

ANALISIS SENTIMEN TERHADAP KEBIJAKAN KENAIKAN TARIF PAJAK PERTAMBAHAN NILAI (PPN) 12% DI INDONESIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE PADA MEDIA SOSIAL X

M. Farish Ramadhan Ferdinansyah, Karina Auliasari, Mira Orisa

Teknik Informatika, Institut Teknologi Nasional Malang

Jalan Raya Karanglo km 2 Malang, Indonesia

2118054@scholar.itn.ac.id

ABSTRAK

Kebijakan pemerintah Indonesia untuk menaikkan tarif Pajak Pertambahan Nilai (PPN) dari 11% menjadi 12% mulai 1 Januari 2025 menimbulkan berbagai reaksi dari masyarakat, khususnya di media sosial. Kenaikan ini dianggap menambah beban ekonomi, terutama bagi pelaku UMKM yang khawatir terhadap penurunan daya saing produk akibat naiknya harga jual. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan tersebut dengan memanfaatkan data dari media sosial X, guna memahami persepsi publik secara lebih akurat. Metode yang digunakan adalah *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF), yang dikenal memiliki akurasi tinggi dalam klasifikasi data. Data yang dianalisis merupakan unggahan berbahasa Indonesia yang dikumpulkan dalam periode 7 Oktober 2023 hingga 31 Januari 2025. Proses klasifikasi dibatasi pada dua kategori sentimen, yaitu positif dan negatif. Hasil penelitian menunjukkan bahwa mayoritas masyarakat memberikan sentimen negatif terhadap kebijakan kenaikan tarif PPN, terutama disebabkan kekhawatiran akan dampak ekonomi lanjutan. Hasil pengujian yang dilakukan terhadap data dari media sosial X dengan dua skenario pembagian data, yaitu 70:30 dan 80:20. Hasil pengujian menunjukkan bahwa model mencapai akurasi sebesar 89,37% pada split 70:30 dan 86,12% pada split 80:20.

Kata kunci : Analisis Sentimen, Pajak Pertambahan Nilai (PPN), Confussion Matrix, Machine Learning, Support Vector Machine (SVM)

1. PENDAHULUAN

Kebijakan pemerintah terkait kenaikan tarif Pajak Pertambahan Nilai (PPN) menjadi isu hangat di Indonesia. Sesuai amanat Undang-Undang Nomor 7 Tahun 2021 tentang Harmonisasi Peraturan Perpajakan, tarif PPN akan naik dari 11% menjadi 12% mulai 1 Januari 2025 [1]. PPN sendiri dikenakan atas pertambahan nilai barang atau jasa dalam proses produksi hingga ke konsumen akhir [2]. Kenaikan tarif ini berdampak luas, baik bagi pengusaha maupun konsumen. Pengusaha, khususnya pelaku UMKM, merasa terbebani karena harus menanggung selisih pajak dan menaikkan harga jual, yang dapat menurunkan daya saing produk mereka [3][4].

PPN menyumbang sekitar 35% dari total penerimaan pajak, namun pertumbuhannya lambat sejak 2012 dan kembali melemah pada 2020. Kenaikan tarif menjadi 12% bertujuan mengurangi defisit anggaran dan meningkatkan penerimaan negara, meskipun tantangan global seperti naiknya harga bahan baku tetap menjadi hambatan [1]. Kebijakan ini memicu beragam opini publik yang disampaikan melalui media sosial X, baik dalam bentuk sentimen positif maupun negatif [3]. Analisis sentimen penting dilakukan untuk memahami sikap masyarakat terhadap kebijakan ini [2].

Penelitian ini bertujuan melakukan analisis sentimen terhadap kebijakan kenaikan tarif PPN 12% menggunakan metode Support Vector Machine (SVM) pada data dari media sosial X. SVM dipilih karena memiliki akurasi tinggi dalam klasifikasi sentimen, sehingga mampu menggambarkan secara

akurat sejauh mana kebijakan ini diterima atau ditolak oleh publik.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian oleh Nur Fitriyah, Budi Warsito, dan Di Asih I Maruddani (2020) menggunakan metode Support Vector Machine (SVM) untuk menganalisis sentimen terhadap Gojek dari 1.500 tweet, dengan kernel linear dan RBF. Hasilnya, pelabelan manual dan sentiment scoring sama-sama menghasilkan akurasi 79,19%, dan kernel RBF lebih unggul dibanding linear [5].

Muh. Fitra Rizki, Karina Auliasari, dan Renaldi Primaswara Prasetya (2021) meneliti sentimen terkait cyberbullying di media sosial dengan SVM. Dari kurang dari 100 tweet bertagar #cebong dan #kadrun, sistem menghasilkan akurasi 70% dengan kecepatan pemrosesan rata-rata 0,000989 data/ms dan waktu eksekusi 101100,2 ms [6].

Raflizar Deswandi Yahya, Suryo Adi Wibowo, dan Nurlaily Vendyansyah (2024) menggunakan SVM untuk mendeteksi ujaran kebencian terkait Pemilu 2024. Dari 100 tweet, model mencapai *recall* 76%, *precision* 96%, dan akurasi 81% [7].

Styawati, Nirwana Hendrastuty, dan rekan (2024) menganalisis opini masyarakat terhadap Program Kartu Prakerja di Twitter menggunakan SVM. Hasil pengujian menunjukkan bahwa kernel linear lebih unggul dengan akurasi 98,67%, *precision* 98%, *recall* 99%, dan F1-score 98% [8].

Sri Lestari dan Safira Berliani (2023) meneliti sentimen isu “Pecat Sri Mulyani” di Twitter dengan Naive Bayes dan SVM. Dari 1.958 tweet, metode SVM mencapai akurasi 95,13% sementara Naive Bayes memperoleh 96,14% [9].

2.2. Text Mining

Text mining adalah teknik analisis yang digunakan untuk mengekstraksi informasi berupa pola, tren, atau insight bermakna dari data teks tidak terstruktur, seperti komentar di media sosial atau dokumen digital [10]. Proses ini melibatkan beberapa tahapan inti—mulai dari preprocessing teks (seperti tokenisasi dan TF-IDF), hingga klasifikasi menggunakan algoritma pembelajaran mesin.

2.3. Analisis Sentimen

Analisis sentimen merupakan cabang ilmu yang berperan dalam menganalisis, memproses, dan mengekstrak data teks dari berbagai sumber, seperti layanan, produk, individu, organisasi, peristiwa, atau topik tertentu [15]. Analisis sentimen dilakukan untuk mengevaluasi opini serta kecenderungan suatu pendapat terhadap suatu topik, baik dalam aspek positif maupun negatif. Analisis sentimen dapat diterapkan pada berbagai bidang, termasuk ekonomi, politik, sosial, dan hukum.

2.4. Term Frequency (TF)

Term Frequency (TF) menunjukkan seberapa sering suatu kata muncul dalam sebuah dokumen tertentu. Semakin sering kata tersebut muncul dalam dokumen, semakin tinggi nilai TF-nya, yang berarti bobot kata tersebut juga akan semakin besar. Pada persamaan (1) menunjukkan perhitungan dari *Term Frequency* (TF) [7].

$$TF = 1 + \log(F_{t,d}), F_{t,d} > 0 \quad (1)$$

Keterangan :

TF = Term Frequency

F_{t,d} = Frequency Term pada dokumen

2.5. Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) adalah penghitungan yang mencerminkan sejauh mana suatu kata tersebar luas dalam koleksi dokumen tertentu. IDF menggambarkan tingkat ketersediaan suatu kata dalam seluruh dokumen [7]. Perhitungan *Inverse Document Frequency* (IDF) dapat dilakukan menggunakan persamaan (2).

$$IDF = \log\left(\frac{D}{df}\right) \quad (2)$$

Keterangan :

IDF = Inverse Document Frequency.

D = Jumlah keseluruhan dokumen.

Df = Jumlah dokumen yang mengandung term.

2.6. Confussion Matrix

Proses evaluasi dilakukan untuk mengukur kualitas suatu *classifier*. Penilaian terhadap performa

sistem klasifikasi menjadi langkah penting dalam menentukan sejauh mana tingkat akurasi prediksi yang dihasilkan. Dengan evaluasi yang tepat, dapat diketahui seberapa baik model dalam mengklasifikasikan data secara akurat [7]. Terdapat 3 metode Evaluasi yang dapat digunakan, yaitu *Accuracy*, *Precision*, *Recall*. Nilai-nilai ini diperoleh dari *confussion matrix* seperti pada Tabel 1.

Tabel 1. Confussion Matrix

Predicted Class	True Class	
	Positif	Negatif
Positif	True Positif (TP)	False Positif (FP)
Negatif	False Positif (FP)	True Negatif (TN)

Keempat parameter diatas digunakan untuk menghitung 3 metode evaluasi yakni :

- Recall* adalah metrik yang membandingkan jumlah data relevan yang berhasil diklasifikasikan dengan benar oleh model terhadap seluruh data yang sebenarnya relevan. Nilai *recall* yang tinggi menunjukkan bahwa model mampu mendeteksi sebagian besar data relevan tanpa banyak yang terlewat. *Recall* dapat dihitung melalui rumusan berikut :

$$Recall = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

- Precision* adalah rasio antara jumlah data yang benar-benar relevan dari semua data yang diklasifikasikan sebagai relevan oleh model. Metrik ini mengukur ketepatan model dalam memprediksi kategori tertentu. Nilai *precision* yang tinggi menunjukkan bahwa sebagian besar hasil klasifikasi relevan dan model memiliki tingkat kesalahan rendah. *Precision* dapat dihitung dengan rumus:

$$Precision = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

- Accuracy* adalah metrik yang mengukur seberapa besar proporsi prediksi yang benar dari keseluruhan data yang diuji. Metrik ini mencerminkan kinerja umum model dalam mengklasifikasikan data ke dalam kategori yang tepat, baik positif maupun negatif. Nilai *accuracy* yang tinggi menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. *Accuracy* dapat dihitung dengan rumus:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \times 100\% \quad (5)$$

2.7. Machine Learning

Machine Learning (ML) adalah cabang kecerdasan buatan yang memungkinkan sistem belajar dari data tanpa pemrograman eksplisit [10]. ML digunakan untuk menemukan pola dan memprediksi data baru. Analisis sentimen mengevaluasi opini dalam teks untuk menentukan kecenderungan positif atau negatif, dan dapat diterapkan di berbagai bidang. Pre-processing diperlukan untuk membersihkan data

dari noise, terutama pada teks media sosial yang tidak terstruktur [12][13].

2.8. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma yang banyak digunakan untuk klasifikasi, termasuk dalam analisis sentimen. SVM efektif dalam menangani data berdimensi tinggi dan memisahkan data ke dalam kategori tertentu, sehingga relevan untuk penelitian ini. SVM merupakan metode pembelajaran yang dapat digunakan untuk klasifikasi, regresi, dan prediksi sistem. Tujuan utamanya adalah mengukur frekuensi kemunculan kata dan mengklasifikasikan kalimat ke dalam sentimen positif atau negatif. Keunggulan SVM terletak pada kemampuannya mengelola ruang input berdimensi tinggi dan merepresentasikan dokumen sebagai vektor [13]. Perhitungan metode *Support Vector Machine* dapat dilakukan menggunakan persamaan (6) :

$$f(x) = w^T x + b \quad (6)$$

Keterangan :

$f(x)$ = Fungsi keputusan

w = Vektor bobot

x = Vektor fitur input

b = Bias (intersep)

3. METODE PENELITIAN

3.1. Pengumpulan Data

Data dalam penelitian ini dikumpulkan dari media sosial X menggunakan library tweet-harvests dengan metode crawling. Kata kunci yang digunakan meliputi: ppn, ppn naik, ppn 12%, kenaikan ppn, dan kebijakan ppn naik, dalam rentang waktu 7 Oktober 2023 hingga 31 Januari 2025. Hasil crawling disimpan dalam format .csv, dengan total 1.223 baris data yang berhasil dikumpulkan.

3.2. Pre-Processing Data

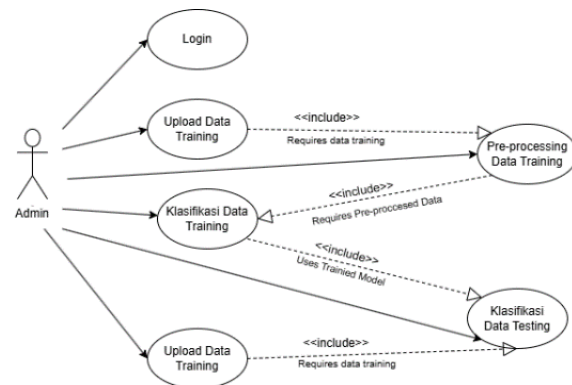


Gambar 1. Alur Pre-Processing

Sebelum masuk ke proses klasifikasi, Langkah pertama yang perlu dilakukan sebelum proses klasifikasi adalah *pre-processing* teks. *Pre-processing* data mencakup beberapa tahapan, antara lain pembersihan data (*cleaning*), *case folding*, tokenisasi, normalisasi, penghapusan *stopwords*, dan *stemming*. Proses *pre-processing* yang dijelaskan ini dapat dilihat pada Gambar 1.

3.3. Use Case Diagram

Use Case Diagram yang akan dikembangkan dalam sistem ini dapat dilihat pada Gambar 2.

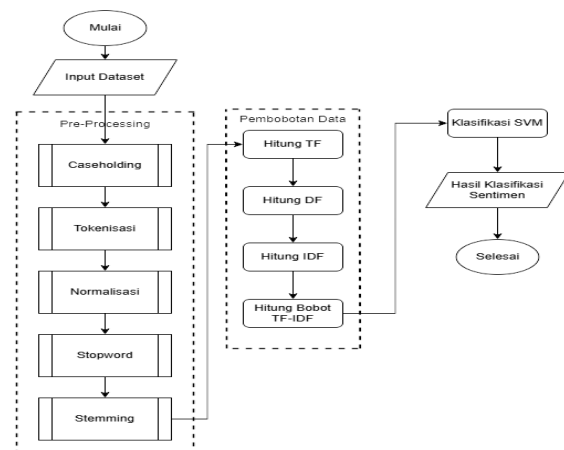


Gambar 2. Use Case Diagram

Gambar 2 menunjukkan alur sistem analisis sentimen berbasis SVM dengan *Admin* sebagai peran utama. Proses dimulai dari login, lalu *Admin* mengunggah dataset training untuk dipre-processing dan digunakan melatih model. Dataset testing juga dipre-processing sebelum diklasifikasikan oleh model. Setiap tahap harus diselesaikan secara berurutan agar proses berjalan sistematis.

3.4. Flowchart Sistem

Flowchart sistem yang akan dikembangkan dapat dilihat pada Gambar 3.



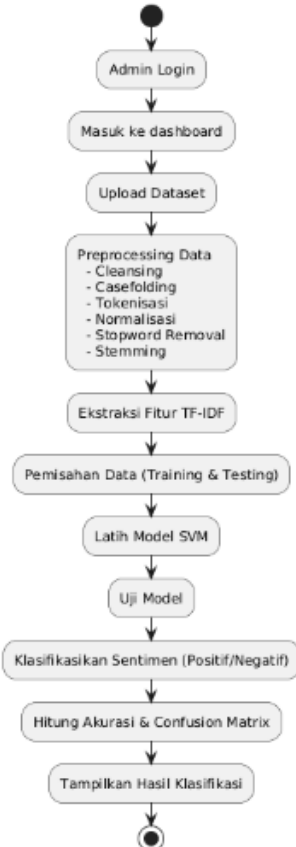
Gambar 3. Flowchart Sistem

Gambar 3 menunjukkan alur sistem analisis sentimen yang dimulai dari input data, dilanjutkan dengan *pre-processing* untuk membersihkan teks, kemudian pembobotan menggunakan TF-IDF, dan

diakhiri dengan klasifikasi menggunakan algoritma SVM untuk menghasilkan output berupa sentimen positif atau negatif.

3.5. Activity Diagram

Activity Diagram yang akan dikembangkan dalam sistem ini dapat dilihat pada Gambar 4.

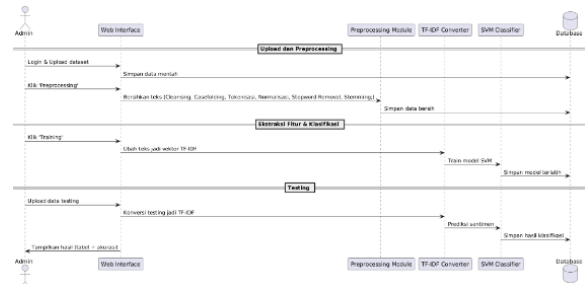


Gambar 4. Activity Diagram

Gambar 4 menunjukkan activity diagram sistem analisis sentimen kebijakan kenaikan tarif PPN 12%. Proses dimulai dari login Admin, upload dataset, preprocessing (*cleansing*, *casefolding*, tokenisasi, normalisasi, *stopword removal*, *stemming*), lalu konversi TF-IDF, kemudian pembagian data *training/testing*, hingga pelatihan dan pengujian SVM. Hasil klasifikasi sentimen positif/negatif ditampilkan bersama akurasi dan confusion matrix.

3.6. Sequence Diagram

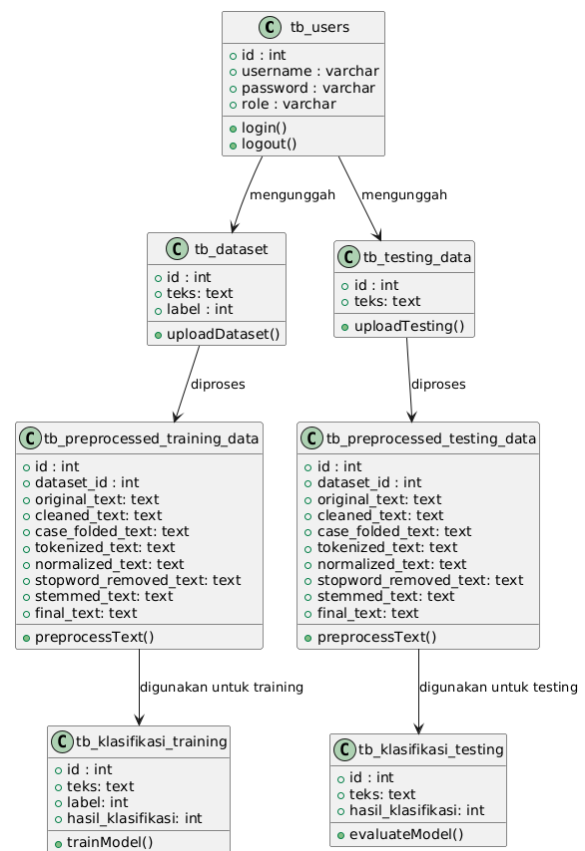
Sequence Diagram yang akan dikembangkan dalam sistem ini dapat dilihat pada Gambar 5, memperlihatkan *sequence diagram* sistem analisis sentimen PPN 12% menggunakan SVM. Diagram ini menggambarkan interaksi antara *Admin*, *Web Interface*, *Preprocessing Module*, *TF-IDF Converter*, *SVM Classifier*, dan *Database*. Proses dimulai dari login dan upload dataset, dilanjutkan dengan preprocessing, konversi TF-IDF, pelatihan model SVM, hingga pengujian *data testing*. Hasil klasifikasi sentimen positif atau negatif beserta akurasi disimpan dan ditampilkan secara terstruktur.



Gambar 5. Sequence Diagram

3.7. Class Diagram

Class Diagram yang akan dikembangkan dalam sistem ini dapat dilihat pada Gambar 6.

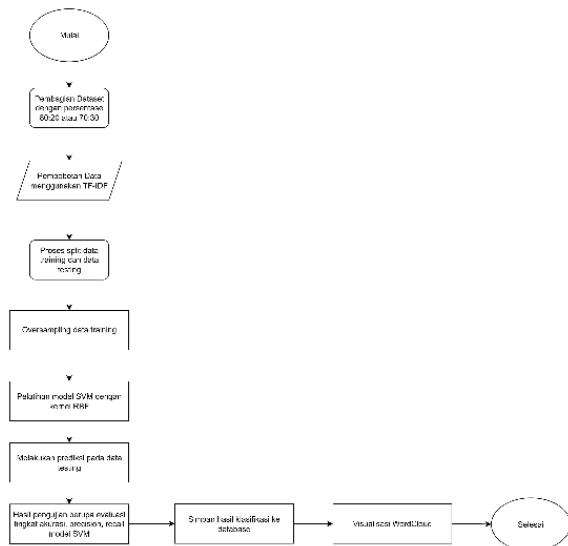


Gambar 6. Class Diagram

Gambar 6 menampilkan class diagram sistem analisis sentimen PPN 12% berbasis SVM. Diagram terdiri dari delapan kelas utama yang merepresentasikan tabel database mulai dari pengguna, dataset, hingga hasil klasifikasi. Setiap kelas memiliki atribut sesuai tabel serta fungsi inti seperti upload, preprocessing, training, dan evaluasi, sehingga membentuk alur data sistematis dari input hingga keluaran.

3.8. Flowchart Metode Support Vector Machine

Flowchart metode *Support Vector Machine* yang akan dikembangkan dapat dilihat pada Gambar 6.



Gambar 7. Flowchart Metode SVM

Gambar 7 menunjukkan alur sistem analisis sentimen hingga menghasilkan klasifikasi positif dan negatif. Proses diawali dengan pre-processing data, meliputi *cleaning*, *case folding*, tokenisasi, *stopword*

removal, dan *stemming*, untuk menyaring teks relevan dan menghilangkan kata tidak baku. Data bersih kemudian diberi bobot menggunakan metode TF-IDF, lalu diklasifikasikan menggunakan algoritma SVM. Hasil akhirnya adalah kategori sentimen positif atau negatif.

4. HASIL DAN PEMBAHASAN

4.1. Implementasi Support Vector Machine

Sebanyak 1.223 data teks dari media sosial X digunakan sebagai dataset dalam penelitian ini dengan topik pembahasan seputar kenaikan tarif PPN 12% di Indonesia dengan menggunakan keyword ppn, ppn naik, dan ppn 12%. Dataset tersebut kemudian dilakukan pelabelan oleh anotator mencakup dua kategori sentimen: positif dan negatif. Dengan distribusi sebagai berikut:

- Positif: 675 data (55.19%)
- Negatif: 548 data (44.81%)

Dataset tersebut diambil melalui proses *crawling* dengan menggunakan library *tweet harvest*. Untuk memudahkan penelitian, data yang telah diperoleh disimpan dalam format CSV. Tabel 2 merupakan contoh dataset yang telah diberikan label oleh anotator.

Tabel 2. Pelabelan Data Oleh Anotator

No.	Teks	Label
1.	PPN 12% membantu mendukung pembangunan nasional dan memperkuat fondasi ekonomi. #PPN12Persen https://t.co/ysakKBoHGa	Positif
2.	Masih ada hasilnya kok. Kemarin #TolakPPN12Persen yang digalakkan @barengwarga berhasil ngerubah aturan PPN 12% walau ya emang kampretnya the last minute banget. Makanya ayok sesama warga saling support apalagi kalau ada gerakan turun bareng	Negatif
...	...	...
1223.	PPN adalah investasi kita untuk masa depan yang lebih cerah. #PPN12Persen #EkonomiKuat #StabilitasNasional #IndonesiaMaju https://t.co/YisdVn4uuS	Positif

4.2. Pre-Processing Dataset

Langkah awal sebelum klasifikasi adalah pre-processing teks, yang bertujuan membersihkan dan menyiapkan data agar siap dianalisis. Tahapan pre-processing meliputi standarisasi format data (*cleaning*), *case folding*, *tokenization*, *normalization*, penghapusan kata umum (*stopword removal*), dan *stemming*. Proses ini dilakukan sebagai berikut:

a. Cleansing

Tabel 3 merupakan contoh tahapan *cleansing*, pada tahap ini dilakukan pembersihan data seperti menghilangkan mention dan menghilangkan punctuasi. Teks mengalami pembersihan dan penyederhanaan melalui tahapan ini.

Tabel 3. Tahapan *Cleaning*

Input Proses	Output Proses
PPN 12% membantu mendukung pembangunan nasional dan memperkuat fondasi ekonomi. #PPN12Persen https://t.co/ysakKBoHGa	PPN 12% membantu mendukung pembangunan nasional dan memperkuat fondasi ekonomi. PPN12Persen

b. Casefolding

Tabel 4 merupakan contoh tahapan *casefolding*, yaitu mengubah seluruh karakter menjadi huruf kecil. Inputan pada proses ini adalah hasil dari proses *cleansing*. Perubahan dapat dilihat pada kata Kalo yang dikonversi menjadi huruf kecil.

Tabel 4. Tahapan *Casefolding*

Input Proses	Output Proses
PPN 12% membantu mendukung pembangunan nasional dan memperkuat fondasi ekonomi. PPN12Persen	ppn membantu mendukung pembangunan nasional dan memperkuat fondasi ekonomi ppnpersen

c. Tokenisasi

Tabel 5 merupakan contoh tahapan tokenisasi. Tokenisasi menguraikan teks menjadi unit-unit kata. Tujuan tahapan ini adalah untuk mempermudah pengolahan teks dengan cara memisahkannya menjadi elemen-elemen yang lebih kecil.

Tabel 5. Tahapan Tokenisasi

Input Proses	Output Proses
ppn membantu mendukung pembangunan nasional dan memperkuat fondasi ekonomi ppnpersen	'ppn', 'membantu', 'mendukung', 'pembangunan', 'nasional', 'dan', 'memperkuat', 'fondasi', 'ekonomi', 'ppnpersen'

d. Normalisasi

Tabel 6 merupakan contoh tahapan normalisasi. Pada tahapan ini teks akan diubah menjadi bentuk dasar bahasa baku. Tujuan tahapan ini adalah untuk menyelaraskan bentuk penulisan kata. Sehingga kata konsisten dengan kosakata yang digunakan oleh model.

Tabel 6. Tahapan Normalisasi

Input Proses	Output Proses
ppn membantu mendukung pembangunan nasional dan memperkuat fondasi ekonomi ppnpersen	ppn membantu mendukung pembangunan nasional dan memperkuat fondasi ekonomi ppnpersen

e. Stopword Removal

Tabel 7 tahap Stopword Removal, kata-kata umum seperti “saya”, “dengan”, “ini”, “bukan”, dan “yang” dihapus karena tidak memberikan informasi yang signifikan untuk analisis.

Tabel 7. Tahapan Stopword Removal

Input Proses	Output Proses
ppn membantu mendukung pembangunan nasional dan memperkuat fondasi ekonomi ppnpersen	ppn membantu mendukung pembangunan nasional memperkuat fondasi ekonomi ppnpersen

f. Stemming

Tabel 8 merupakan tahapan *stemming*. Pada tahap ini dilakukan pembersihan data dengan reduksi kata ke bentuk dasar melalui penghapusan imbuhan. Pada contoh diatas, terjadi perubahan kata memproduksi menjadi Kata "kebijakan" disederhanakan menjadi "bijak", Kata "menekan" dan "ditekan" "tekan", Serta "pengeluaran" distem menjadi "keluar", karena stemmer menganggapnya sebagai bentuk turunan dari kata tersebut.

Tabel 8. Tahapan Stemming

Input Proses	Output Proses
ppn membantu mendukung pembangunan nasional memperkuat fondasi ekonomi ppnpersen	ppn bantu dukung bangun nasional kuat fondasi ekonomi ppnpersen

4.3. Hasil Klasifikasi Data Training

Setelah melakukan pengujian pada model SVM, setiap teks menerima pelabelan. Contoh hasil pengujian data antara anotator dan sistem ditunjukkan Tabel 9.

Tabel 9. Hasil Klasifikasi Data Training

No.	Teks	Label Asli	Hasil Klasifikasi
1.	Kebijakan PPN 12% ini seharusnya ditunda sampai kondisi ekonomi masyarakat membaik. Penambahan pajak di tengah ketidakpastian ekonomi tidak adil bagi rakyat. #EkonomiBerkelanjutan https://t.co/beritaindonesia	Negatif	Negatif
2.	Dengan PPN 12% negara punya peluang membangun lebih banyak fasilitas publik. Sekarang tinggal kawal pelaksanaannya. #ReformasiFiskal	Positif	Positif
3.	Saya khawatir dengan kebijakan PPN 12% yang dapat memperburuk kesenjangan sosial. Masyarakat kecil akan semakin terbebani oleh pajak yang membengkak. #KesejahteraanBersama https://t.co/politikupdate #PPN12Persen	Negatif	Positif

4.4. Pengujian Metode Data Training

Dalam penelitian ini, dataset dibagi ke dalam dua skema rasio, yaitu 70:30 dan 80:20, untuk menguji kinerja model klasifikasi. Akurasi sistem dievaluasi dengan membandingkan hasil klasifikasi data uji terhadap label anotator menggunakan confusion matrix.

a. Rasio 70% Data Training : 30% Data Testing

Tabel 10 menampilkan confusion matrix yang menunjukkan kemampuan model dalam membedakan kelas negatif dan positif. Dari 154 data negatif, 116 diklasifikasikan dengan benar dan 38 salah. Untuk 213 data positif, 212 dikenali dengan tepat dan hanya 1 yang salah klasifikasi. Secara keseluruhan, model menghasilkan 328 prediksi benar dan 39 kesalahan,

dengan akurasi tinggi sebesar 89,37%. Hasil ini menunjukkan bahwa model sangat baik dalam mengenali kedua kelas, khususnya dalam mengidentifikasi data positif dengan sangat akurat.

Tabel 10. Confusion Matrix Rasio 70:30

True Class	Predicted Class		
	Negatif	Positif	Total
Negatif	116	38	154
Positif	1	212	213
Total	117	250	367

Nilai *recall*, *precision*, dan *accuracy* akan dihitung berdasarkan *confusion matrix* yang telah dibuat. Berikut adalah hasil evaluasinya:

Tabel 11. Matrix Evaluasi Rasio 70:30

Label	Precision	Recall
Negatif	$\frac{116}{117} \times 100\% = 99.14\%$	$\frac{116}{154} \times 100\% = 75.32\%$
Positif	$\frac{212}{250} \times 100\% = 84.80\%$	$\frac{212}{213} \times 100\% = 99.53\%$
Akurasi	$\frac{116 + 212}{367} \times 100\% = \frac{328}{367} \times 100\% = 89.37\%$	

Pada Tabel 11, model menunjukkan performa yang sangat baik dalam membedakan data positif dan negatif. *Precision* untuk kelas negatif mencapai 99,14% dengan *recall* 75,32%, sedangkan *precision* pada kelas positif sebesar 84,80% dengan *recall* 99,53%. Secara keseluruhan, model memperoleh akurasi sebesar 89,37%, yang menunjukkan bahwa 89,37% dari total data berhasil diklasifikasikan dengan benar.

b. Rasio 80% Data Training : 20% Data Testing

Pada rasio 80% data training dan 20% data testing, Tabel 12 menampilkan confusion matrix yang menunjukkan performa model sangat baik dalam membedakan data positif dan negatif. Dari 99 data negatif, 65 diklasifikasikan dengan benar dan 34 salah. Sementara itu, dari 146 data positif, 146 dikenali dengan benar dan 0 salah klasifikasi. Secara keseluruhan, model menghasilkan 211 prediksi benar dan 34 salah, dengan akurasi sebesar 86,12%.

Tabel 12. Confussion Matrix Rasio 80:20

True Class	Predicted Class		
	Negatif	Positif	Total
Negatif	65	34	99
Positif	0	146	146
Total	65	137	245

Nilai *recall*, *precision*, dan *accuracy* akan dihitung berdasarkan *confusion matrix* yang telah dibuat. Berikut adalah hasil evaluasinya:

Tabel 13. Matrix Evaluasi Rasio 80:20

Label	Precision	Recall
Negatif	$\frac{65}{65} \times 100\% = 100.00\%$	$\frac{65}{99} \times 100\% = 65.66\%$
Positif	$\frac{132}{137} \times 100\% = 81.11\%$	$\frac{146}{146} \times 100\% = 100.00\%$
Akurasi	$\frac{65 + 146}{245} \times 100\% = \frac{211}{245} \times 100\% = 86.12\%$	

Tabel 13 menunjukkan hasil evaluasi model dengan performa sangat baik. *Precision* tertinggi terdapat pada kelas negatif sebesar 100,00% dan *recall* 65,66%, menandakan mayoritas data negatif dikenali dengan tepat. Pada kelas positif, *precision* mencapai 81,11% dan *recall* 100,00%, menunjukkan hampir seluruh data positif diklasifikasikan dengan benar. Secara keseluruhan, akurasi model sebesar 86,12%, mengindikasikan sekitar 86% data berhasil diklasifikasikan dengan akurat.

4.5. Pengujian Metode Data Testing

Penelitian ini mengevaluasi model klasifikasi sentimen dengan rasio data 70:30 dan 80:20, di mana rasio 70:30 memberikan hasil terbaik. Model diuji pada 100 data baru terkait respons masyarakat terhadap kenaikan PPN 12% di media sosial X. Hasilnya, 73 teks bernada positif dan 27 negatif, menunjukkan mayoritas pengguna merespons positif, meskipun masih ada yang menanggapi negatif. Temuan ini menekankan pentingnya sosialisasi kebijakan yang lebih merata.

4.6. Pengujian Fungsional

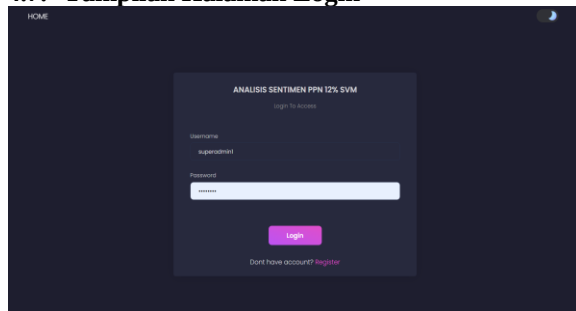
Pengujian Fungsional dilakukan dengan menggunakan pendekatan *Blackbox*.

Tabel 14. Pengujian Fungsional

No.	Fitur	Aktivitas Pengujian	Hasil yang diharapkan	Hasil
1	Login	Login dengan data yang telah terdaftar	Berhasil login dan diarahkan ke halaman Dashboard	Valid
		Login dengan informasi yang tidak terdaftar di sistem.	Login Gagal	Valid
		Login menggunakan kata sandi yang tidak valid.	Login Gagal	Valid
2	Dashboard	Verifikasi data yang ditampilkan pada dashboard sesuai dengan data actual dalam sistem.	Menampilkan data terbaru sesuai dengan data yang baru ditambahkan	Valid
3	Upload Dataset	Unggah file .csv berisi data kolom teks dan label	File berhasil diunggah	Valid
		Unggah file dengan format selain .csv	File Tidak dapat diunggah	Valid
4	Pre-Processing	Klik tombol "Preprocessing" setelah file diunggah	Berhasil di lakukan Pre-Processing	Valid
		Klik Klik tombol "Preprocessing" sebelum file diunggah	Tidak dapat melakukan preprocessing	Valid
		Klik tombol "Download Hasil Preprocessing" setelah melakukan preprocessing	Berhasil Mendownload file hasil pre-processing	Valid

No.	Fitur	Aktivitas Pengujian	Hasil yang diharapkan	Hasil
		Klik tombol "Download Hasil Preprocessing" sebelum melakukan preprocessing	Tidak dapat melakukan download hasil pre-processing	Valid
5	Klasifikasi Training	Klik tombol "Mulai Training" setelah preprocessing selesai tanpa memilih nilai split rasio	Tidak dapat melakukan training	Valid
		Klik tombol "Mulai Training" setelah preprocessing selesai dan memilih nilai split rasio	Berhasil melakukan klasifikasi dengan sesuai dengan rasio yang dipilih	Valid
		Klik tombol "Mulai Training" sebelum mengupload data training	Tidak dapat melakukan training	Valid
		Klik tombol "Mulai Training" sebelum melakukan pre-processing data	Tidak dapat melakukan training	Valid
6	Data Testing	Unggah file .csv berisi data kolom teks	File berhasil diunggah	Valid
		Unggah file dengan format selain .csv	File Tidak dapat diunggah	Valid
		Unggah file .csv berisi data tanpa kolom teks	File tidak dapat diunggah	Valid
		Klik tombol "Submit" setelah preprocessing selesai dan memilih nilai split rasio.	Berhasil dilakukan testing	Valid
		Klik tombol "Hapus Semua Data" setelah proses Testing selesai.	Berhasil menghapus semua data hasil Testing	Valid

4.7. Tampilan Halaman Login



Gambar 8. Halaman Login

Gambar 8 menunjukkan halaman login yang digunakan oleh pengguna yang telah terdaftar melalui halaman *register*. Setelah berhasil login, pengguna dapat mengakses *dashboard* dan melihat hasil analisis sentimen. Khusus untuk *superadmin*, akun ditambahkan secara *manual* melalui program, sedangkan akun admin ditambahkan oleh superadmin melalui halaman kelola pengguna.

4.8. Tampilan Halaman Dashboard

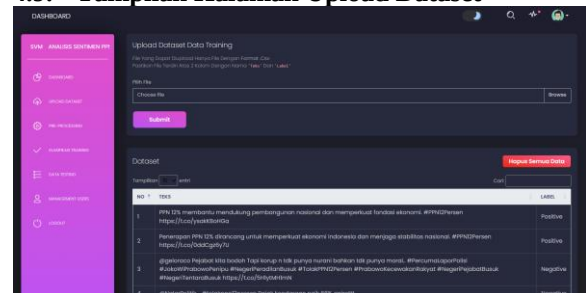


Gambar 9. Halaman Dashboard

Gambar 9 menunjukkan halaman dashboard yang merupakan halaman utama aplikasi. Pada halaman ini terdapat rangkuman dari aplikasi, yaitu berupa total

data yang terdapat pada sistem serta hasil klasifikasi untuk data training dan data testing.

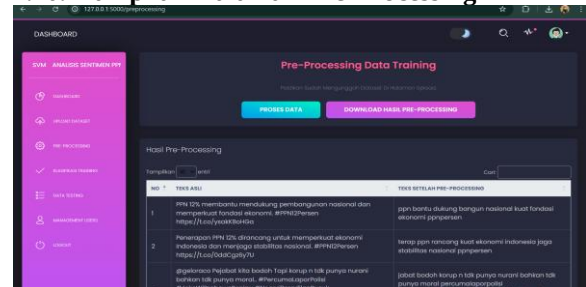
4.9. Tampilan Halaman Upload Dataset



Gambar 10. Halaman Upload Dataset

Gambar 10 menunjukkan halaman upload dataset yang digunakan untuk upload dataset. Data yang dapat di-upload hanya data berformat .csv yang memiliki kolom dengan nama 'teks' dan 'label'. Data pada halaman inilah yang akan digunakan dalam proses pelatihan data pada halaman klasifikasi *training*.

4.10. Tampilan Halaman Pre-Processing

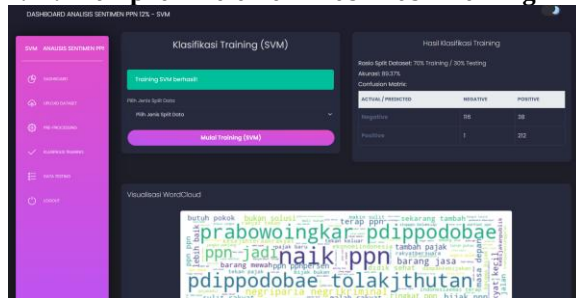


Gambar 11. Halaman Pre-Processing

Gambar 11 adalah representasi visual dari halaman *preprocessing*, di mana data yang telah diunggah akan diproses lebih lanjut. Oleh sebab itu, halaman ini hanya dapat berfungsi jika pengguna telah

melakukan *upload* data *training*. Setelah proses *preprocessing* selesai, pengguna dapat melakukan *download* hasil *preprocessing* dalam format *.xlsx*.

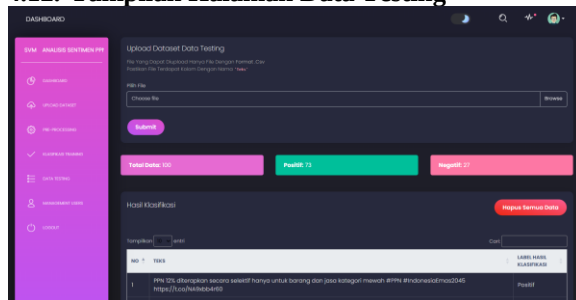
4.11. Tampilan Halaman Klasifikasi Training



Gambar 12. Halaman Klasifikasi Training

Gambar 12 berfungsi untuk melakukan *training* dataset. Halaman ini hanya dapat dijalankan jika terdapat dataset yang telah ter-*upload* dan data tersebut telah dilakukan *pre-processing*. Pada tahapan klasifikasi ini pengguna dapat memilih jenis rasio split data yang akan dilakukan pelatihan oleh model.

4.12. Tampilan Halaman Data Testing



Gambar 13. Halaman Data Testing

Gambar 13 menampilkan halaman pengujian data. Halaman ini memiliki fungsi yang serupa dengan halaman Klasifikasi Training. Namun, pada halaman ini data yang di-*upload* hanya memiliki kolom teks saja tanpa label karena akan difungsikan untuk melakukan testing metode yang sudah dilakukan pelatihan pada halaman klasifikasi *training*.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa model SVM mampu menganalisis sentimen terhadap kebijakan kenaikan PPN 12% dengan akurasi tinggi, yaitu 89,37% (split 70:30) dan 86,12% (split 80:20). Model lebih efektif dalam mengenali sentimen positif, sejalan dengan dominasi respons positif di media sosial X. Meski hasilnya baik, penelitian ini masih terbatas oleh jumlah data dan ketidakseimbangan label. Untuk peningkatan ke depan, disarankan melakukan penambahan data dan menerapkan teknik *balancing* seperti *oversampling* atau *undersampling* guna meningkatkan generalisasi dan mengurangi bias.

DAFTAR PUSTAKA

- [1] Putri, I. M. (2024). Kenaikan Ppn 12% Dan Dampaknya Terhadap Ekonomi. *Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA)*, 8(2), 934-944.
- [2] Irmayani, W. (2024). PERSEPSI PUBLIK TERHADAP KENAIKAN PPN 12%: PENDEKATAN SENTIMEN PADA KOMENTAR YOUTUBE. *Jurnal Khatulistiwa Informatika*, 12(2), 112–118.
- [3] Fauziah, N. (2024). Analisis Sentimen Publik Terhadap Kenaikan Tarif PPN di Indonesia dengan Pendekatan VADER. *Jurnal Akuntansi dan Keuangan*, 12(2), 228-238.
- [4] Kodu, A. R., Heri, H., Hasibuan, R. J. F., & Sunarsih, U. (2025). Pandangan kenaikan tarif PPN 12% dengan menggunakan sistem Coretax: Implikasi pada UMKM di Indonesia. *Jurnal Masharif Al-Syariah: Jurnal Ekonomi dan Perbankan Syariah*, 10(1).
- [5] Fitriyah, N., Warsito, B., & Di Asih, I. M. (2020). Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM). *Jurnal Gaussian*, 9(3), 376-390.
- [6] M. Fitra Rizki, K. Auliasari, dan R. Primaswara Prasetya, "ANALISIS SENTIMENT CYBERBULLYING PADA SOSIAL MEDIA TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," 2021.
- [7] R. D. Yahya, S. A. Wibowo, dan N. Vendyansyah, "ANALISIS SENTIMEN UNTUK DETEKSI UJARAN KEBENCIAN PADA MEDIA SOSIAL TERKAIT PEMILU 2024 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," 2024.
- [8] Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). Analisis sentimen masyarakat terhadap program kartu prakerja pada twitter dengan metode support vector machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150-155.
- [9] Berliani, S., & Lestari, S. (2024). Analisis Sentimen Masyarakat Terhadap Isu Pecat Sri Mulyani Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine. *Jurnal Sains dan Teknologi*, 5(3), 951-960.
- [10] Gemilang, W. S., Purwantoro, P., & Carudin, C. (2023). ANALISIS SENTIMEN PENGGUNA INSTAGRAM PADA CALON PRESIDEN 2024 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2849-2855.
- [11] Arsi, P., & Waluyo, R. (2021). Analisis sentimen wacana pemindahan ibu kota Indonesia menggunakan algoritma Support Vector Machine (SVM). *J. Teknol. Inf. dan Ilmu Komput*, 8(1), 147.

- [12] Sari, F. V., & Wibowo, A. (2019). Analisis sentimen pelanggan toko online Jd. Id menggunakan metode Naïve Bayes Classifier berbasis konversi ikon emosi. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 10(2), 681-686.
- [13] Rahayu, S., & Yamasari, Y. (2024). Klasifikasi Penyakit Stroke dengan Metode Support Vector Machine (SVM). *Journal of Informatics and Computer Science (JINACS)*, 5(03), 440-446.
- [14] M. Mega, M. Olhang, S. Achmadi, dan F. X. Ariwibisono, "ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP COVID-19 DI INDONESIA MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (NBC)," 2020.
- [15] Gemilang, W. S., Purwantoro, P., & Carudin, C. (2023). ANALISIS SENTIMEN PENGGUNA INSTAGRAM PADA CALON PRESIDEN 2024 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2849-2855.