

# Komparasi Algoritma Support Vector Machine dan Decision Tree Dalam Analisis Sentimen Publik Terhadap Penerapan PPN 12%

Djalu Bintang Putra<sup>1\*</sup>, Ryan Randy Suryono<sup>2</sup>

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: <sup>1\*</sup>djalu\_bintang\_putra@teknokrat.ac.id, <sup>2</sup>@ryan@teknokrat.ac.id

Email Penulis Korespondensi: djalu\_bintang\_putra@teknokrat.ac.id

Submitted: 20/03/2025; Accepted: 04/06/2025; Published: 06/06/2025

**Abstrak**—Penerapan kebijakan Pajak Pertambahan Nilai (PPN) 12% di Indonesia menimbulkan berbagai reaksi dari masyarakat, baik yang positif maupun negatif. Untuk memahami pandangan publik, peneliti membandingkan kinerja dari dua algoritma, yaitu Support Vector Machine (SVM) dan Decision Tree, dalam menganalisis sentimen di media sosial. Data sebanyak 7.965 tweet dari platform X (Twitter) dikumpulkan menggunakan teknik web scraping dan diproses melalui beberapa tahap seperti pembersihan teks, tokenisasi, penghapusan kata yang tidak penting, stemming, serta penyeimbangan data dengan metode SMOTE. Selain itu, untuk meningkatkan akurasi model. Hasil evaluasi penelitian didapatkan hasil SVM memiliki akurasi 80% lebih tinggi dibandingkan Decision Tree yang hanya mencapai 68%. Dari hasil tersebut, dapat disimpulkan bahwa SVM lebih efektif dalam menganalisis sentimen publik terkait kebijakan PPN 12%. Temuan ini dapat menjadi acuan bagi pemerintah dan pihak terkait untuk memahami pandangan masyarakat dan merancang kebijakan yang lebih sesuai. Penelitian ini juga memberikan peluang untuk pengembangan lebih lanjut dengan mencoba algoritma lain atau teknik pemrosesan data yang lebih canggih agar analisis sentimen menjadi lebih akurat dan efektif di masa depan.

**Kata Kunci:** Analisis Sentimen; PPN; *Support Vector Machine*; *Decision Tree*; *SMOTE*

**Abstract**—The implementation of the 12% Value-Added Tax (VAT) policy in Indonesia has generated various reactions from the public, both positive and negative. To understand public perception, researchers compared the performance of two algorithms, namely Support Vector Machine (SVM) and Decision Tree, in analyzing sentiment on social media. A total of 7,965 tweets were collected from the X (Twitter) platform using web scraping techniques and processed through several stages, including text cleaning, tokenization, stopword removal, stemming, and data balancing using the SMOTE method to improve model accuracy. The evaluation results showed that SVM achieved 80% accuracy, higher than Decision Tree, which only reached 68%. Based on these findings, it can be concluded that SVM is more effective in analyzing public sentiment regarding the 12% VAT policy. These findings can serve as a reference for the government and relevant stakeholders to better understand public opinion and design more suitable policies. This study also provides opportunities for further development by exploring other algorithms or more advanced data processing techniques to enhance the accuracy and effectiveness of sentiment analysis in the future.

**Keywords:** Sentiment Analysis; VAT; Support Vector Machine; Decision Tree; SMOTE

## 1. PENDAHULUAN

Salah satu sumber utama penerimaan negara adalah pajak pertambahan nilai (PPN), yang dikenakan atas konsumsi Barang Kena Pajak (BKP) dan Jasa Kena Pajak (JKP) nasional. Berdasarkan Undang-Undang Nomor 7 Tahun 2021, tarif PPN yang sebelumnya sebesar 10% naik secara bertahap hingga mencapai 12%. Tujuan dari kebijakan ini adalah untuk meningkatkan penerimaan negara dan menyesuaikan sistem perpajakan dengan standar yang lebih kompetitif di seluruh dunia [1]. Namun, pelaksanaan kebijakan ini diperkirakan memiliki dampak yang signifikan terhadap berbagai sektor ekonomi, termasuk industri, perdagangan, dan konsumsi masyarakat. Oleh karena itu, penting untuk memahami bagaimana masyarakat menanggapi kebijakan ini, terutama di era digital yang semakin berkembang. Seiring dengan meningkatnya penggunaan internet, media sosial telah menjadi salah satu sarana utama bagi masyarakat untuk mengekspresikan opini dan pandangan mereka terhadap isu-isu yang sedang berkembang. Pada tahun 2024, jumlah pengguna internet di Indonesia diperkirakan mencapai 221.563.479 jiwa dari total populasi 278.696.200 jiwa pada tahun 2023. Berdasarkan survei penetrasi internet Indonesia 2024, tingkat penetrasi internet telah mencapai 79,5%, meningkat sebesar 1,4% dibandingkan periode sebelumnya. Media sosial seperti WhatsApp, Telegram, X (sebelumnya Twitter), Facebook, dan Instagram memainkan peran penting dalam memfasilitasi diskusi publik mengenai berbagai isu, termasuk kebijakan ekonomi [2].

Di antara berbagai platform media sosial, X menjadi salah satu wadah utama bagi masyarakat dalam menyampaikan pendapat mereka terhadap isu-isu yang sedang trending. Fitur seperti hashtag (#) dan pencarian kata kunci memungkinkan pengguna untuk mengikuti diskusi secara real-time, sehingga opini publik dapat dengan cepat terbentuk dan tersebar luas. Dengan adanya kemudahan dalam menyampaikan pendapat di media sosial, analisis sentimen terhadap kebijakan PPN 12% menjadi relevan untuk dilakukan guna memahami persepsi masyarakat. Isu kenaikan Pajak Pertambahan Nilai (PPN) menjadi topik yang ramai diperbincangkan di berbagai platform media sosial, salah satunya X. Kebijakan ini menimbulkan beragam opini dan sentimen dari masyarakat. Di satu sisi, peningkatan tarif PPN diharapkan dapat meningkatkan pendapatan negara untuk mendukung pembangunan nasional. Namun, di sisi lain, kebijakan tersebut berpotensi menekan daya beli masyarakat serta memengaruhi harga barang dan jasa di pasar domestik. Peneliti bermaksud untuk melakukan analisis sentiment terkait isu tersebut. Analisis sentimen merupakan proses ekstraksi, pengolahan, dan pemahaman data teks yang tidak terstruktur secara otomatis untuk

mengidentifikasi serta memperoleh informasi mengenai sentimen yang terkandung dalam suatu pernyataan, pendapat, atau opini [3].

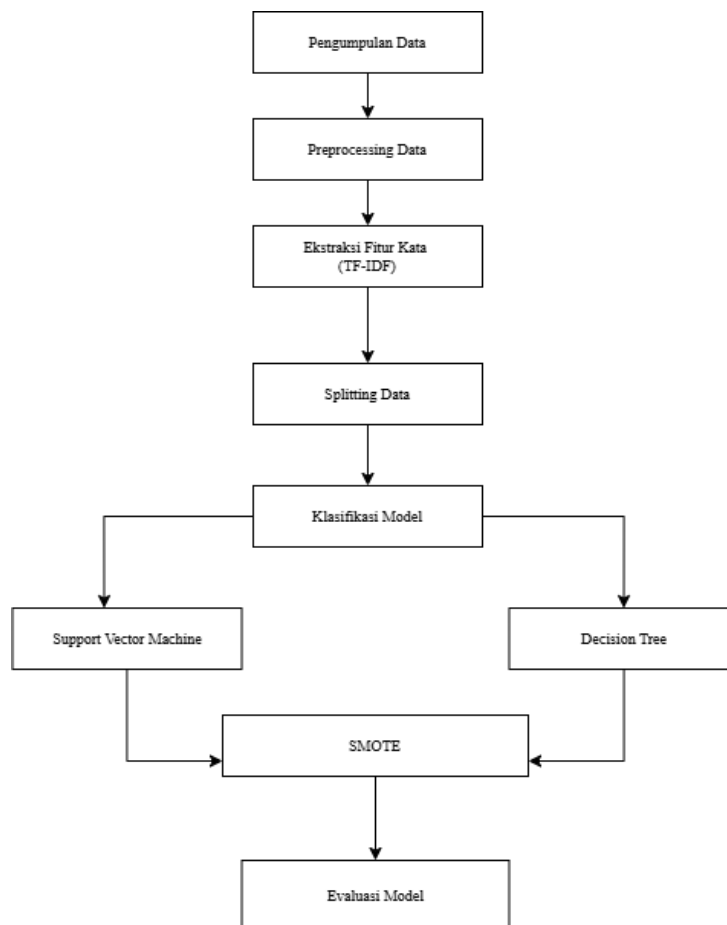
Dalam beberapa penelitian sebelumnya, algoritma *Support Vector Machine* (SVM) telah digunakan secara luas dalam analisis sentimen di berbagai bidang. Salah satu studi yang membahas analisis sentimen terhadap wacana pemindahan ibu kota Indonesia menggunakan algoritma SVM dengan metode pemberian bobot *Term Frequency-Inverse Document Frequency* (TF-IDF). Hasil penelitian tersebut menunjukkan bahwa dari 1.236 data yang terdiri atas 404 data sentimen positif dan 832 data sentimen negatif, diperoleh akurasi sebesar 96,68%, presisi 95,82%, recall 94,04%, dan *Area Under the Curve* (AUC) sebesar 0,979 [4]. Selain itu, penelitian lain yang menganalisis sentimen masyarakat terhadap maskapai penerbangan di platform X (sebelumnya Twitter) juga menggunakan algoritma SVM dengan metode tuning parameter. Hasil penelitian ini menunjukkan akurasi sebesar 84,37%, sementara penggunaan metode *10-fold cross-validation* menghasilkan akurasi sebesar 80,41% [5]. Sebuah penelitian yang menganalisis respons masyarakat terhadap kebijakan Pembatasan Sosial Berskala Besar (PSBB) selama pandemi *covid-19* menggunakan perbandingan tiga algoritma klasifikasi, yaitu *Decision Tree*, *Naïve Bayes*, dan *K-Nearest Neighbors* (KNN), menemukan bahwa *Decision Tree* memberikan performa terbaik. Algoritma ini mencapai tingkat akurasi sebesar 84,78%, precision 84,78%, dan recall 100% [6]. Penelitian lain yang menerapkan algoritma *Decision Tree* dalam analisis sentimen pada marketplace di Indonesia menggunakan metode *cross validation* menunjukkan akurasi sebesar 70,27%, sementara metode *split validation* menghasilkan akurasi 66,95%. Berdasarkan temuan tersebut, metode *cross validation* dianggap sebagai salah satu pendekatan yang efektif dalam penerapan algoritma *Decision Tree* [7]. Lebih lanjut, penelitian yang membandingkan kinerja algoritma SVM dan *Decision Tree* dalam analisis sentimen terhadap program Merdeka Belajar Kampus Merdeka (MBKM) menunjukkan bahwa SVM memiliki akurasi lebih tinggi, yakni 84,76%, dibandingkan dengan *Decision Tree* yang hanya memperoleh akurasi 72,86% [8]. Selain itu, studi lain yang membahas analisis sentimen terhadap program Kartu Prakerja menggunakan SVM dengan dua jenis kernel, yaitu *linear* dan *RBF*, menunjukkan hasil evaluasi yang beragam. Kernel *linear* menghasilkan akurasi 96,67%, precision 98%, recall 99%, dan *F1-score* 98%, sedangkan kernel *RBF* menunjukkan performa lebih tinggi dengan akurasi 98,34%, precision 97%, recall 98%, dan *F1-score* 98% [9]. Studi lain yang membandingkan algoritma SVM dan *Naïve Bayes* menggunakan dataset berjumlah 2.852 data menunjukkan bahwa SVM memiliki akurasi 98% sebelum balancing dan turun menjadi 97% setelah balancing. Sementara itu, *Naïve Bayes* mencatat akurasi 97% sebelum balancing dan turun menjadi 90% setelah balancing, mengindikasikan bahwa SVM lebih unggul dalam klasifikasi sentimen [10].

Berdasarkan penelitian-penelitian sebelumnya, algoritma SVM dan *Decision Tree* terbukti memiliki performa yang baik dalam analisis sentimen, meskipun hasilnya dapat bervariasi tergantung pada metode pengolahan data yang digunakan. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja algoritma SVM dengan *Decision Tree* dalam menganalisis sentimen masyarakat terhadap kebijakan kenaikan PPN 12% di media sosial X. Melalui perbandingan ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai efektivitas kedua algoritma serta model yang lebih optimal dalam mengolah opini publik terkait kebijakan ekonomi. Hasil penelitian ini diharapkan dapat memberikan wawasan bagi pemangku kebijakan dalam memahami persepsi masyarakat terhadap kebijakan fiskal, khususnya dalam penerapan PPN

## 2. METODOLOGI PENELITIAN

### 2.1 Metode Penelitian

Dalam penelitian ini, pendekatan yang digunakan melibatkan pengumpulan komentar publik dari media sosial Twitter/X dengan menggunakan kata kunci PPN 12%, yang menghasilkan total 7.996 data. Selanjutnya, dilakukan proses preprocessing dan pembersihan data (*data cleaning*) untuk menghilangkan elemen-elemen yang tidak relevan (*noise*) seperti tanda baca, angka, *URL*, dan *stopwords*, serta mengubah semua teks menjadi format huruf kecil (*lowercase*). Proses ini bertujuan untuk mempersiapkan data agar lebih terstruktur dan siap untuk tahap pelabelan sentimen. Setelah proses preprocessing selesai, data diklasifikasikan ke dalam tiga kategori sentimen utama, yaitu positif, negatif, dan netral berdasarkan konteks dan makna dari setiap komentar. Tahap analisis sentimen kemudian dilakukan dengan menerapkan dua metode klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Decision Tree* (DT), untuk mengevaluasi performa masing-masing model dalam mengklasifikasikan data komentar. Untuk mengatasi ketidakseimbangan distribusi kelas dalam dataset, diterapkan teknik SMOTE (*Synthetic Minority Over-sampling Technique*). Teknik ini secara sintesis menambah data pada kelas minoritas agar distribusi kelas menjadi lebih seimbang, sehingga model dapat melakukan klasifikasi dengan lebih adil dan akurat. Evaluasi performa dilakukan sebelum dan sesudah penerapan SMOTE dengan membandingkan metrik akurasi, *precision*, *recall*, dan *F1-score*, serta visualisasi melalui confusion matrix. Alur penelitian dapat dilihat pada Gambar 1.



**Gambar 1.** Tahapan Penelitian.

## 2.2 Pengumpulan Data

Tahap pengumpulan data merupakan langkah awal yang sangat penting dalam proses text mining. Pada tahap ini, data berbentuk teks dikumpulkan dari berbagai sumber sebagai bahan dasar untuk dianalisis lebih lanjut guna mendapatkan informasi dan pengetahuan baru. Dalam penelitian ini, proses pengumpulan data dilakukan melalui platform media sosial Twitter, di mana data diekstraksi menggunakan *Application Programming Interface* (API) yang disediakan oleh Twitter [11]. Selain itu, proses ekstraksi juga didukung oleh pemanfaatan pustaka pemrograman seperti *Tweepy* atau pustaka sejenis yang berfungsi sebagai alat bantu dalam pengambilan data secara otomatis. Tahapan selanjutnya adalah proses preprocessing data untuk dilakukan pembersihan data yang meliputi karakter, dan kata-kata yang tidak diperlukan.

## 2.3 Preprocessing Data

*Preprocessing* data merupakan tahap awal yang berfungsi untuk melakukan pembersihan dan penyiapan data mentah sebelum data tersebut diproses lebih lanjut dalam kegiatan analisis, seperti *text mining* maupun penerapan *algoritma machine learning*. Khusus dalam analisis teks, proses preprocessing memiliki tujuan utama yaitu menghapus unsur-unsur yang tidak memiliki makna signifikan, seperti simbol, angka, dan kata-kata yang sering muncul namun tidak relevan. Selain itu, preprocessing juga menyelaraskan format teks dan menyederhanakan struktur kalimat agar data lebih terorganisir dan siap digunakan oleh algoritma analisis [12].

## 2.4 Pelabelan Data

Pelabelan data merujuk pada proses penetapan kategori atau label pada data mentah (*raw data*), yang bertujuan untuk memberi informasi tentang kelas atau kategori yang dimiliki oleh data tersebut. Dalam konteks text mining dan analisis sentimen, pelabelan data umumnya dilakukan dengan mengidentifikasi apakah teks termasuk dalam kategori *positif*, *negatif*, atau *netral*. Proses pelabelan ini sangat krusial karena algoritma dalam *supervised learning* memerlukan data yang sudah diberi label untuk melatih model, sehingga model tersebut dapat melakukan klasifikasi data baru dengan akurat [13].

## 2.5 Pembagian Data

Pembagian data atau *splitting data* adalah langkah untuk membagi dataset menjadi beberapa bagian, yang umumnya dilakukan dalam analisis data dan machine learning. Biasanya, data dibagi menjadi dua bagian utama, yaitu data latih

(*training data*) dan data uji (*testing data*). Tujuan utama dari proses ini adalah untuk melatih model dengan menggunakan data latih dan mengujinya menggunakan data uji, sehingga model dapat menunjukkan kemampuannya dalam menggeneralisasi terhadap data baru yang belum pernah dilihat sebelumnya, bukan hanya menghafal data latih. Pembagian data sangat penting untuk mencegah *overfitting*, di mana model terlalu menyesuaikan diri dengan data latih dan tidak dapat bekerja dengan baik pada data uji yang berbeda. Teknik pembagian data yang umum digunakan adalah *train-test split* dengan rasio tertentu (seperti 80:20 atau 70:30) atau dengan menerapkan *cross-validation* untuk memperoleh hasil yang lebih konsisten dan dapat diandalkan [14].

## 2.6 Fitur Ekstraksi (TF-IDF).

Tujuan dari TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah untuk menilai kata-kata yang paling relevan dalam suatu dokumen dan membedakannya dari dokumen lain dalam kumpulan data. Metode ini menggabungkan dua komponen utama perhitungan bobot: frekuensi kata yang muncul dalam dokumen tertentu dan frekuensi kebalikan dari dokumen yang mengandung kata tersebut [15].

## 2.7 Penerapan Algoritma SVM dan Decision Tree

Untuk klasifikasi sentimen teks, algoritma *Support Vector Machine* (SVM) dan *Decision Tree* digunakan karena keduanya memiliki kemampuan untuk menangani data dengan struktur yang kompleks dan tidak linear. SVM memisahkan kelas sentimen dengan menemukan batas pemisah optimal, sementara *Decision Tree* membangun model yang lebih mudah dipahami dengan menggunakan struktur cabang keputusan. Penggunaan kedua algoritma ini dimulai dengan pemrosesan data teks yang telah dibersihkan. Kemudian, teknik TF-IDF digunakan untuk mengubah data menjadi vektor numerik. Selanjutnya, data dibagi menjadi dua bagian, data pelatihan ( $X_{train}$  dan  $y_{train}$ ) dan data pengujian ( $X_{test}$  dan  $y_{test}$ ). Kedua bagian ini digunakan untuk mengevaluasi ketepatan model dalam mengklasifikasikan sentimen yang berkaitan dengan penerapan PPN 12% di Indonesia.

## 2.8 Penyeimbangan Data (SMOTE)

SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan salah satu metode dalam pembelajaran mesin yang digunakan untuk mengatasi masalah ketidakseimbangan kelas dalam dataset. Ketidakseimbangan ini sering terjadi ketika jumlah data pada kelas minoritas secara signifikan lebih rendah dibandingkan dengan kelas mayoritas. Dalam penelitian ini, SMOTE bekerja dengan cara meningkatkan jumlah sampel pada kelas minoritas melalui pembuatan data sintesis baru, sehingga distribusi kelas dalam dataset menjadi lebih seimbang. Teknik ini bertujuan untuk memperbaiki performa model yang dapat terpengaruh oleh ketidakseimbangan kelas, khususnya dalam klasifikasi [16].

## 2.9 Evaluasi Model

Pada tahap ini, penelitian menyajikan hasil dari seluruh proses yang telah dilakukan, termasuk implementasi algoritma *Support Vector Machine* (SVM) dan *Decision Tree* dalam analisis sentimen terhadap penerapan PPN 12% di Indonesia. Analisis ini bertujuan untuk memperoleh temuan serta pemahaman yang lebih mendalam mengenai persepsi publik terkait kebijakan tersebut. Evaluasi kinerja model dilakukan dengan menggunakan empat metrik utama, yaitu akurasi, presisi, *recall*, dan *f1-score*. Hasil yang diperoleh dari masing-masing algoritma akan dibandingkan untuk menentukan metode yang lebih optimal dalam mengklasifikasikan sentimen masyarakat terhadap penerapan kebijakan PPN 12%.

# 3. HASIL DAN PEMBAHASAN

## 3.1 Pengumpulan Data

Penelitian ini menggunakan data sekunder dari Twitter yang dikumpulkan dengan *Python* melalui Twitter API, menggunakan key dan access token untuk autentikasi. Data dikumpulkan dengan kata kunci “PPN 12%” dalam periode sekitar tiga bulan, mencakup sebelum dan sesudah kebijakan diberlakukan, guna menganalisis perubahan sentimen publik. Proses preprocessing mencakup *cleaning*, tokenisasi, penghapusan *stopword*, dan *stemming*, untuk memastikan kualitas data sebelum dianalisis menggunakan algoritma *Support Vector Machine* (SVM) dan *Decision Tree*. Total 7.965 tweet berhasil dikumpulkan dan disimpan dalam format *.csv*.

## 3.2 Preprocessing Data

*Preprocessing* adalah tahapan penting yang harus dilakukan sebelum dapat mengimplementasikan algoritma klasifikasi pada dataset. Tahapan dalam *preprocessing* adalah sebagai berikut :

### a. *Cleaning* Dataset

*Cleaning* dataset adalah proses pembersihan data yang bertujuan untuk menghilangkan atau memperbaiki data yang kotor, tidak relevan, atau tidak konsisten dalam sebuah dataset. Proses ini mencakup beberapa tahapan, seperti menghapus data duplikat, menghapus karakter khusus, memperbaiki kesalahan ejaan, serta menghapus data kosong atau data yang tidak sesuai dengan kebutuhan analisis. *Cleaning* dataset sangat penting dilakukan untuk



meningkatkan kualitas data sehingga analisis yang dilakukan menjadi lebih akurat dan dapat diandalkan [17]. Hasil data Cleaning dapat dilihat pada Tabel 1.

**Tabel 1.** Contoh Hasil Cleaning Data

Data Mentah	Cleaning Data
@MurtadhaOne1 Waktu produksi kena PPN 12 % setelah sampai distributor kena PPN 12 % sampai di jual ke retail sidomart dan salfamart kena lagi PPN 12 % Bener gk sih itunganya ini... jadi berapa kali kena PPN..???	Waktu produksi kena PPN setelah sampai distributor kena PPN sampai di jual ke retail sidomart dan salfamart kena lagi PPN Bener gk sih itunganya ini jadi berapa kali kena PPN
@PaltiWest2024 Karena pengumuman mendadak diakhir tahun sementara yg dibawah sudah siap dgn PPN 12%. Tapi tenang yg sudah kena PPN 12% akan dikembalikan demikian kata DJP. Caranya gimana? Au ah elap.	Karena pengumuman mendadak diakhir tahun sementara yg dibawah sudah siap dgn PPN Tapi tenang yg sudah kena PPN akan dikembalikan demikian kata DJP Caranya gimana Au ah elap
@KompasTV PPN 12% hanya berlaku untuk barang dan jasa mewah demi menjaga keadilan ekonomi dan daya beli rakyat.	PPN hanya berlaku untuk barang dan jasa mewah demi menjaga keadilan ekonomi dan daya beli rakyat
Investasi untuk Masa Depan: PPN 12% Demi Fasilitas Publik Lebih Baik. #PPNberkeadilan	Investasi untuk Masa Depan PPN Demi Fasilitas Publik Lebih Baik PPNberkeadilan

b. *Casefolding*

*Casefolding* adalah proses mengubah semua huruf dalam teks menjadi huruf kecil (lowercase) sebagai bagian dari tahap preprocessing dalam pengolahan data teks. Tujuan utama *casefolding* adalah untuk menyamakan bentuk kata yang sebenarnya memiliki makna yang sama, tetapi ditulis dengan variasi huruf besar dan kecil, misalnya kata "PPN" dan "ppn" akan diperlakukan sama setelah proses *casefolding*. Langkah ini penting untuk menjaga konsistensi data sehingga dapat mengurangi kompleksitas data yang diproses dalam analisis lebih lanjut, seperti text mining atau analisis sentimen. Hasil dari *casefolding* ditunjukkan pada Tabel 2.

**Tabel 2.** Hasil *Casefolding*

Hasil Cleaning	Casefolding
Waktu produksi kena PPN setelah sampai distributor kena PPN sampai di jual ke retail sidomart dan salfamart kena lagi PPN Bener gk sih itunganya ini jadi berapa kali kena PPN	waktu produksi kena ppn setelah sampai distributor kena ppn sampai di jual ke retail sidomart dan salfamart kena lagi ppn bener gk sih itunganya ini jadi berapa kali kena ppn
Karena pengumuman mendadak diakhir tahun sementara yg dibawah sudah siap dgn PPN Tapi tenang yg sudah kena PPN akan dikembalikan demikian kata DJP Caranya gimana Au ah elap	karena pengumuman mendadak diakhir tahun sementara yg dibawah sudah siap dgn ppn tapi tenang yg sudah kena ppn akan dikembalikan demikian kata djp caranya gimana au ah elap
PPN hanya berlaku untuk barang dan jasa mewah demi menjaga keadilan ekonomi dan daya beli rakyat	ppn hanya berlaku untuk barang dan jasa mewah demi menjaga keadilan ekonomi dan daya beli rakyat
Investasi untuk Masa Depan PPN Demi Fasilitas Publik Lebih Baik PPNberkeadilan	investasi untuk masa depan ppn demi fasilitas publik lebih baik ppnberkeadilan

c. Normalisasi

Normalisasi kata merupakan proses standarisasi teks dengan menyelaraskan kata-kata tidak baku atau memiliki variasi penulisan ke dalam bentuk yang sesuai dengan kaidah bahasa. Dalam *text mining*, normalisasi berperan dalam meningkatkan kualitas data teks dengan mengurangi perbedaan penulisan yang memiliki makna serupa. Proses ini sangat penting dalam pemrosesan bahasa alami (NLP) guna meningkatkan ketepatan dalam berbagai tugas analisis teks, seperti klasifikasi sentimen, pengelompokan teks, dan ekstraksi informasi [18]. Beberapa metode yang digunakan dalam normalisasi kata meliputi:

1. Penghapusan karakter khusus: Menghilangkan tanda baca atau simbol yang tidak relevan.
2. Koreksi ejaan: Memperbaiki kesalahan penulisan agar sesuai dengan bentuk baku.
3. Transformasi kata tidak baku: Mengonversi kata slang atau tidak resmi menjadi bentuk standar.
4. Ekspansi singkatan atau akronim: Mengubah bentuk singkatan menjadi versi lengkapnya.

Hasil dari normalisasi data ditunjukkan pada Tabel 3.

**Tabel 3.** Hasil Normalisasi

Hasil Cleaning	Normalisasi
Waktu produksi kena PPN setelah sampai distributor kena PPN sampai di jual ke retail sidomart dan salfamart	waktu produksi kena ppn setelah sampai distributor kena ppn sampai di jual ke retail sidomart dan salfamart



Hasil Cleaning	Normalisasi
salfamart kena lagi PPN Bener gk sih itunganya ini jadi berapa kali kena PPN Karena pengumuman mendadak diakhir tahun sementara yg dibawah sudah siap dgn PPN Tapi tenang yg sudah kena PPN akan dikembalikan demikian kata DJP Caranya gimana Au ah elap PPN hanya berlaku untuk barang dan jasa mewah demi menjaga keadilan ekonomi dan daya beli rakyat Investasi untuk Masa Depan PPN Demi Fasilitas Publik Lebih Baik PPNBerkeadilan	kena lagi ppn benar tidak sih itunganya ini jadi berapa kali kena ppn karena pengumuman mendadak diakhir tahun sementara yang dibawah sudah siap dengan ppn tapi tenang yang sudah kena ppn akan dikembalikan demikian kata djp caranya bagaimana tau ah elap ppn hanya berlaku untuk barang dan jasa mewah demi menjaga keadilan ekonomi dan daya beli rakyat investasi untuk masa depan ppn demi fasilitas publik lebih baik ppnberkeadilan

d. Tokenisasi.

Tujuan utama tokenisasi adalah memudahkan pemrosesan dan analisis teks dengan mengubah teks mentah menjadi bentuk yang lebih terstruktur. Proses ini menjadi langkah awal dalam berbagai tugas *text mining*, seperti analisis sentimen, klasifikasi teks, dan ekstraksi informasi. Karena dengan memecah teks menjadi unit-unit kata atau token, sistem dapat lebih mudah untuk mengenali pola, menghitung frekuensi kata, serta membangun representasi fitur yang dibutuhkan dalam pemodelan dan pengambilan keputusan berbasis teks. Contoh hasil dari Tokenisasi dapat dilihat pada Tabel 4.

Tabel 4. Hasil Tokenisasi

Hasil Normalisasi	Tokenisasi
waktu produksi kena ppn setelah sampai distributor kena ppn sampai di jual ke retail sidomart dan salfamart kena lagi ppn benar tidak sih itunganya ini jadi berapa kali kena ppn	['waktu', 'produksi', 'kena', 'ppn', 'setelah', 'sampai', 'distributor', 'kena', 'ppn', 'sampai', 'di', 'jual', 'ke', 'retail', 'sidomart', 'dan', 'salfamart', 'kena', 'lagi', 'ppn', 'benar', 'tidak', 'sih', 'itunganya', 'ini', 'jadi', 'berapa', 'kali', 'kena', 'ppn']
karena pengumuman mendadak diakhir tahun sementara yang dibawah sudah siap dengan ppn tapi tenang yang sudah kena ppn akan dikembalikan demikian kata djp caranya bagaimana tau ah elap	['karena', 'pengumuman', 'mendadak', 'diakhir', 'tahun', 'sementara', 'yang', 'dibawah', 'sudah', 'siap', 'dengan', 'ppn', 'tapi', 'tenang', 'yang', 'sudah', 'kena', 'ppn', 'akan', 'dikembalikan', 'demikian', 'kata', 'djp', 'caranya', 'bagaimana', 'au', 'ah', 'elap']
ppn hanya berlaku untuk barang dan jasa mewah demi menjaga keadilan ekonomi dan daya beli rakyat	['ppn', 'hanya', 'berlaku', 'untuk', 'barang', 'dan', 'jasa', 'mewah', 'demi', 'menjaga', 'keadilan', 'ekonomi', 'dan', 'daya', 'beli', 'rakyat']
investasi untuk masa depan ppn demi fasilitas publik lebih baik ppnberkeadilan	['investasi', 'untuk', 'masa', 'depan', 'ppn', 'demi', 'fasilitas', 'publik', 'lebih', 'baik', 'ppnberkeadilan']

e. Filtering

Pada tahap *filtering*, dilakukan proses penghapusan *stopword*, yaitu kata-kata yang tidak memiliki hubungan atau makna signifikan dalam analisis sentimen. Berikut hasil *stopword removal* yang menunjukkan teks yang telah dibersihkan dari kata-kata umum sehingga hanya menyisakan kata-kata yang berkontribusi terhadap pemahaman konteks dan sentiment dari suatu kalimat atau dokumen yang dapat dilihat pada Tabel 5.

Tabel 5. Filtering/ Stopword Removal

Hasil Normalisasi	Stopword Removal
waktu produksi kena ppn setelah sampai distributor kena ppn sampai di jual ke retail sidomart dan salfamart kena lagi ppn benar tidak sih itunganya ini jadi berapa kali kena ppn	['produksi', 'kena', 'ppn', 'distributor', 'kena', 'ppn', 'jual', 'retail', 'sidomart', 'salfamart', 'kena', 'ppn', 'sih', 'itunganya', 'kali', 'kena', 'ppn']
karena pengumuman mendadak diakhir tahun sementara yang dibawah sudah siap dengan ppn tapi tenang yang sudah kena ppn akan dikembalikan demikian kata djp caranya bagaimana tau ah elap	['pengumuman', 'mendadak', 'diakhir', 'dibawah', 'ppn', 'tenang', 'kena', 'ppn', 'dikembalikan', 'djp', 'au', 'ah', 'elap']
ppn hanya berlaku untuk barang dan jasa mewah demi menjaga keadilan ekonomi dan daya beli rakyat	['ppn', 'berlaku', 'barang', 'jasa', 'mewah', 'menjaga', 'keadilan', 'ekonomi', 'daya', 'beli', 'rakyat']
investasi untuk masa depan ppn demi fasilitas publik lebih baik ppnberkeadilan	['investasi', 'ppn', 'fasilitas', 'publik', 'ppnberkeadilan']

f. Stemming

*Stemming* adalah proses menghapus kata depan, kata imbuhan, kata ganti, dan akhiran dengan tetap memperhatikan standar kamus besar Bahasa Indonesia (KBBI). Peneliti melakukan *stemming* menggunakan *library* Sastrawi, yang mana *library sastrawi* merupakan pustaka yang berfungsi untuk pemrosesan Bahasa



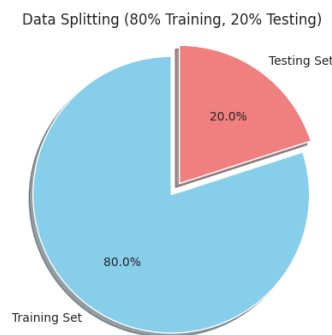
dalam teks berdasarkan kata-kata yang terkandung di dalamnya. Kata-kata dengan makna positif dilabelkan Positif (Satu), kata-kata dengan makna negatif dilabelkan Negatif, dan teks yang tidak menunjukkan sentimen yang jelas dilabelkan Netral. Hasil dari pengkategorian ini dapat dilihat pada **Tabel 7**, dan hasil perbandingan ketiga kategori sentimen disajikan dalam **Gambar 5**.

**Tabel 7.** Pelabelan

Data Mentah	Kategori
@MurtadhaOne1 Waktu produksi kena PPN 12 % setelah sampai distributor kena PPN 12 % sampai di jual ke retail sidomart dan salfamart kena lagi PPN 12 % Bener gk sih itunganya ini... jadi berapa kali kena PPN..???	Negatif
@PaltiWest2024 Karena pengumuman mendadak diakhir tahun sementara yg dibawah sudah siap dgn PPN 12%. Tapi tenang yg sudah kena PPN 12% akan dikembalikan demikian kata DJP. Caranya gimana? Au ah elap.	Positif
@KompasTV PPN 12% hanya berlaku untuk barang dan jasa mewah demi menjaga keadilan ekonomi dan daya beli rakyat.	Positif
Investasi untuk Masa Depan: PPN 12% Demi Fasilitas Publik Lebih Baik. #PPNBerkeadilan	Netral

### 3.1 Pembagian Data

Dalam tahap pembagian data, data set dikelompokkan menjadi, data *testing* dan *training*. Penelitian ini menerapkan komposisi 80:20, di mana 80% dari total dataset digunakan untuk pelatihan (*training*), dan 20% lainnya diterapkan guna pengujian (*testing*). Pembagian dataset bertujuan guna menjamin model algoritma dapat digunakan dengan baik sebelum diuji. Ilustrasi pembagian ditunjukkan dalam Gambar 6 sebagai visualisasi pembagian data.



**Gambar 4.** Visualisasi pembagian data

### 3.2 Penerapan Model *Support Vector Machine* dan *Decision Tree*

Penelitian ini menggunakan metode *Support Vector Machine* (SVM) dan *Decision Tree* untuk melakukan klasifikasi sentimen pada teks. Data teks yang telah melalui tahap *preprocessing* akan dikonversi ke dalam bentuk *vektor numerik* menggunakan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) guna merepresentasikan setiap kata dalam bentuk numerik yang dapat diproses oleh model. Setelah proses transformasi, data dibagi menjadi data pelatihan ( $X_{train}$  dan  $y_{train}$ ) dan data pengujian ( $X_{test}$  dan  $y_{test}$ ) dengan komposisi yang telah ditentukan. Data pelatihan digunakan untuk membangun model klasifikasi, sedangkan data pengujian digunakan untuk mengevaluasi performa model dalam mengklasifikasikan sentimen. Penerapan metode *Support Vector Machine* dan *Decision Tree* ini bertujuan untuk menganalisis efektivitas masing-masing algoritma dalam menentukan sentimen publik terhadap penerapan kebijakan PPN 12%.

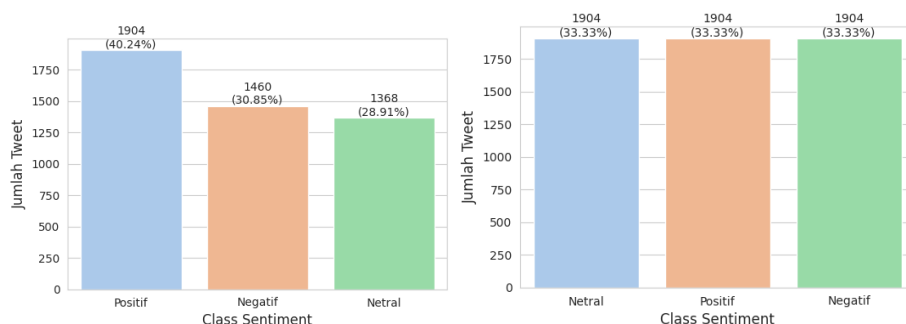
#### 3.2.1 Pembobotan TF-IDF

Pada tahap ini, penelitian menggunakan metode *Support Vector Machine* (SVM) dan *Decision Tree* untuk melakukan klasifikasi sentimen pada teks. Data yang telah melalui proses *preprocessing* dikonversi ke dalam bentuk vektor numerik menggunakan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) agar dapat diolah oleh model pembelajaran mesin. Setelah proses transformasi, data dibagi menjadi dua subset, yaitu data pelatihan ( $X_{train}$  dan  $y_{train}$ ) yang digunakan untuk melatih model, serta data pengujian ( $X_{test}$  dan  $y_{test}$ ) yang digunakan untuk mengukur performa model dalam mengklasifikasikan sentimen. Proses ini bertujuan untuk mengevaluasi efektivitas SVM dan *Decision Tree* dalam menganalisis opini publik terkait penerapan PPN 12%.

#### 3.2.2 Penyeimbangan data (SMOTE)

Penyeimbangan data diperlukan untuk mengatasi ketidakseimbangan kelas dalam dataset, yang dapat mempengaruhi performa model klasifikasi. Berdasarkan hasil pelabelan, jumlah data sentimen positif sebanyak 1.904, sentimen negatif sebanyak 1.460, dan sentimen netral sebanyak 1.368. Ketidakseimbangan ini dapat menyebabkan model lebih cenderung mengenali kelas mayoritas (positif), sementara kelas minoritas (negatif dan netral) berisiko kurang

terwakili dalam proses pembelajaran mesin. Untuk mengatasi masalah ini, penelitian menerapkan teknik *oversampling* menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE bekerja dengan menghasilkan data sintetis untuk kelas minoritas, sehingga jumlahnya lebih seimbang dengan kelas mayoritas dalam dataset [19]. Dengan menerapkan SMOTE, model diharapkan dapat lebih adil dalam mengklasifikasikan ketiga kategori sentimen.



**Gambar 5.** Perbandingan kategori kelas sentimen sebelum dan sesudah dilakukan SMOTE.

Pada Gambar 5, didapatkan data sebelum proses *balancing* yang berlabel netral berjumlah 1.460 dan label negatif berjumlah 1.368 serta positif 1.904. Kemudian dilakukan proses *balancing* sehingga membuat data menjadi seimbang untuk kelas netral dan negatif berjumlah 1.904. Selanjutnya dilakukan pelatihan pemodelan menggunakan model algoritma *Support Vector Machine* (SVM) dan *Decision Tree*.

### 3.3 Evaluasi Hasil

Proses evaluasi hasil dilakukan untuk melakukan pemeriksaan lanjutan terkait model yang telah dibuat sebelumnya.

#### 3.3.1 Algoritma *Support Vector Machine* (SVM)

Dalam penelitian ini, dilakukan analisis komparatif antara algoritma *Support Vector Machine* (SVM) dan *Decision Tree* untuk klasifikasi sentimen publik terhadap kebijakan PPN 12%. Data yang digunakan berupa komentar atau opini publik yang dikumpulkan dari berbagai sumber, kemudian diproses menggunakan teknik pre-processing seperti tokenisasi, *stopword removal*, *stemming*, dan *vectorization* (TF-IDF). Model SVM diimplementasikan menggunakan fungsi *SVC* dari *library Scikit-Learn* dengan parameter default serta pengaturan *random\_state=0* untuk memastikan replikasi hasil yang konsisten. Tabel 9 menampilkan hasil pengolahan data menggunakan algoritma SVM sebelum dan sesudah dilakukan proses penyeimbangan data.

**Tabel 8.** Hasil Model SVM sebelum dan sesudah SMOTE

Model	Class	Accuracy	Precision	Recall	F1-Score
<i>Before Smote</i>	Negatif	74%	74%	80%	77%
	Netral		62%	53%	57%
	Positif		81%	83%	82%
<i>After Smote</i>	Negatif	80%	78%	87%	82%
	Netral		77%	71%	74%
	Positif		85%	83%	84%

Berdasarkan table 9, dapat disimpulkan penerapan metode penyeimbangan data atau *SMOTE* pada algoritma *Support Vector Machine*, berhasil meningkatkan kinerja algoritma yang ditunjukkan dengan peningkatan akurasi dari 74% menjadi 80% serta perbaiki nilai *precision*, *recall*, dan *F1-Score* secara signifikan pada seluruh kelas kategori, khususnya pada kelas kategori Netral yang sebelumnya memiliki performa terendah.

#### 3.3.2 Algoritma *Decision Tree*

Algoritma *Decision Tree* adalah metode pembelajaran mesin yang menggunakan struktur pohon untuk membuat keputusan berdasarkan serangkaian aturan yang berasal dari fitur data. Setiap node internal dalam pohon mewakili suatu pengujian terhadap atribut (misalnya, apakah suatu nilai lebih besar dari ambang batas tertentu), setiap cabang mewakili hasil pengujian tersebut, dan setiap daun (leaf) mewakili label kelas atau nilai target. Proses ini berlanjut secara rekursif hingga semua data terklasifikasi atau hingga memenuhi kriteria penghentian tertentu [20]. model *Decision Tree* diterapkan menggunakan fungsi *DecisionTreeClassifier* dari *library* yang sama, dengan parameter default. Tabel 10 menampilkan hasil pengolahan data menggunakan algoritma *Decision Tree* sebelum dan setelah dilakukan proses penyeimbangan data.

**Tabel 9.** Hasil Model Decision Tree Sebelum dan Sesudah SMOTE

Model	Class	Accuracy	Precision	Recall	F1-Score
<i>Before Smote</i>	Negatif	64%	64%	60%	62%
	Netral		54%	59%	57%

Model	Class	Accuracy	Precision	Recall	F1-Score
After Smote	Positif		72%	71%	71%
	Negatif	68%	73%	65%	69%
	Netral		59%	72%	65%
	Positif		76%	69%	72%

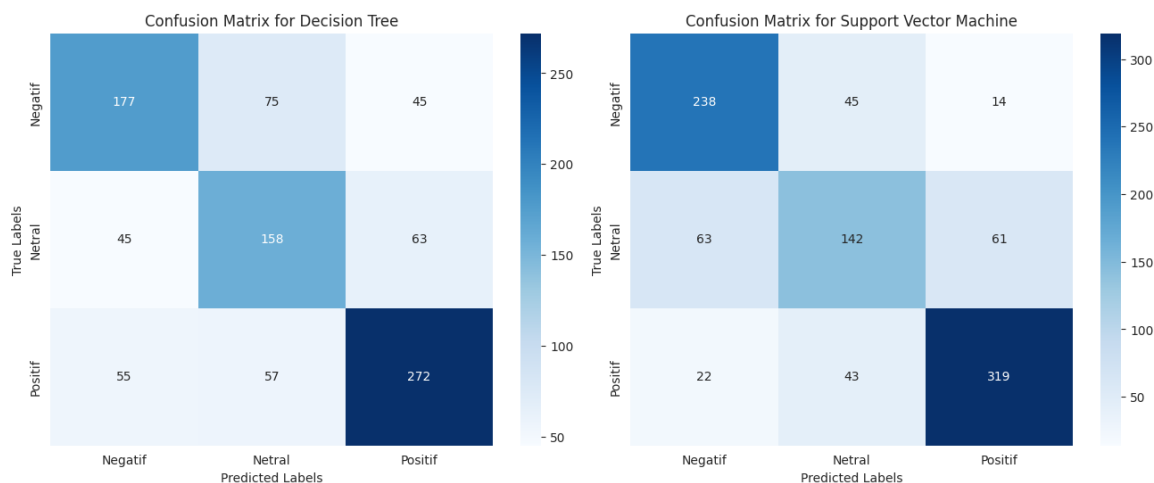
Berdasarkan table 10, dapat disimpulkan bahwa penerapan metode penyeimbangan data atau SMOTE pada algoritma *Decision Tree* juga mengalami peningkatan akurasi dari 64% menjadi 69%, serta terdapat perbaikan nilai *F1-Score* untuk semua kelas kategori, khususnya pada kelas kategori Netral yang juga mengalami peningkatan pada nilai *Recall* dari 59% menjadi 72%, sehingga menghasilkan model klasifikasi algoritma *Decision Tree* yang lebih seimbang.

### 3.3.3 Confusion Matrix

*Confusion matrix* merupakan suatu metode yang digunakan untuk mengevaluasi performa model klasifikasi dengan menyajikan jumlah prediksi yang benar dan salah dalam bentuk tabel. Matriks ini memberikan gambaran mengenai perbandingan antara hasil klasifikasi yang dilakukan oleh sistem dengan klasifikasi yang sebenarnya. Dalam *confusion matrix*, terdapat empat kategori hasil klasifikasi, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *True Positive* (TP) dan *True Negative* (TN). Dalam penelitian ini, *confusion matrix* yang dihasilkan memiliki ukuran 3x3, sesuai dengan jumlah kelas yang dianalisis.

### 3.3.4 Sebelum Proses *Balancing* (SMOTE)

Ditampilkan pada gambar 7 hasil dari perbandingan *Confusion matrix* hasil kinerja dari model *Support Vector Machine* (SVM) dan *Decision Tree* sebelum diberlakukan SMOTE dalam pembagian sentimen publik ke dalam tiga kategori utama, yaitu Negatif (Satu), Netral (Nol), dan Positif (Dua). Model ini menunjukkan performa algoritma *Support Vector Machine* yang cukup baik dalam mengidentifikasi sentimen positif dan negatif, di mana sebanyak 238 data berhasil dikelompokkan dengan valid sebagai negatif, meskipun masih terdapat 45 data negatif yang salah dikategorikan sebagai netral serta 14 data negatif yang keliru diprediksi sebagai positif. Sementara itu, dalam kategori netral (nol), model mampu mengelompokkan 142 data dengan benar, namun terdapat 63 data netral yang salah diidentifikasi sebagai negatif, serta 61 data lainnya yang justru diklasifikasikan ke dalam kategori positif, yang mengindikasikan adanya kesulitan dalam membedakan sentimen netral dari kategori lainnya. Dalam klasifikasi sentimen positif, model berhasil mengelompokkan 319 data secara akurat, sekalipun terdapat data positif berjumlah 22 yang salah diperkirakan sebagai negatif dan 43 data positif yang keliru dikategorikan sebagai netral.



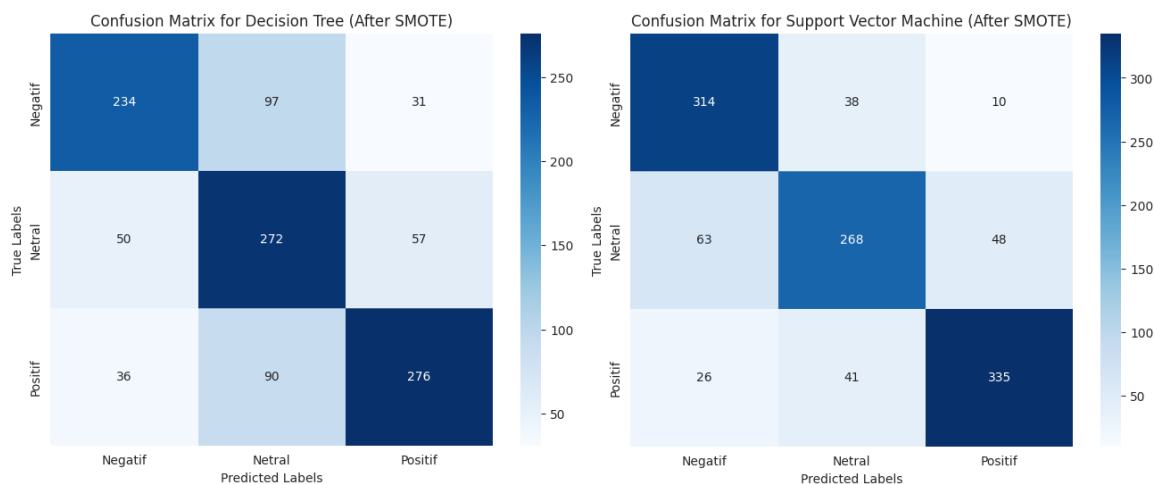
Gambar 6. Perbandingan *Confusion Matrix Support Vector Machine* dan *Decision Tree* Sebelum SMOTE.

Sementara itu, hasil *Confusion matrix* yang merepresentasikan evaluasi kinerja model *Decision Tree* sebelum proses SMOTE dalam mengelompokkan pendapat public, yang diklasifikasikan dalam tiga kelas utama, yaitu Negatif (Satu), Netral (Nol), dan Positif (Dua). Pada kategori Positif, model menunjukkan kinerja yang cukup baik, model ini mampu mengelompokkan 272 data secara akurat. Meski begitu, masih terdapat 55 data positif yang salah diidentifikasi sebagai Negatif dan 57 data positif yang diprediksi sebagai Netral. Untuk kategori Negatif, model berhasil mengenali 177 data dengan benar. Namun, terdapat 75 data negatif yang keliru diprediksi sebagai Netral, serta 45 data negatif yang justru diklasifikasikan sebagai Positif. Hal ini mengindikasikan bahwa model mengalami kesulitan dalam membedakan antara sentimen Negatif dan Netral. Sementara itu, pada kelas Netral, model dapat mengklasifikasikan 158 data dengan benar. Sayangnya, 45 data netral salah terklasifikasi sebagai Negatif dan 63 data netral diprediksi sebagai Positif, yang menunjukkan bahwa model masih belum optimal dalam mengenali karakteristik spesifik

sentimen Netral. Kondisi ini memperlihatkan adanya tumpang tindih karakteristik antara sentimen Netral dengan kelas lainnya.

### 3.3.5 Setelah Proses *Balancing* (SMOTE)

Hasil evaluasi menggunakan confusion matrix berukuran 3x3 setelah balancing data dengan SMOTE menunjukkan peningkatan performa model, khususnya dalam klasifikasi kategori Netral yang sebelumnya menjadi tantangan utama. Pada model *Support Vector Machine* (SVM), kategori Negatif berhasil diklasifikasikan dengan baik sebanyak 314 data, meskipun masih terdapat kesalahan klasifikasi ke kelas Netral (38 data) dan Positif (10 data). Kategori Netral menunjukkan peningkatan akurasi dengan 268 data diklasifikasikan secara tepat, walaupun 63 data masih salah dikenali sebagai Negatif dan 48 sebagai Positif. Kategori Positif juga menunjukkan performa baik, dengan 335 data diklasifikasikan benar, meskipun 26 data salah dikenali sebagai Negatif dan 41 sebagai Netral. Secara keseluruhan, SVM menunjukkan performa yang lebih konsisten pasca-SMOTE, dengan distribusi data yang lebih seimbang membantu model mengenali pola sentimen secara lebih akurat. Sementara itu, model *Decision Tree* juga menunjukkan peningkatan setelah penerapan SMOTE. Untuk kategori Negatif, sebanyak 234 data berhasil diklasifikasikan dengan benar, namun masih terdapat 97 dan 31 data yang salah diprediksi sebagai Netral dan Positif. Kategori Netral diklasifikasikan dengan benar sebanyak 272 data, namun masih terdapat kesalahan signifikan ke kategori Negatif (50 data) dan Positif (57 data). Pada kategori **Positif**, model berhasil mengklasifikasikan 276 data dengan benar, namun masih kesulitan membedakan data yang cenderung Netral, dengan 90 data salah diprediksi ke kelas tersebut.



**Gambar 7.** Perbandingan *Confusion Matrix Support Vector Machine* dan *Decision Tree* Setelah SMOTE

Dari kedua confusion matrix, terlihat bahwa penerapan SMOTE berhasil meningkatkan performa kedua model, terutama dalam mengidentifikasi kategori Netral. Namun, *Decision Tree* masih cenderung lebih rentan terhadap kesalahan klasifikasi antar kategori yang memiliki karakteristik serupa. Dengan demikian, meskipun SMOTE efektif dalam mengatasi ketidakseimbangan data, optimalisasi lanjutan seperti tuning hyperparameter, pemilihan fitur yang lebih relevan, dan teknik preprocessing lanjutan masih diperlukan untuk mencapai kinerja klasifikasi yang lebih akurat dan konsisten.

## 4. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa metode balancing data menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) memiliki pengaruh yang signifikan terhadap peningkatan akurasi kedua algoritma klasifikasi yang digunakan. Pada algoritma *Support Vector Machine* (SVM), akurasi awal yang sebesar 74% meningkat menjadi 80% setelah dilakukan penyeimbangan data. Demikian pula, algoritma *Decision Tree* mengalami peningkatan akurasi dari 64% menjadi 68%. Hasil ini membuktikan bahwa penerapan SMOTE mampu mengatasi ketidakseimbangan data, sehingga meningkatkan kinerja model dalam mengklasifikasikan sentimen publik. Dengan nilai akurasi akhir yang lebih tinggi, algoritma SVM dapat menjadi pilihan yang lebih optimal dalam analisis sentimen publik, termasuk dalam penerapan kebijakan seperti PPN 12%.

## REFERENCES

- [1] Republik Indonesia, “Undang-Undang No. 7 Tahun 2021 tentang Harmonisasi Peraturan Perpajakan (HPP),” *Republik Indones.*, vol. 12, no. November, pp. 1–68, 2021.
- [2] Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), “Internet Indonesia,” 2024.
- [3] N. Fauziah, M. Alkautsar, Y. Suryaman, and F. F. Roji, “Pelabelan VADER Dalam Menganalisis Persepsi Masyarakat Terhadap Kenaikan Tarif PPN di Indonesia,” *J. Akunt. dan Keuang. JAK*, vol. 12, no. 2, pp. 228–238, 2024.
- [4] P. Arsi and R. Waluyo, “Analisis Sentimen Wacana Pemandangan Ibu Kota Indonesia Menggunakan Algoritma Support



- Vector Machine (SVM),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021.
- [5] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, pp. 18–26, 2021.
- [6] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, “Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 2, no. 2, pp. 91–97, 2021.
- [7] A. T. Octa Nuryawan, M. Hasbullah, M. Rizal, M. F. Rajab, and N. Agustina, “Algoritma Decision Tree Untuk Analisis Sentimen Public Terhadap Marketplace Diindonesia,” *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 5, no. 1, pp. 18–25, 2023.
- [8] L. A. Pramesti and N. Pratiwi, “Analisis Sentimen Twitter Terhadap Program MBKM Menggunakan Decision Tree dan Support Vector Machine,” *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1145–1154, 2023.
- [9] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021.
- [10] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023.
- [11] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, “Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine,” *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 542, 2022.
- [12] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, “Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022.
- [13] S. F. N. Hadju and R. Jayadi, “Sentiment analysis of Indonesian e-commerce product reviews using support vector machine based term frequency inverse document frequency,” *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 17, pp. 4316–4325, 2021.
- [14] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023.
- [15] F. D. Adhiatma and A. Qoiriah, “Penerapan Metode TF-IDF dan Deep Neural Network untuk Analisa Sentimen pada Data Ulasan Hotel,” *J. Informatics Comput. Sci.*, vol. xx, pp. 183–193, 2022.
- [16] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, “Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks,” *Appl. Sci.*, vol. 13, no. 6, pp. 1–43, 2023.
- [17] M. H. Wicaksono, M. D. Purbolaksono, and S. Al Faraby, “Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily,” *eProceedings Eng.*, vol. 10, no. 3, pp. 3591–3600, 2023.
- [18] R. Riyaddulloh and A. Romadhony, “Normalisasi Teks Bahasa Indonesia Berbasis Kamus Slang Studi Kasus: Tweet Produk Gadget Pada Twitter,” *eProceedings Eng.*, vol. 8, no. 4, pp. 4216–4228, 2021.
- [19] A. Syukron, S. Sardiarinto, E. Saputro, and P. Widodo, “Penerapan Metode Smote Untuk Mengatasi Ketidakeimbangan Kelas Pada Prediksi Gagal Jantung,” *J. Teknol. Inf. dan Terap.*, vol. 10, no. 1, pp. 47–50, 2023.
- [20] E. Priyanti, “Penerapan Decision Tree Untuk Klasifikasi Tingkat Pendapatan,” *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 7, no. 1, pp. 7–12, 2022.