel 3. 2 Cleaning.....	24
Tabel 3. 3 Tokenizing	24
Tabel 3. 4 Stopword Removal.....	25
Tabel 3. 5 Stemming.....	25
Tabel 4. 1 Deskripsi dataset	29
Tabel 4. 2 parameter pengujian SVM.....	44



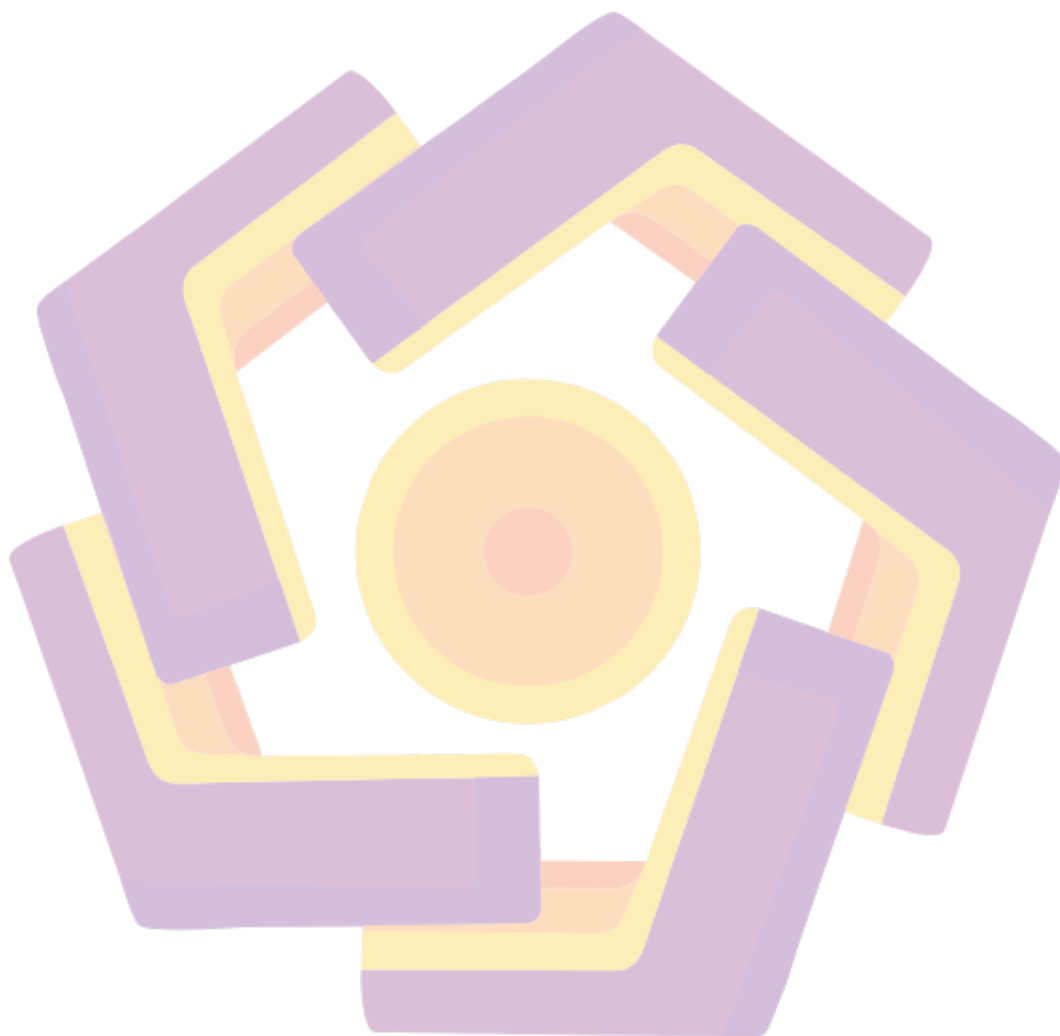
DAFTAR GAMBAR

Gambar 3. 1 Alur penelitian.....	21
Gambar 4. 1 model Naïve Bayes.....	30
Gambar 4. 2 kode untuk menampilkan Confusion Matrix pada model Naïve Bayes	31
Gambar 4. 3 Confusion Matrix Naïve Bayes	32
Gambar 4. 4 kode distribusi label Naïve Bayes	33
Gambar 4. 5 hasil distribusi label Naïve Bayes.....	34
Gambar 4. 6 kode WordCloud negatif.....	35
Gambar 4. 7 Wordcloud negatif.....	35
Gambar 4. 8 kode WordCloud positif.....	36
Gambar 4. 9 WordCloud positif.....	36
Gambar 4. 10 model Naïve Bayes + SMOTE.....	37
Gambar 4. 11 kode Confusion Matrix model Naïve Bayes + SMOTE	38
Gambar 4. 12 Confusion Matrix Naïve Bayes + SMOTE	39
Gambar 4. 13 kode distribusi label Naïve Bayes + SMOTE	40
Gambar 4. 14 hasil distribusi label Naïve Bayes + SMOTE.....	41
Gambar 4. 15 kode WordCloud negatif.....	42
Gambar 4. 16 WordCloud negatif.....	42
Gambar 4. 17 kode WordCloud positif.....	43
Gambar 4. 18 WordCloud positif.....	43
Gambar 4. 19 parameter terbaik dan score cross validation terbaik SVM Kernel Linear	45
Gambar 4. 20 model SVM Kernel Linear.....	46
Gambar 4. 21 kode Confusion Matrix SVM Kernel Linear	47
Gambar 4. 22 Confusion Matrix SVM Kernel Linear	48
Gambar 4. 23 kode distribusi label SVM Kernel Linear	49
Gambar 4. 24 hasil distribusi label SVM Kernel Linear	50
Gambar 4. 25 kode WordCloud negatif.....	51
Gambar 4. 26 WordCloud negatif	51

Gambar 4. 27 kode WordCloud positif.....	52
Gambar 4. 28 WordCloud positif.....	53
Gambar 4. 29 parameter terbaik dan score cross validation terbaik SVM Kernel RBF.....	54
Gambar 4. 30 model SVM Kernel RBF.....	55
Gambar 4. 31 kode Confusion Matrix SVM Kernel RBF	56
Gambar 4. 32 Confusion Matrix SVM Kernel RBF.....	57
Gambar 4. 33 kode distribusi label SVM Kernel RBF	58
Gambar 4. 34 hasil distribusi label SVM Kernel RBF	59
Gambar 4. 35 kode WordCloud negatif.....	60
Gambar 4. 36 WordCloud negatif.....	60
Gambar 4. 37 kode WordCloud positif.....	61
Gambar 4. 38 WordCloud positif.....	61
Gambar 4. 39 parameter terbaik dan score cross validation terbaik SVM Kernel Polynomial	62
Gambar 4. 40 model SVM Kernel Polynomial	63
Gambar 4. 41 kode Confusion Matrix SVM Kernel Polynomial.....	64
Gambar 4. 42 Confusion Matrix SVM Kernel Polynomial	65
Gambar 4. 43 kode distribusi label SVM Kernel Polynomial	66
Gambar 4. 44 hasil distribusi label SVM Kernel Polynomial.....	67
Gambar 4. 45 kode WordCloud negatif.....	68
Gambar 4. 46 WordCloud negatif	68
Gambar 4. 47 kode WordCloud positif.....	69
Gambar 4. 48 WordCloud positif.....	69
Gambar 4. 49 parameter terbaik dan score cross validation terbaik SVM Kernel Sigmoid.....	70
Gambar 4. 50 model SVM Kernel Sigmoid.....	71
Gambar 4. 51 kode Confusion Matrix SVM Kernel Sigmoid	72
Gambar 4. 52 Confusion Matrix SVM Kernel Sigmoid.....	73
Gambar 4. 53 kode distribusi label SVM Kernel Sigmoid	74
Gambar 4. 54 hasil distribusi label SVM Kernel Sigmoid	75

Gambar 4. 55 kode WordCloud negatif.....	76
Gambar 4. 56 WordCloud negatif	76
Gambar 4. 57 kode WordCloud positif.....	77
Gambar 4. 58 WordCloud positif.....	78
Gambar 4. 59 parameter terbaik dan score cross validation terbaik SVM Kernel RBF+SMOTE.....	79
Gambar 4. 60 model SVM Kernel RBF+SMOTE	81
Gambar 4. 61 kode Confusion Matrix SVM Kernel RBF+SMOTE	82
Gambar 4. 62 Confusion Matrix SVM Kernel RBF+SMOTE	83
Gambar 4. 63 kode distribusi label SVM Kernel RBF+SMOTE.....	84
Gambar 4. 64 hasil distribusi label SVM Kernel RBF+SMOTE.....	85
Gambar 4. 65 kode WordCloud negatif.....	86
Gambar 4. 66 WordCloud negatif	86
Gambar 4. 67 kode WordCloud positif.....	87
Gambar 4. 68 WordCloud positif.....	88

DAFTAR LAMPIRAN

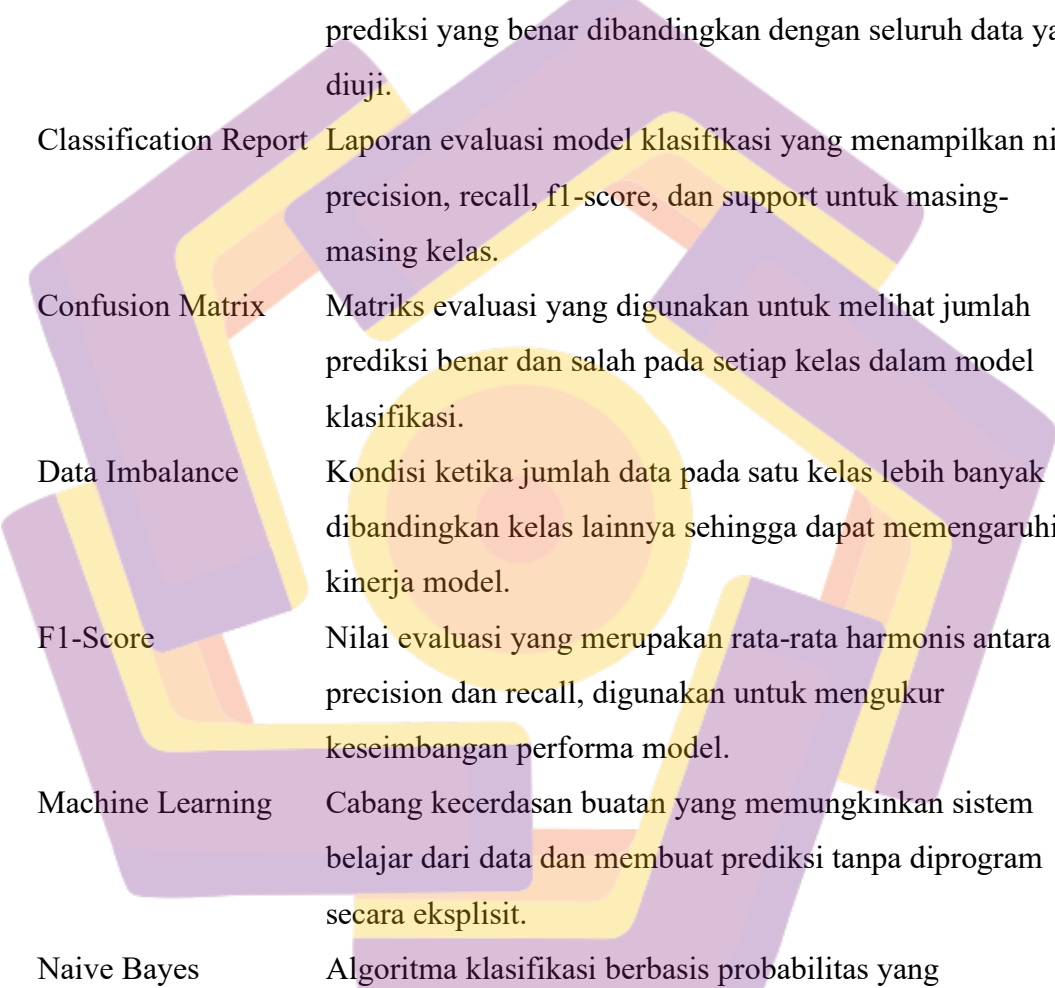


DAFTAR LAMBANG DAN SINGKATAN

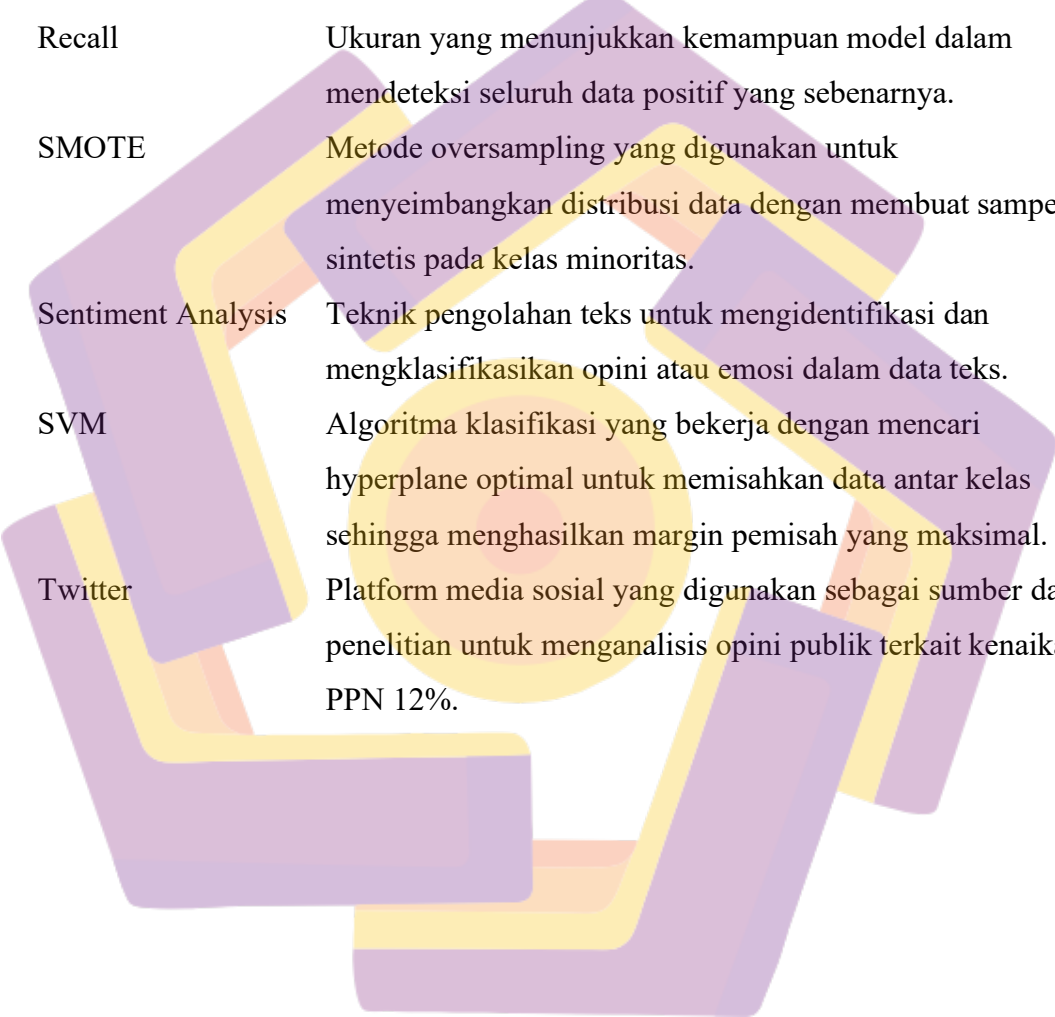
CSV	Comma Separated Values
F1-Score	Harmonic Mean of Precision and Recall
ML	Machine Learning
NB	Naive Bayes
PPN	Pajak Pertambahan Nilai
RBF	Radial Basis Function
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
TF-IDF	Term Frequency – Inverse Document Frequency



DAFTAR ISTILAH



Algoritma	Serangkaian langkah atau prosedur sistematis yang digunakan untuk menyelesaikan suatu permasalahan atau melakukan proses komputasi tertentu.
Accuracy (Akurasi)	Ukuran kinerja model yang menunjukkan persentase prediksi yang benar dibandingkan dengan seluruh data yang diuji.
Classification Report	Laporan evaluasi model klasifikasi yang menampilkan nilai precision, recall, f1-score, dan support untuk masing-masing kelas.
Confusion Matrix	Matriks evaluasi yang digunakan untuk melihat jumlah prediksi benar dan salah pada setiap kelas dalam model klasifikasi.
Data Imbalance	Kondisi ketika jumlah data pada satu kelas lebih banyak dibandingkan kelas lainnya sehingga dapat memengaruhi kinerja model.
F1-Score	Nilai evaluasi yang merupakan rata-rata harmonis antara precision dan recall, digunakan untuk mengukur keseimbangan performa model.
Machine Learning	Cabang kecerdasan buatan yang memungkinkan sistem belajar dari data dan membuat prediksi tanpa diprogram secara eksplisit.
Naive Bayes	Algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi independensi antar fitur.
Negative Sentiment	Kelas sentimen yang menunjukkan opini atau persepsi negatif terhadap suatu topik.
Positive Sentiment	Kelas sentimen yang menunjukkan opini atau persepsi positif terhadap suatu topik.



Precision	Ukuran yang menunjukkan proporsi prediksi positif yang benar dibandingkan seluruh prediksi positif yang dihasilkan model.
Preprocessing	Tahapan pembersihan dan persiapan data teks sebelum dilakukan proses analisis, seperti cleaning, case folding, tokenization, stopwords removal, dan stemming.
Recall	Ukuran yang menunjukkan kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya.
SMOTE	Metode oversampling yang digunakan untuk menyeimbangkan distribusi data dengan membuat sampel sintetis pada kelas minoritas.
Sentiment Analysis	Teknik pengolahan teks untuk mengidentifikasi dan mengklasifikasikan opini atau emosi dalam data teks.
SVM	Algoritma klasifikasi yang bekerja dengan mencari hyperplane optimal untuk memisahkan data antar kelas sehingga menghasilkan margin pemisah yang maksimal.
Twitter	Platform media sosial yang digunakan sebagai sumber data penelitian untuk menganalisis opini publik terkait kenaikan PPN 12%.

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat di Twitter terkait kebijakan kenaikan Pajak Pertambahan Nilai (PPN) sebesar 12% serta membandingkan kinerja algoritma Naive Bayes dan Support Vector Machine (SVM) dalam klasifikasi sentimen. Data dikumpulkan dari Twitter dan melalui tahapan *preprocessing* yang meliputi pembersihan teks, *case folding*, *tokenization*, *stopword removal*, dan *stemming*. Data kemudian dibagi menjadi data latih dan data uji untuk proses evaluasi model. Hasil eksperimen menunjukkan bahwa model SVM yang dioptimasi menggunakan kernel Radial Basis Function (RBF) memperoleh kinerja terbaik dengan tingkat akurasi sebesar 79,93%. Berdasarkan *classification report*, SVM menunjukkan performa lebih konsisten dalam mengidentifikasi kelas Negative (f1-score = 0,86) dibandingkan kelas Positive (f1-score = 0,65), dengan nilai weighted average f1-score sebesar 0,80. Sementara itu, model Naive Bayes memperoleh akurasi sebesar 74,07%, dengan f1-score sebesar 0,83 pada kelas Negative dan 0,46 pada kelas Positive, serta weighted average f1-score sebesar 0,71. Berdasarkan hasil tersebut, dapat disimpulkan bahwa SVM memiliki performa yang lebih unggul dibandingkan Naive Bayes dalam analisis sentimen Twitter terkait kebijakan kenaikan PPN 12%.

Kata kunci : PPN 12%, Naive Bayes, Support Vector Machine

ABSTRACT

This study investigates public sentiment on Twitter regarding the 12% Value Added Tax (VAT) increase policy and compares the performance of Naive Bayes and Support Vector Machine (SVM) algorithms for sentiment classification. The dataset was collected from Twitter and processed through text preprocessing stages, including cleaning, case folding, tokenization, stopwords removal, and stemming. The data were then divided into training and testing sets for model evaluation. Experimental results indicate that the optimized SVM model with the Radial Basis Function (RBF) kernel achieved the best performance, with an accuracy of 79.93%. The classification report shows that SVM performed more consistently in identifying the Negative class (f1-score = 0.86) compared to the Positive class (f1-score = 0.65), with an overall weighted average f1-score of 0.80. In contrast, the Naive Bayes model obtained an accuracy of 74.07%, with an f1-score of 0.83 for the Negative class and 0.46 for the Positive class, resulting in a weighted average f1-score of 0.71. These findings demonstrate that SVM outperforms Naive Bayes in sentiment classification of Twitter data related to the VAT increase policy, making it a more effective approach for this task.

Keywords : PPN 12%, Naïve Bayes, Support Vector Machine