

ANALISIS SENTIMEN MASYARAKAT MENGENAI KEBIJAKAN PEMERINTAH MENAIKKAN PPN 12% PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Yola Anggreni Br Sinulaki¹, Khairul², Suria Alamsyah³
Sistem Informasi, STMIK Kaputama, Palembang
E-mail: [*yolaanggreni2583@gmail.com](mailto:yolaanggreni2583@gmail.com)¹

ABSTRAK

Kebijakan kenaikan Pajak Pertambahan Nilai (PPN) dari 10% menjadi 12% menimbulkan beragam respons masyarakat. Penelitian ini menganalisis sentimen masyarakat pada periode Desember 2024 (± 3000 tweet) dan Januari 2025 (± 3000 tweet) menggunakan metode Support Vector Machine (SVM) dengan pendekatan Term Frequency-Inverse Document Frequency (TF-IDF). Tahapan penelitian mengacu pada Knowledge Discovery in Databases (KDD) yang mencakup pengumpulan data, preprocessing, transformasi, data mining, dan evaluasi hasil, dengan implementasi menggunakan Python di Google Colab. Hasil analisis menunjukkan bahwa pada Desember sentimen didominasi negatif (46,9%), sedangkan pada Januari sentimen netral lebih dominan (42,6%) dengan sedikit peningkatan positif (16,7%). Evaluasi model menghasilkan akurasi 84,87% pada Desember dan 79,33% pada Januari, sehingga SVM terbukti cukup efektif dalam klasifikasi sentimen isu kebijakan publik.

Kata kunci

Analisis Sentimen, PPN 12%, Support vector Machine, TF-IDF, KDD.

ABSTRACT

The government's policy to increase the Value Added Tax (VAT) from 10% to 12% has generated various public responses. This study analyzes public sentiment during December 2024 ($\pm 3,000$ tweets) and January 2025 ($\pm 3,000$ tweets) using the Support vector Machine (SVM) method with the Term Frequency-Inverse Document Frequency (TF-IDF) approach. The research follows the Knowledge Discovery in Databases (KDD) stages, including data collection, preprocessing, Transformation, data mining, and Evaluation, implemented with Python on Google Colab. The results show that in December negative sentiment dominated (46.9%), while in January neutral sentiment became more dominant (42.6%) with a slight increase in positive sentiment (16.7%). The model achieved an Accuracy of 84.87% in December and 79.33% in January, indicating that SVM is effective for sentiment classification in public policy issues.

Keywords

Sentiment Analysis, VAT 12%, Support vector Machine, TF-IDF, KDD.

1. PENDAHULUAN

Pemerintah Indonesia telah menerapkan kenaikan tarif Pajak Pertambahan Nilai (PPN) dari 10% menjadi 12% sejak 1 April 2022, sebagai bagian dari Undang-Undang Harmonisasi Peraturan Perpajakan (UU HPP) Nomor 7 Tahun 2021. Kenaikan ini bertujuan untuk mencapai target penerimaan negara demi pemulihan ekonomi nasional. Selain itu, kebijakan ini juga sejalan dengan agenda transformasi fiskal untuk memperluas basis pajak dan menciptakan sistem perpajakan yang lebih berkeadilan (Pradana 2022).

Kebijakan kenaikan PPN ini menuai beragam tanggapan dari masyarakat, yang terbagi antara yang mendukung karena kontribusinya terhadap pembangunan dan yang menolak karena kekhawatiran akan peningkatan pengeluaran (Kharisma et al. 2023). Perbedaan pandangan ini menjadikan analisis sentimen masyarakat menjadi penting untuk dikaji secara ilmiah. Media sosial, khususnya *Twitter* (sekarang dikenal *X*), merupakan platform yang sangat penting bagi masyarakat untuk menyampaikan

pendapat mereka tentang berbagai isu, termasuk kebijakan pemerintah. Platform ini menyediakan data yang kaya untuk menganalisis sentimen dan opini publik (Cholifah et al. 2023).

Untuk menganalisis data dalam jumlah besar dari *Twitter* (X), diperlukan pendekatan analisis sentimen menggunakan metode pembelajaran mesin (machine learning). Algoritma Support Vector Machine (SVM) telah terbukti efektif dalam klasifikasi teks dan analisis sentimen (Yunita and Kamayani 2023). Algoritma ini mampu mengelompokkan data teks ke dalam kategori sentimen positif, negatif, atau netral dengan tingkat akurasi yang baik. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terkait kebijakan kenaikan PPN 12% dan membandingkan hasil analisis sentimen dari dua periode waktu berbeda (1-31 Desember 2024 dan 1-31 Januari 2025) untuk mengidentifikasi perubahan persepsi masyarakat dengan memanfaatkan data dari *Twitter* (X) dan metode SVM.

2. METODE PENELITIAN

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan menerapkan metode *Support Vector Machine* (SVM) sebagai algoritma analisis data teks. Alur penelitian ini mengacu pada tahapan dalam metode KDD (*Knowledge Discovery in Databases*) yang terdiri dari lima tahap utama, yaitu: *Data Selection*, *Preprocessing*, *Transformation*, *Data mining*, dan *Evaluation*. Untuk menunjang proses analisis data, digunakan Google Colab dan bahasa pemrograman *Python* untuk proses pemodelan dan evaluasi klasifikasi sentimen.

a. *Data Selection*

Data dikumpulkan dari media sosial *Twitter* menggunakan proses *Crawling* dengan bantuan token autentikasi. Data yang diambil berupa *tweet* Masyarakat yang mengandung opini atau tanggapan terhadap kebijakan kenaikan PPN 12%.

b. *Preprocessing*

Data yang telah dikumpulkan kemudian melalui tahapan preprocessing, yang meliputi *case folding*, pembersihan data (*cleaning*), tokenisasi, *stopword removal*, dan *stemming*. Tujuan dari pengolahan data ini adalah untuk menghasilkan data yang siap digunakan sebagai input pada tahap pelabelan sentimen dan pelatihan model klasifikasi SVM.

1) *Case Folding*

Tahapan ini dilakukan proses perubahan huruf kapital menjadi huruf kecil dalam dokumen.

2) *Cleaning*

Tahap ini adalah penghapusan karakter non-abjad untuk mengurangi noise seperti titik (.), koma (,), tanda tanya (?), tanda seru (!), serta simbol seperti tanda “@” untuk nama pengguna, hashtag (#), emotikon, dan alamat situs web)

3) *Tokenizing*

Tokenizing merupakan tahap dimana dilakukan proses pemecahan kalimat menjadi kata yang disebut dengan token.

4) *Stopword Removal*

Stopword Removal dilakukan untuk menghilangkan kata yang tidak relevan terhadap kalimat seperti kata penghubung atau imbuhan.

5) *Stemming*

Mengubah kata menjadi bentuk dasarnya (*root word*) menggunakan *library* Sastrawi untuk Bahasa Indonesia.

c. *Transformation*

Tahap ini dilakukan labeling sentimen dengan pendekatan lexicon, yaitu mencocokkan kata pada tweet ke dalam daftar positif dan negatif untuk menentukan kategori sentimen. Selanjutnya, data berlabel diubah menjadi bentuk numerik menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*), yang memberi bobot pada kata sesuai frekuensinya dalam dokumen relatif terhadap seluruh data. Rumus menghitung TF-IDF, yaitu:

1) TF (*Term Frequency*)

$$TF_{(t,d)} = \frac{f_{t,d}}{N_d}$$

Keterangan:

$f_{t,d}$ = Jumlah kemunculan kata t dalam dokumen d

N_d = Total kata dalam dokumen d

2) DF (*Document Frequency*) – IDF (*Inverse Document Frequency*)

$$IDF_{(t)} = \ln\left(\frac{N}{df_t}\right) + 1$$

Keterangan:

N = Jumlah total dokumen

df_t = Total jumlah dokumen yang mengandung kata t

3) TF-IDF

$$TF - IDF(t, d) = TF(t, d) \times IDF(t)$$

d. *Data mining*

$$f(x) = w \cdot x + b$$

Keterangan:

w = Bobot model yang sudah dilatih

x = Nilai dari fitur ke- i (TF-IDF)

b = Bias

e. *Evaluation*

Dalam hal klasifikasi teks terkadang terdapat kesalahan dalam mengklasifikasinya. Tahapan evaluasi akan menggunakan *Confusion Matrix* untuk mengetahui performansi dari setiap kernel algoritma *Support vector Machine* dengan hasil evaluasi berupa nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Rumus untuk menghitungnya:

1) *Accuracy*

$$accuracy = \frac{TP}{TP + TN + FP + FN}$$

Keterangan:

a) TP = *True Positive*

b) TN = *True Negative*

c) FP = *False Positive*

d) FN = *False Negative*

2) *Precision*

$$precision = \frac{TP}{TP + FP}$$

Keterangan:

a) TP (*True Positive*): jumlah data yang diprediksi positif dan memang benar positif

b) FP (*False Positive*): jumlah data yang diprediksi positif namun ternyata salah (negatif)

3) *Recall*

$$recall = \frac{TP}{TP + FN}$$

Keterangan:

a) TP (*True Positive*): jumlah data yang diprediksi positif dan memang benar positif

b) FN (*False Negative*): jumlah data yang seharusnya positif tetapi diprediksi sebagai negatif

4) *F1-Score*

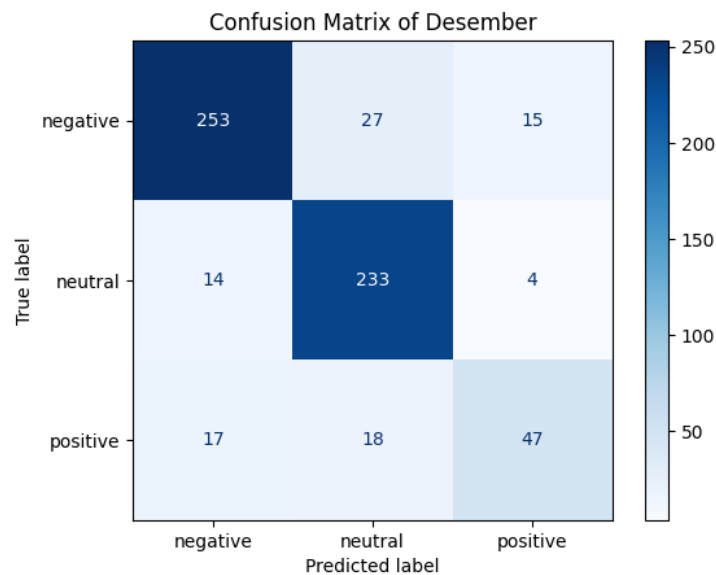
$$F1 - Score = 2x \frac{Precision \times Recall}{Precision + Recall}$$

3. HASIL DAN PEMBAHASAN

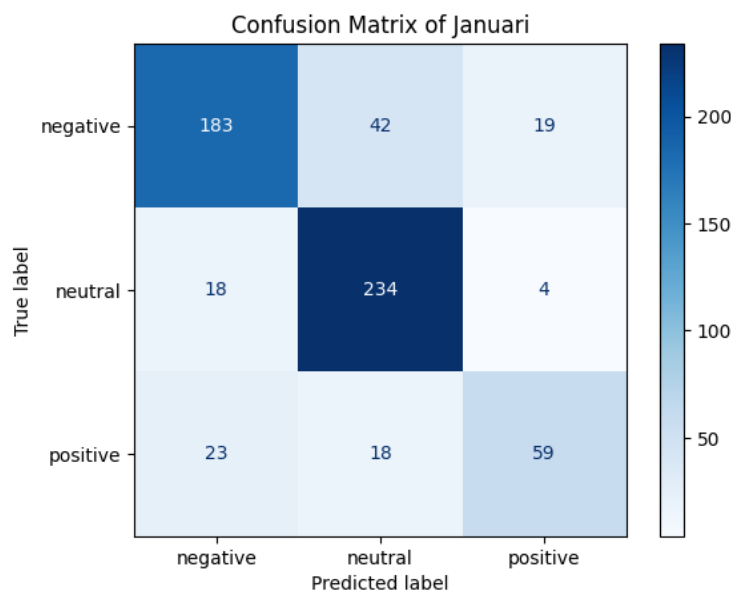
Hasil analisis menunjukkan adanya perbedaan distribusi sentimen antara Desember dan Januari. Pada Desember, sentimen negatif mendominasi (46,9%), sedangkan pada Januari sentimen negatif menurun (40,7%) dengan peningkatan sentimen netral (42,6%) dan positif (16,7%). Hal ini mengindikasikan bahwa respon masyarakat terhadap kebijakan PPN 12% cenderung bergeser dari dominasi negatif menuju opini yang lebih netral dengan sedikit peningkatan dukungan positif.

a. Confusion Matrix

Pada data Desember 2024, model paling baik mengenali sentimen netral dan negatif, sedangkan sentimen positif masih banyak salah terklasifikasi ke negatif maupun netral. Hasil serupa terlihat pada Januari 2025, di mana sentimen netral tetap paling akurat, sementara sentimen negatif sering tertukar dengan netral. Secara keseluruhan, model lebih mudah membedakan sentimen netral dan negatif, namun kesulitan pada sentimen positif karena jumlah datanya lebih sedikit.



Gambar 1. Confusion Matrix Desember



Gambar 2. Confusion Matrix Desember

b. Akurasi

Hasil pengujian metode Support Vector Machine (SVM) menunjukkan bahwa pada dataset Desember 2024, model memperoleh akurasi 84,87%, dengan sentimen negatif mendominasi (F1-Score 0,87), sentimen netral menunjukkan stabilitas terbaik (Recall 0,93; F1-Score 0,88), sementara sentimen positif menjadi kategori terendah (F1-Score 0,64). Pada dataset Januari 2025, akurasi menurun menjadi 79,33%, dengan pergeseran dominasi ke sentimen netral yang memiliki kinerja stabil (Recall 0,91; F1-Score 0,85). Sentimen negatif mengalami penurunan performa (F1-Score 0,78), dan sentimen positif tetap dengan nilai terendah (F1-Score 0,65).

Akurasi: 0.79
Akurasi: 79.33%

Laporan Klasifikasi:

	precision	recall	f1-score	support
negative	0.82	0.75	0.78	244
neutral	0.80	0.91	0.85	256
positive	0.72	0.59	0.65	100
accuracy			0.79	600
macro avg	0.78	0.75	0.76	600
weighted avg	0.79	0.79	0.79	600

Gambar 3. Akurasi, Precision, Recall dan F1-Score Desember

Akurasi: 1.00
Akurasi: 100.00%

Laporan Klasifikasi:

	precision	recall	f1-score	support
negative	1.00	1.00	1.00	130
positive	1.00	1.00	1.00	30
accuracy			1.00	160
macro avg	1.00	1.00	1.00	160
weighted avg	1.00	1.00	1.00	160

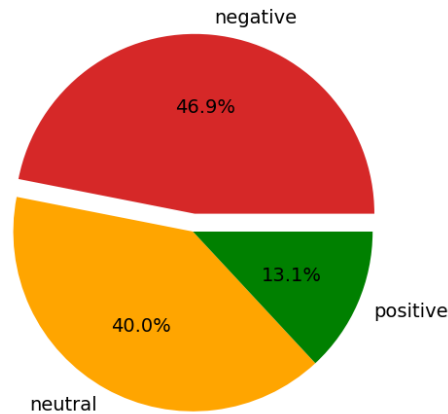
Gambar 4. Akurasi, Precision, Recall dan F1-Score Januari

c. Distribusi Sentimen

Berdasarkan hasil penelitian analisis sentimen masyarakat terhadap kebijakan kenaikan PPN 12% dengan metode Support Vector Machine (SVM), diperoleh beberapa kesimpulan. Pertama, sentimen negatif masih mendominasi pada periode Desember 2024 dan Januari 2025, yang menunjukkan mayoritas masyarakat kurang menerima

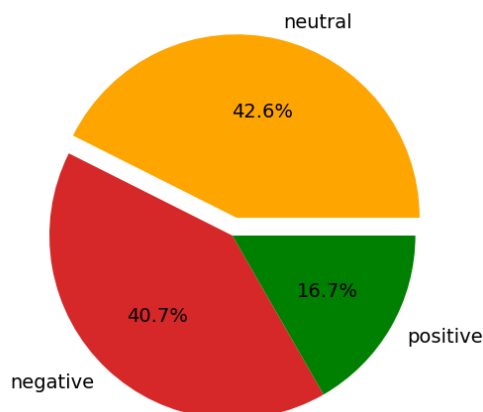
kebijakan tersebut. Kedua, terdapat peningkatan sentimen positif dan netral pada Januari 2025, yang mengindikasikan adanya proses penyesuaian masyarakat. Ketiga, penerapan SVM dengan pendekatan TF-IDF terbukti efektif dalam mengklasifikasikan sentimen, sehingga dapat digunakan untuk menganalisis opini publik terhadap kebijakan ekonomi.

Sentiment Polarity Desember on Tweets Data
(total = ± 3000 tweets)



Gambar 5. Distribusi Sentimen Desember

Sentiment Polarity Januari on Tweets Data
(total = ± 3000 tweets)



Gambar 6. Distribusi Sentimen Januari

d. Word Cloud

Pada Desember 2024, *Word Cloud* didominasi kata “tolak”, “naik”, “persen”, dan “rakyat”, menunjukkan penolakan kuat terhadap kebijakan PPN 12%. Kata-kata bernada negatif seperti “tidak”, “demo”, dan “harga” juga banyak muncul sehingga sentimen masyarakat cenderung kritis. Pada Januari 2025, kata “tolak” masih ada namun tidak sebesar sebelumnya, sementara kata “sudah”, “sebuah”, dan “pajak” lebih menonjol. Hal ini menandakan pembicaraan masyarakat mulai bergeser dari dominasi penolakan menuju diskusi yang lebih netral dan umum terkait pajak.



Gambar 7. *Word Cloud* Desember



Gambar 8. *Word Cloud* Januari



Gambar 9. *Word Cloud* Positif dan Negatif Desember



Gambar 9. *Word Cloud* Positif dan Negatif Desember



Gambar 10. *Word Cloud* Positif dan Negatif Januari



4. KESIMPULAN

Berdasarkan hasil penelitian, sentimen masyarakat terhadap kebijakan kenaikan PPN 12% pada Desember 2024 dan Januari 2025 didominasi oleh sentimen negatif, menunjukkan kurangnya penerimaan terhadap kebijakan tersebut. Namun, terjadi sedikit peningkatan sentimen positif dan netral pada Januari 2025 yang mengindikasikan adanya proses penyesuaian. Metode SVM dengan TF-IDF terbukti efektif dalam mengklasifikasikan sentimen, sehingga dapat digunakan untuk menganalisis opini publik terkait kebijakan ekonomi.

5. DAFTAR PUSTAKA

- Cholifah, Chorirotun, Hanny Hikmayanti Handayani, and Ayu Ratna Juwita. 2023. "Analisis Sentimen Twitter Terhadap Uu Omnibus Law Menggunakan Metode Support Vector Machine (Svm) Dan Naïve Bayes Classifier (Nbc)." *Jurnal Informatika Teknologi dan Sains* 4(4): 483–88. doi:10.51401/jinteks.v4i4.2191.
- Kharisma, Nurul, Imahda Khorir Furqon, Uin K H Abdurrahman, and Wahid Pekalongan. 2023. "Analisis Dampak Kenaikan Tarif Pajak Pertambahan Nilai (PPN) Terhadap Masyarakat Dan Inflasi Di Indonesia." *Jurnal Sahmiyya* 2: 295–303.
- Pradana, Rifky Zahran. 2022. "Problematisasi Pemulihan Ekonomi Nasional Pasca Pandemi Dengan Pemberlakuan Kenaikan Tarif Ppn." *Japhtn-Han* 1(2). doi:10.55292/japhtnhan.v1i2.33.
- Yunita, Rani, and Mia Kamayani. 2023. "Perbandingan Algoritma SVM Dan Naïve Bayes Pada Analisis Sentimen Penghapusan Kewajiban Skripsi." *Indonesian Journal of Computer Science*. doi:10.33022/ijcs.v12i5.3415.