Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 3, Maret 2025, Hal. 831-843
DOI: https://doi.org/10.52436/1.jpti.710
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Evaluasi Opini Publik di Media Sosial X terhadap Kebijakan Pajak Pertambahan Nilai 12% di Indonesia Menggunakan Naive Bayes dan Decision Tree

Eka Putri Adamansyah[1], Aditia Yudhistira*[2]

[1]Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
[2]Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
Email: [1]eckaputri0@gmail.com, [2]aditiayudhistira@teknokrat.ac.id

Abstrak

Penerapan kebijakan Pajak Pertambahan Nilai (PPN) 12% di Indonesia telah memicu beragam tanggapan dari masyarakat, khususnya di media sosial. Penelitian ini bertujuan untuk mengevaluasi pandangan publik terhadap kebijakan tersebut dengan memanfaatkan data dari media sosial X. Data dikumpulkan melalui teknik crawling, menghasilkan 1.815 tweet yang relevan dengan diskusi mengenai PPN 12%. Tahapan analisis meliputi preprocessing data serta pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), diikuti dengan klasifikasi sentimen menggunakan algoritma *Naive Bayes* dan *Decision Tree*. Hasil penelitian menunjukkan bahwa meskipun algoritma Decision Tree memiliki akurasi yang lebih tinggi (93,44%) dibandingkan *Naive Bayes* (92,68%), namun *Naive Bayes* lebih efisien dalam menangani dataset yang lebih besar. Dari seluruh tweet yang dianalisis, 94,54% mengandung sentimen negatif terkait kekhawatiran tentang dampak ekonomi dan peningkatan beban pajak, sementara 5,46% mengandung sentimen positif yang umumnya menyoroti potensi peningkatan penerimaan negara dan pembangunan. Penelitian ini menyediakan wawasan bagi pemerintah dalam memahami persepsi publik serta merancang strategi komunikasi yang lebih efektif terkait kebijakan perpajakan.

Kata kunci: *Analisis Sentimen, Decision Tree, Media Sosial X, Naive Bayes Pajak Pertambahan Nilai*

Evaluation of Public Opinion on Social Media X Regarding the 12% Value Added Tax Policy in Indonesia Using Naive Bayes and Decision Tree

Abstract

The implementation of the 12% Value Added Tax (VAT) policy in Indonesia has sparked various reactions from The implementation of the 12% Value Added Tax (VAT) policy in Indonesia has sparked various reactions from the public, particularly on social media. This study aims to evaluate public opinion on the policy by utilizing data from social media platform X. Data were collected using web crawling techniques, resulting in 1,815 tweets relevant to discussions about the 12% VAT. The analysis stages include data preprocessing and word weighting using Term Frequency-Inverse Document Frequency (TF-IDF), followed by sentiment classification using Naive Bayes and Decision Tree algorithms. The findings show that although the Decision Tree algorithm achieved a higher accuracy rate (96.97%) compared to Naive Bayes (94.49%), Naive Bayes was more efficient in handling larger datasets. Of the tweets analyzed, 94.54% contained negative sentiment related to concerns about economic impacts and increased tax burdens, while 5.46% contained positive sentiment, primarily highlighting the potential for increased state revenue and development. This study provides valuable insights for the government in understanding public perception and developing more effective communication strategies regarding tax policies.

Keywords: *Decision Tree, Naive Bayes, Sentiment Analysis, Social Media X, Value-Added Tax*

1. PENDAHULUAN

Pajak Pertambahan Nilai (PPN) merupakan salah satu sumber pendapatan negara yang utama dalam perekonomian Indonesia. Salah satu upaya Kementerian Keuanga dalam meningkatan pendapatan pemerintah Indonesia yaitu melalui kebijakan kenaikan pajak dan mengatur sistem perpajakan yang lebih efisien dan adil. Salah satu kebijakan terbaru yang direncakan pemerinta yaitu adanya perubahan tarif PPN yang naik menjadi 12% [1]-[2]. Kebijakan perubahan tarif PPN menimbulkan reaksi yang beragam dari masyarakat baik dari sisi pelaku bisnis, konsumen, maupun pemerintah itu sendiri. Dalam masa digital sekarang terutama penggunaan media sosial menjadi salah satu *platform* yang digunakan oleh masyarakat untuk mengekspresikan pendapat baik

p-ISSN: 2775-4227
e-ISSN: 2775-4219

pendapat yang bersifat positif maupun negatif terhadap kebijakan pemerintah [3]. Sehubungan dengan meningkatnya jumlah pengguna media sosial hal ini dapat mempengaruhi analisis terhadap sentimen yang berkembang di *platform* tersebut menjadi sangat penting. Menurut data yang dipublikasikan oleh situs Databoks, yang bersumber dari laporan We Are Social, media sosial X saat ini menjadi fpilihan terbanyak di kalangan generasi Z. Saat ini, terdapat 564,1 juta pengguna X di seluruh dunia. Pada Juli 2023, jumlah pengguna X di Indonesia mengalami peningkatan, menjadikannya negara dengan pengguna terbanyak keempat secara global. Sebelumnya, Indonesia berada di peringkat keenam pada Mei 2023. Persentase pengguna X di Indonesia mencapai 71,2%, dengan total pengguna mencapai 25,25 juta per Juli 2023[1]-[2]. Sentimen atau pendapat yang diungkapkan oleh masayarakat di media sosial tidak hanya mencerminkan opini publik namun dapat mempengaruhi citra pemerintah, keputusan bisnis, dan perkembangan kebijakan lebih lanjut [4]. Oleh karena itu penting untuk melakukan analisis sentimen terhadap berbagai respon yang muncul di media sosial mengenai kebijakan kenaikan PPN 12% khususnya pada media X atau sebelumnya dikenal dengan nama Twitter.

Media sosial X merupakan salah satu platform yang banyak digunakan oleh masyarakat Indonesia untuk menyampaikan opini atau pendapat perihal berbagai isu publik [5]. Dalam analisis sentimen yang diungkapkan oleh masyarakat di media sosial X dapat memberikan gambaran yang jelas mengenai kebijakan PPN 12% diterima oleh masyarakat [6]-[7]. Sehingga perlu dilakukan analisis sentimen terhadap opini kebijakan Pajak PPN 12% menggunakan algortima dalam *machine learning* diantaranya menggunakan algoritma Naive Bayes dan Desician Tree. Analisis sentime terhadap kebijakan kenaikan PPN 12% diharapkan dapat menjadi rekomendasi apakah kebijakan pemerintah tersebut diperlu atau perlu dikaji ulang untuk diterapkan. Analisis sentimen ini bertujuan untuk mengklasifikasikan apakah suatu teks atau postingan pengguna media sosial mengandung sentimen positif, negatif, atau netral terhadap kebijakan yang sedang dianalisis. Beberapa penelitian terkait terhadap opini *public* tentang pajak diantara menganalisis sentiment analisis pajak menggunakan metode vader hasil peneitian tersebut menyatakan bahwa insentif perpajakan memiliki potensi untuk mendorong pertumbuhan ekonomi [8]. Selanjutnya pada penelitian lainnya yang membahas dampak kenaikan pajak dapat memengaruhi pertumbungan perekonomian [9]-[10]. Selanjutnya penelitian terkait kenaikan pajak menggunakan menggunakan Bert [11] dimana penelitian tersebut menunjukkan hasil akurasi sebesar 83%. Selanjutnya penelitian terkait yaitu membahas tantang kenaikan pajak menggunakan algoritma. Kemudian penelitian sebelumnya relavan pada penelitian ini juga membahas tentang kenaikan pajak menggunakan metode KNN yaitu sebesar 66,67% [12]. Selanjunya pada penelitian terkait membahas analisis sentimen terhadap kenaikan PPN 12% menggunakan algoritma *decision tree* dengan tingkat akurasi sebesar 41,34% [7]. Dengan demikian penelitian ini dapat dikembangkan lagi dengan menggunakan berbagai teknik dan algoritma diantaranya adalah metode *Naive Bayes* dan *Decision Tree*. Kedua metode ini banyak digunakan dalam analisis sentimen karena kemampuannya yang baik dalam mengklasifikasikan data teks dalam jumlah besar. Metode Naive Bayes adalah salah satu pendekatan probabilistik yang memanfaatkan asumsi independensi antar fitur untuk melakukan klasifikasi [13]. Algoritma *Naive Bayes* terbukti efektif dalam berbagai aplikasi analisis sentimen karena kecepatan dan kemudahan dalam penerapannya. Sementara algoritma *Decision Tree* merupakan metode berbasis pohon keputusan yang membagi data menjadi beberapa cabang berdasarkan atribut tertentu untuk memperoleh hasil yang paling optimal [14]. Keunggulan utama dari *Decision Tree* adalah kemampuannya untuk menghasilkan model yang mudah dipahami sehingga hasil klasifikasi dapat dimengerti oleh pengambil keputusan. Fokus utama dalam penelitian ini adalah untuk menganalisis sentimen pengguna media sosial X terhadap kebijakan kenaikan tarif PPN 12% di Indonesia menggunakan metode *Naive Bayes* dan *Decision Tree*. Hasil analisis sentimen ini diharapkan dapat memperoleh wawasan yang lebih mendalam mengenai bagaimana persepsi masyarakat terhadap kebijakan tersebut dan bagaimana dampak terhadap sektor-sektor tertentu dalam perekonomian di Indonesia.

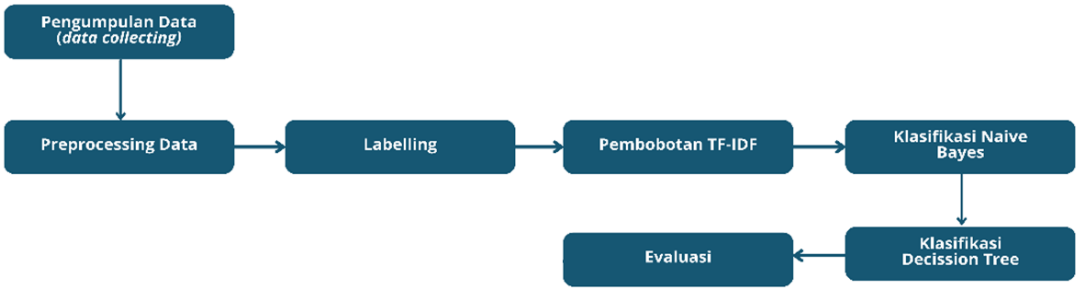
Hasil penelitian ini dapat menjadi referensi bagi pihak-pihak terkait seperti pemerintah, akademisi, dan pelaku bisnis dalam merumuskan kebijakan yang lebih responsif terhadap opini publik yang berdampak pada perkembangan ekonomi di masyarakat. Dengan demikian, penelitian ini tidak hanya relevan dari segi akademis, tetapi juga memiliki dampak praktis dalam meningkatkan kualitas pengambilan keputusan di Indonesia. Penelitian ini diharapkan dapat memberikan gambaran yang jelas tentang bagaimana masyarakat bereaksi terhadap perubahan kebijakan tentang pajak sehingga analisis sentimen dapat memberikan informasi yang berguna bagi pengambilan kebijakan. Penelitian ini bertujuan untuk mengevaluasi sentimen publik terhadap kebijakan PPN 12% menggunakan algoritma *Naive Bayes* dan *Decision Tree* serta membandingkan efektivitas kedua metode dalam analisis sentimen media sosial.

2. METODE PENELITIAN

Metode penelitian yang akan digunakan dalam analisis sentimen pada masyarakat di Media Sosial X menggunakan metode *Naive Bayes* dan *Decision Tree.* Penyelesaian penelitian ini ada beberapa tahapan yang dilakukan yang dapat dilihat pada Gambar 1. Pada Metode penelitian dalam penelitian ini dirancang untuk

p-ISSN: 2775-4227
e-ISSN: 2775-4219

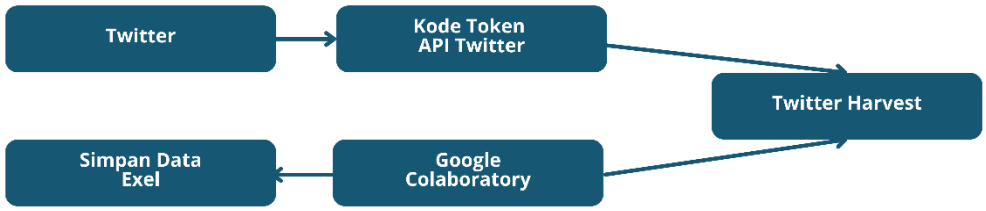
memberikan pemahaman mendalam terkait sentimen terhadap kebijakan pajak pertambahan nilai (ppn) 12% yang diambil dari media sosial X dengan membandingkan kinerja algoritma *Naive Bayes* dan *Decission Tree*. Pada Gambar 1 dijelaskan tahapan-tahapan yang meliputi pengumpulan data (*data collecting*), preproses data (*preprocessing data*), *labelling data*, pemodelan, evaluasi, dan visualisasi.



Gambar 1. Metode Penelitian

2.1. Pengumpulan Data (*Data Collecting*)

Data yang digunakan dalam penelitian ini merupakan data mentah yang dikumpulkan dari media sosial X menggunakan teknik crawling. Dataset diperoleh melalui scrapping menggunakan library Tweet Harvest dalam bahasa pemrograman Python. Proses crawling membutuhkan auth token dari Twitter dan berhasil mengumpulkan sebanyak 1.815 tweet yang relevan. Kata kunci yang digunakan dalam pencarian meliputi "kebijakan PPN 12%", "Pajak", dan "PPN", dengan penerapan filter bahasa Indonesia (lang:id) untuk memastikan data sesuai dengan kebutuhan penelitian. Dataset yang terkumpul mencakup rentang waktu antara tahun 2023 hingga 2024 karena ketersediaan data yang relevan dan tren diskusi di media social terhadap Kebijakan Pajak Pertambahan Nilai (PPN) 12% di Indonesia dalam periode waktu tersebut. Tahapan proses *crawling* data Pajak Pertambahan Nilai 12% dapat dilihat pada Gambar 2.



Gambar 2. Proses *Crawling*

2.2. *Preprocessing* Data

Setelah mendapat data mentah langkah selanjutnya yaitu *preprocessing* data. *Preprocessing* data adalah langkah penting dalam analisis sentimen untuk memastikan data siap digunakan untuk diolah lebih lanjut atau dianalisis. Tahapan ini adalah pengolahan data mentah menjadi data bersih, berikut tahapannya :

1. *Cleansing*
   Proses *cleansing* merupakan tahap pembersihan data teks dari komponen yang tidak diperlukan atau dianggap sebagai noise. Komponen yang akan dihilangkan antara lain karakter HTML, simbol emoticon, hashtag (#), username account (@username), retweet (RT), link URL, dan alamat website [15].
2. *Case Folding*
   Proses *case folding* merupakan fase mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*) yang bertujuan untuk memastikan bahwa perbedaan dalam kapitalisasi tidak mempengaruhi analisis teks. Misalnya, "Data" dan "data" akan dianggap sama setelah case folding. Hal ini membantu dalam menyederhanakan data dan mengurangi variasi kata yang tidak diperlukan [16].

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3. *Tokenize*
   Tahap *tokenization* adalah proses memecahkan kalimat menjadi kata-kata individu. Proses ini melibatkan pemisahan seluruh karakter dari kumpulan dokumen menjadi bagian-bagian kecil berupa kata. Bagian-bagian kecil ini disebut sebagai token [15].
4. *Stopword*
   Tahap *stopword* adalah proses menghapus kata-kata yang dianggap tidak relevan atau terlalu umum dalam teks atau dokumen. *Stopword* biasanya berupa kata-kata umum yang sering muncul dalam banyak kalimat dalam jumlah besar, seperti kata penghubung, contohnya: "adalah", "ke", "di", "dari", "dan"[17].
5. *Stemming*
   Tahap *stemming* merupakan tahap transformasi kata-kata yang sebelumnya terdapat imbuhan menjadi bentuk kata dasarnya yang kemudian dapat diperkirakan bahwa kata yang mengalami transformasi memiliki arti dan makna yang sama dengan kata dasarnya[18].

2.3. *Labelling* Data

Pada tahap *labelling* dataset dikategorikan menjadi 3 label yaitu positif, netral, dan negatif. Proses penentuan label ini dilakukan melalui fungsi polaritas akan menghasilkan nilai -0.5 dan mendekati 0.5 dapat menghasilkan polaritas positif, sedangkan polaritas >-0.5 menghasilkan nilai netral dan mendekati <= -0.5 maka menghasilkan negatif yang terkandung dalam komentar di Twitter . Proses *labelling* dimana data dibagi menjadi dua kategori label yaitu positif dan negatif. Setelah proses labeling jumlah tweet terdiri dari 1715 bernilai negatif, 99 bernilai positif, dan 0 bernilai netral.

2.4 Pembobotan TF-IDF

Tahap TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan teknik yang digunakan untuk mengubah teks menjadi representasi numerik yang dapat digunakan oleh model *machine learning*. Metode ini menilai pentingnya kata dalam sebuah dokumen relatif terhadap kumpulan dokumen. *Term Frequency* (TF) adalah kemunculan sebuah kata dalam suatu dokumen sedangkan *Inverse Document Frequency* (IDF) adalah jumlah seluruh dokumen yang mengandung kata tertentu. Dengan menggabungkan TF dan IDF, TF-IDF dapat menghitung total bobot dari kata dalam sebuah dokumen, memberikan nilai yang lebih tinggi untuk kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain.

1. *Term Frequency* (TF): TF mengukur frekuensi kemunculan sebuah kata (*term*) dalam dokumen dapat menggunakan persamaan berikut:

$$TF(t,d) = \frac{Jumlah\ kemunculan\ termt\ dalam\ dokumen\ d}{total\ jumlah\ term\ dalam\ dokumen\ d} \quad (1)$$

   TF adalah rasio antara jumlah kemunculan term t dalam dokumen d dengan total jumlah term dalam dokumen tersebut, mengukur seberapa sering sebuah term muncul dalam dokumen.
2. *Inverse Document Frequency* (IDF): IDF mengukur seberapa banyak dokumen dalam koleksi D yang mengandung term t dapat menggunakan persamaan berikut:

$$IDF(t,D) = log\ \frac{Total\ jumlah\ dokumen\ dalam\ koleksi\ D}{Jumlah\ dokumen\ yang\ mengandung\ term\ t+1} \quad (2)$$

   IDF adalah logaritma dari rasio total jumlah dokumen dalam koleksi D dengan jumlah dokumen yang mengandung term t ditambah satu. Ini memberikan bobot lebih tinggi pada term yang jarang muncul di banyak dokumen mengurangi pengaruh term yang umum.
3. Gf TF-IDF: TF-IDF adalah hasil perkalian antara *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) dapat menggunakan persamaan berikut:

$$TF - IDF\ (t,d,D) = TF(t,d)\ XIDF(t,d) \quad (3)$$

   TF-IDF menghitung total bobot dari kata dalam sebuah dokumen, memberikan nilai yang lebih tinggi untuk kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain.
   Keterangan dalam rumus tersebut adalah sebagai berikut:
   - t adalah term atau kata kunci tertentu.
   - d adalah dokumen tertentu.
   - D adalah koleksi seluruh dokumen.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.5 Model Klasifikasi *Naive Bayes*

Model klasifikasi menggunakan algoritma *Naive Bayes* merupakan gambaran matematis yang digunakan oleh suatu sistem untuk membagi atau mengkategorikan data ke dalam berbagai kelas atau kategori berdasarkan fitur yang ada. Algoritma ini populer dalam bidang *machine learning* karena efisiensi dan keandalannya. Tujuan utama dari model klasifikasi seperti *Naive Bayes* untuk menghasilkan prediksi yang tepat dan konsisten untuk data yang belum pernah dilihat sebelumnya [19]. Model Naive Bayes diyakini mampu meningkatkan akurasi dalam klasifikasi [20]. Pemilihan model *Naive Bayes* karena model ini mampu menangani data teks dengan efisien dan bekerja baik dalam klasifikasi teks yang besar.

2.6 Model Klasifikasi *Decision Tree*

Model *Decision Tree* merupakan suatu metode yang memanfaatkan struktur pohon atau hirarki dalam membuat keputusan. Setiap simpul dalam pohon mewakili suatu fitur dan setiap cabang mewakili aturan keputusan. Model *Decision Tree* sangat terkenal dalam bidang *machine learning* karena kemudahan interpretasi dan fleksibilitas dalam menangani berbagai jenis data. Tujuan utama dari model klasifikasi menggunakan algoritma *decision tree* untuk menghasilkan prediksi yang akurat dan konsisten untuk data [21]. Model *Decision Tree* dipilih karena memiliki kemampuan dalam memahami hubungan yang lebih kompleks antar variabel.

2.7 *SMOTE*

SMOTE (Synthetic Minority Oversampling Technique) merupakan metode yang digunakan untuk mengatasi ketidakseimbangan kelas dalam sebuah dataset. Ketika salah satu kelompok data memiliki jumlah yang jauh lebih besar dibandingkan kelompok lainnya, ketidakseimbangan ini dapat menyebabkan model lebih cenderung memprediksi kelas mayoritas. Dengan menerapkan SMOTE, pola-pola penting dari kelas minoritas dapat teridentifikasi secara lebih baik, sehingga menghasilkan prediksi yang lebih akurat [22]-[23].

2.8 Evaluasi Kinerja Model

Evaluasi kinerja model dalam analisis sentimen melibatkan penilaian hasil eksperimen sistem serta tanggapan dari para responden. Metode evaluasi ini menggunakan berbagai metrik standar yaitu nilai akurasi, presisi, *recall*, *F1-Score* dan *confusion matrix* [24]. Askurasi menggambarkan performa keseluruhan model, sedangkan presisi dan recall mengukur ketepatan prediksi dan kemampuan model dalam menemukan kasus positif. F1-Score memberikan keseimbangan antara kedua metrik tersebut. *Confusion matrix* memberikan wawasan tentang jenis kesalahan yang dibuat oleh model. Dengan menggunakan metrik-metrik evaluasi ini secara bersama-sama, evaluasi memberikan gambaran yang komprehensif tentang kinerja model dalam mengidentifikasi sentimen, mendukung pengambilan keputusan yang akurat dalam penggunaan model di dunia nyata. *Confusion matrix* terdiri dari akurasi, presisi, *recall* dan *F1 score*. Berikut parameter untuk evaluasi model yang digunakan dalam penelitian ini.

1. *Accuracy*: Persentase prediksi yang benar dari semua prediksi yang dibuat oleh model dapat mengunakan persamaan dapat menggunakan persamaan berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

2. *Precision*: Proporsi prediksi positif yang benar dari semua prediksi positif yang dibuat oleh model dapat mengunakan persamaan dapat menggunakan persamaan berikut:

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

3. *Recall* : Proporsi prediksi positif yang benar dari semua kasus sebenarnya positif dapat mengunakan persamaan dapat menggunakan persamaan berikut:

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

4. F1-Score: Harmonik rata-rata dari presisi dan recall, memberikan keseimbangan antara keduanya dapat mengunakan persamaan dapat menggunakan persamaan berikut:

$$F1 - Score = 2X \frac{Precision\ X\ Recall}{Precision+Recall} \quad (7)$$

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3. HASIL DAN PEMBAHASAN

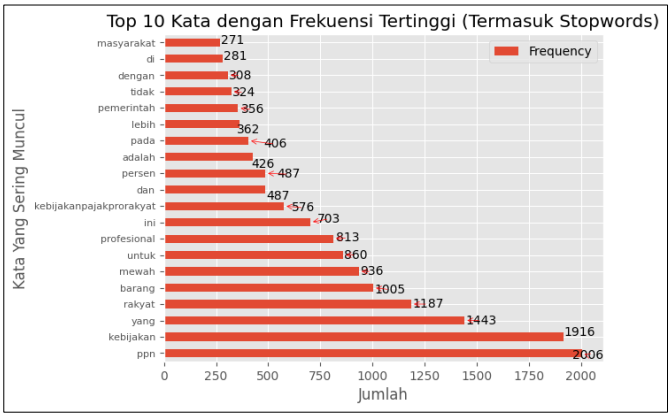
3.1 Pengumpulan Data

Pengumpulan data dilakukan dengan metode *crawlin*g yang dikumpulkan dari media sosial X menggunakan *Python* di *google colaboratory*. *Keyword* yang digunakan dalam pengumpulan data meliputi "Kebijakan PPN 12%" dan "PPN 12%". Jumlah tweet sebanyak 1815 tweet dengan data yang terdiri dari rentang waktu antara tahun 2023 hingga 2024. Data tersebut kemudian disimpan dalam format *Excel* sebelum diolah lebih lanjut untuk analisis sentimen. Proses *crawlling* dilakukan menggunakan *google colab* dengan bahasa pemrograman *Python* menggunakan *library* pandas sebagai dasar perintah untuk pengambilan data. Dataset yang digunakan dalam pemodelan dapat dilihat pada Tabel 1.

Tabel 1. Sampel Hasil Pengumpulan Data

Username	Teks Tweet
DroneEmpritOffc	Pemerintah dapat apresiasi di sektor kesehatan dan ekonomi namun kebijakan seperti PPN 12% picu ketidakpuasan. Isu HAM kebebasan berpendapat dan kabinet gemuk perburuk citra pemerintah. Perlu komunikasi lebih efektif karena ketidakselarasan citra positif dan realitas ini.

Dari data yang tersedia, terdapat kata-kata yang sering muncul yang ditweet akan ditampilkan pada Gambar 3.



Gambar 3. Frekuensi Kata Sebelum Preprocessing

Berdasarkan analisis frekuensi kata, ditemukan bahwa kata ”ppn” merupakan kata yang paling sering muncul dengan jumlah 2006 kali, diikuti oleh kata ”kebijakan” sebanyak 1916 kali dan kata ”yang” sebanyak 1443 kali. Kata-kata ini menunjukan fokus utama pembicaraan berkisar pada topik ppn, dengan penggunaan kata hubungan dan keterangan tempat. Selain itu , kata seperti ”rakyat”(1187 kali), ”barang”(1005 kali) dan ”mewah”(936 kali).

3.2 Preprocessing Data

Preprocessing merupakan langkah penting untuk membersihkan dan menyiapkan data teks sebelum melakukan analisis. Tujuan pada tahap ini untuk meningkatkan kualitas data, menghapus data yang tidak relevan, dan memformat teks agar leibih efisien. Selain itu tahap *preprocessing* untuk mengubah data menjadi format yang lebih mudah dan efektif untuk pengolahan selanjutnya. Berikut adalah beberapa langkah umum dalam *preprocessing* data untuk analisis sentimen. Gambar 4 merupakan tahap *wordcloud* sebelum *preprocessing.*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Gambar 4. Wordcloud Sebelum Preprocessing

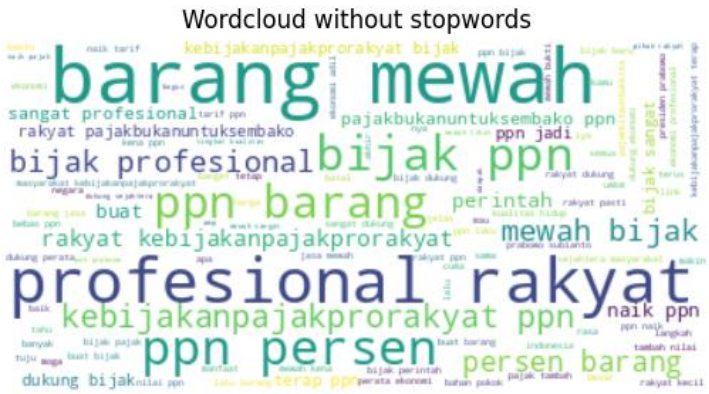
Pada Tabel 2 menampilkan setiap tahapan dan hasil dari *preprocessing* dari salah satu contoh teks tweet.

Tabel 2. Data Tweet

Proses	Output
Teks Asli	Pemerintah dapat apresiasi di sektor kesehatan dan ekonomi namun kebijakan seperti PPN 12% picu ketidakpuasan. Isu HAM kebebasan berpendapat dan kabinet gemuk perburuk citra pemerintah. Perlu komunikasi lebih efektif karena ketidakselarasan citra positif dan realitas ini.
Cleaning	Pemerintah dapat apresiasi di sektor kesehatan dan ekonomi namun kebijakan seperti ppn picu ketidakpuasan isu ham kebebasan berpendapat dan kabinet gemuk perburuk citra pemerintah perlu komunikasi lebih efektif karena ketidakselarasan citra positif dan realitas ini.
Case Folding	Pemerintah dapat apresiasi di sektor kesehatan dan ekonomi namun kebijakan seperti ppn picu ketidakpuasan isu ham kebebasan berpendapat dan kabinet gemuk perburuk citra pemerintah perlu komunikasi lebih efektif karena ketidakselarasan citra positif dan realitas ini.
Tokenize	['pemerintah', 'dapat', 'apresiasi', 'di', 'sektor', 'kesehatan', 'dan', 'ekonomi', 'namun', 'kebijakan', 'seperti', 'ppn', 'picu', 'ketidakpuasan', 'isu', 'ham', 'kebebasan', 'berpendapat', 'dan', 'kabinet', 'gemuk', 'perburuk', 'citra', 'pemerintah', 'perlu', 'komunikasi', 'lebih', 'efektif', 'karena', 'ketidakselarasan', 'citra', 'positif', 'dan', 'realitas', 'ini']
Stopword	['pemerintah', 'apresiasi', 'sektor', 'kesehatan', 'ekonomi', 'kebijakan', 'ppn', 'picu', 'ketidakpuasan', 'isu', 'ham', 'kebebasan', 'berpendapat', 'kabinet', 'gemuk', 'perburuk', 'citra', 'pemerintah', 'perlu', 'komunikasi', 'efektif', 'ketidakselarasan', 'citra', 'positif', 'realitas']
Stemming	Perintah apresiasi sektor sehat ekonomi bijak ppn picu ketidakpuasan isu ham bebas dapat kabinet gemuk buruk citra perintah perlu komunikasi efektif ketidakselarasan citra positif realitas

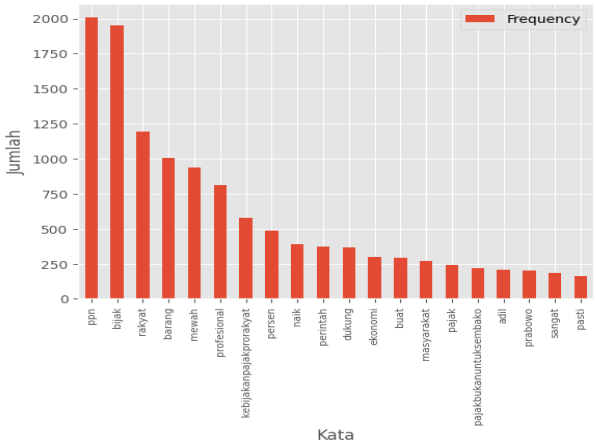
Tabel 2. menunjukan tahapan *preprocessing* data teks yang meliputi beberapa langkah utama *cleaning* untuk menghapus karakter khusus dan elemen tidak relevan, *case folding* untuk mengubah teks menjadi huruf kecil, *tokenize* untuk memisahkan teks menjadi daftar kata, *Stopword* untuk menghapus kata-kata umum yang tidak signifikan dan *stemming* untuk mengubah kata menjadi bentuk dasarnya. Proses ini bertujuan untuk meningkatkan kualitas data teks sehingga dapat digunakan secara efektif dalam analisis lebih lanjut pada tahap klasifikasi sentimen.*Wordcloud* setelah *preprocessing* dapat dilihat pada Gambar 5.

p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 5. Wordcloud Setelah Preprocessing

Gambar 5. merupakan *wordcloud* yang dihasilkan setelah proses *preprocessing* data teks. *Wordcloud* menampilkan kata-kata yang paling sering muncul dalam data dengan ukuran huruf yang menampilkan frekuensi kemunculannya. Kata-kata seperti "ppn," "profesional," "rakyat," "barang," dan "mewah" muncul dengan ukuran besar, menunjukkan bahwa topik pembicaraan utama berfokus kebijakan pajak pertambahan nilai (ppn) 12 % di Indonesia. *Word cloud* membantu memvisualisasikan tema dominan dalam data dengan cara yang mudah dipahami. Frekuensi kata dapat dilihat pada Gambar 6



Gambar 6. Frekuensi Kata Setelah Dilakukan Preprocessing Data

Setelah melalui proses *preprocessing* frekuensi kata yang paling sering muncul menunjukkan fokus utama pada topik Pajak Pertambahan Nilai. Kata seperti “ppn” “bijak” “rakyat” dan “barang” memiliki frekuensi tinggi. Analisis ini memperkuat relevansi data dengan tema kebijakan pajak pertambahan nilai di Indonesia.

3.3 Labelling Data

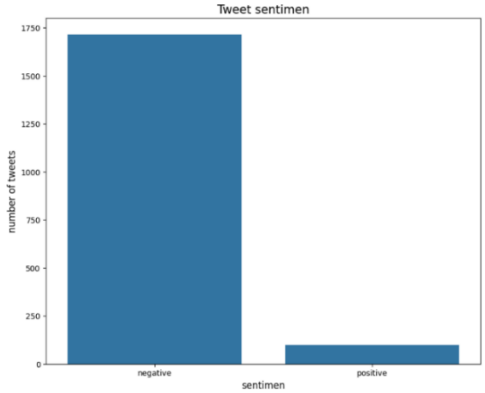
Setelah melalui tahap *pre-processing* data yang telah dikumpulkan sebanyak 2827 dan akan melalui proses pelabelan dengan cara membagi data menjadi 2 kategori sentimen positif dan negatif. Hasil pelabelan data dilihat pada Tabel 3**.**

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Tabel 3. Hasil Dari Tahapan Labelling

Tweet	Skor Sentiment	Pelabelan Sentiment
Warganet juga banyak bahas beberapa kebijakan yang dianggap kontroversial misalnya program makan bergizi gratis kabinet gemuk dan kebijakan ppn	5.46%	POSITIVE
pemerintah dapat apresiasi di sektor kesehatan dan ekonomi namun kebijakan seperti ppn picu ketidakpuasan isu ham kebebasan berpendapat dan kabinet gemuk perburuk citra pemerintah perlu komunikasi lebih efektif karena ketidakselarasan citra positif dan realitas ini	94.54%	NEGATIVE
	0%	NEUTRAL

Tahap selanjutnya melakukan proses *labelling* dimana data dibagi menjadi dua kategori label yaitu positif dan negatif. Sentimen terdiri dari 1715 bernilai negatif dan 99 kumpulan label bernilai positif sedangkan 0 bernilain netral. Hasil jumlah sentimen pada komentar dan presentase nilai sentimen bisa dilihat pada Gambar 7.



Gambar 7. Perbandingan Sentimen Positif dan Negatif

Selanjunya data yang telah dilabeli kemudian dianalisis secara mendalam untuk mengetahui pembagianjumlah sentimen dalam dataset yang tersedia. Analisis ini sangat penting untuk memahami pola dan proporsi sentimen positif, negatif, dan netral yang muncul dalam diskusi atau pembahasan terkait Pajak Pertambahan Nilai di *platform* media sosial X. Hasil pembagian memberikan gambaran yang lebih jelas tentang pandangan umum, baik dukungan, kritik, maupun opini netral yang berhubungan dengan kebijakan pajak pertambahan nilai di indonesia.

3.4 Pembobotan TF-IDF

Proses selanjutnya pemberian nilai terhadap setiap term pada komentar yang telah melewati proses sebelumnya yaitu *preprocessing*. Tujuan dari pembobotan pemberian nilai pada *term* akan digunakan sebagai input pada proses klasifikasi. Gambar 8 menunjukkan hasil pembobotan dengan menggunakan TF – IDF.

	02	10	100	11	12	16	2021	2024	2025	30	...	ya	yaa	yacht	yakin	yang	yg	yong	youtube	yuk	zalim
0	0.0	0.0	0.000000	0.0	0.041916	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.000000	0.0	0.034210	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.000000	0.0	0.050514	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.085034	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.227598	0.0	0.036428	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.061323	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.000000	0.0	0.041684	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.140339	0.0	0.0	0.0	0.0	0.0

Gambar 8. Hasil Pembobotan TF-IDF

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.5 Klasifikasi Naive Bayes

Hasil klasifikasi menggunakan model *Naive Bayes Classifier* (NB) sebelum dilakukan SMOTE dan sesudah proses SMOTE dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Naive Bayes Classifier*

	Accuracy	**Precision**	**Recall**	**F1-Score**
Sebelum SMOTE	0.9272	0.5732	0.5378	0.5474
Setelah SMOTE	0.9268	0.9287	0.9268	0.9267

Hasil evaluasi model *Naive Bayes Classifier* (NB) menunjukkan beberapa hal menarik. Nilai akurasi model NB sebelum SMOTE dengan nilai akurasi sebesar 92,72% dapat memprediksi sentimen tweet dengan benar dari total data uji. Sementara itu nilai presisi sebesar 57,32%, recall bernilai 53,78, *F1 Score* bernilai 54,74% yang merepresentasikan kualitas prediksi model secara keseluruhan. Kemudian setelah mengalami proses SMOTE maka nilai akurasi model *Naive Bayes Classifier* (NB) sebesar 92,68%, presisi dengan nilai 92,87%, recall dengan nilai 92,68% dan *F1 Score* bernilai 92,67. Berdasarkan hasil klasifikasi tersebut menunjukkan bahwa kinerja model menggunakan naive bayes seblum SMOTE menunjukkan kinerja kurang baik memadai dalam menangani kelas yang tidak seimbang. Akurasi yang tinggi disebabkan oleh keberhasilan model memprediksi kelas moyoritas sedangkan rendahnya nilai presisi, recall dan F1 score menunjukkan bahwa model tidak mempu memprediksi kelas miniritas secara efektif. Sehingga penting melakukan penyeimbangan data menggunakan SMOTE dalam meningkatkan kinerja model untuk data yang tidak seimbang. Kemudian kinerja model setelah mengalami proses SMOTE menunjukkan hasil kinerja yang signifikan dalam presisi, recall, dan F1 Score. Ini menunjukkan bahwa proses SMOTE berhasil meningkatkan kemampuan model dalam menangani kelas minoritas sehingga memberikan hasil prediksi yang lebih handal dan mencerminkan untuk seluruh data.

3.6 Klasifikasi Decission Tree

Hasil klasifikasi menggunakan model *Decission Tree Classifier* (TD) sebelum dilakukan SMOTE dan sesudah proses SMOTE dapat dilihat pada Tabel 5.

Tabel 5. Hasil Decission Tree Classifier

	Accuracy	**Precision**	**Recall**	**F1-Score**
Sebelum SMOTE	0.9762	0.9434	0.8212	0.8675
Setelah SMOTE	0.9344	0.9362	0.9344	0.9343

Hasil evaluasi model Decission Tree (NB) menunjukkan beberapa hal menarik. Nilai akurasi model Decission Tree sebelum SMOTE dengan nilai akurasi sebesar 97,62% dapat memprediksi sentimen tweet dengan benar dari total data uji. Sementara nilai presisi 94,34%, Recall bernilai 82,12% dan nilai F1 Score adalah 86,75% yang merepresentasikan kualitas prediksi model secara keseluruhan. Kemudian setelah mengalami proses SMOTE maka nilai akurasi model Decission Tree Classifier sebesar 93,44%, presisi bernilai 93,62%, recall bernilai 93,44% dan F1 Score bernilai 93,43% ini menunjukkan bahwa model memiliki kualitas prediksi setelah melalui proses SMOTE. Hasil evaluasi model Decision Tree memberikan wawasan mendalam tentang performa model sebelum SMOTE dan sesudah SMOTE. Model sebeblum SMOTE menunjukkan bahwa hasil prediksi sangat akurat dalam memprediksi kelas mayoritas, tetapi tidak terlalu baik dalam menangani kelas minoritas, seperti yang tertermin dari nilai recall yang lebih rndah. Model setelah SMOTE menunjukkan bahwa model sedikit mengalami penurunan dalam akurasi, namun kualitas keseluruhan prediksi meningkat secara signifikan terutama dalam hal mendeteksi kelas minoritas yang terlihat dari kenaikan pada nilai recall dan F1 Score, hal ini menunjukkan bahwa model menjadi lebih seimbang dalam memprediksi semua sentimen dengan lebih baik.

3.7 Evaluasi Kinerja Model

Evaluasi kinerja model menggunakan c*onfusion matrix* yang merupakan alat pengukuran dalam bentuk matrix yang dipakai untuk memperoleh tingkat keakuratan klasifikasi kelas-kelas berdasarkan algoritma yang digunakan. Evaluasi kinerja model *Naive Bayes* dan *Decission Tree* sebelum SMOTE ditunjukkan pada Tabel 6.

Tabel 6. Evaluasi kinerja *Naive Bayes* dan *Decission Tree* sebelum SMOTE

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Model	Akurasi	Presisi	Recall	F1-score
Naive Bayes	92,72%	57,32%	53,78%	54,74%
Decission Tree	97,62%	94,34%	82,12%	86,75%

Dari data yang sudah diperoleh pada proses pelabelan menunjukkan sentimen negatif sebanyak 94,54% sementara sentimen positif sebanyak 5,46%. ketidakseimbangan antara 2 kelas tersebut diatasi melalui optimasi metode *SMOTE* untuk menyeimbangkan jumlah data antara positif, negatif dan netral. Kesimbangan ini membantu model dalam mempelajari karakteristik dari setiap kelas secara lebih seimbang. *Dalam* tahap evalusi dua algoritma klasifikasi yaitu Naive Bayes dan decission Tree menunjukkan bahwa *Decission Tree* memberikan jumlah paling tinggi daripada Naive Bayes. Proses ini membandingkan pembagian data sebanyak 80% untuk training serta 20% untuk testing. *Decission Tree* memberikan akurasi 97,62% dan *Naive Bayes* memberikan akurasi sebanyak 92,72%. Berdasrkan model algoritma yang diterapkan, model ini menggunakan data yang telah dioptimalkan dengan metode *SMOTE* serta membandingkan dengan data yang belum optimal. Kemudian evaluasi kinerja model *Naive Bayes* dan *Decission Tree* setelah SMOTE ditunjukkan pada Tabel 7.

Tabel 7. Evaluasi kinerja *Naive Bayes* dan *Decission Tree* setelah SMOTE

Model	Akurasi	Presisi	Recall	F1-score
s	92,68%	92,87%	92,68%	92,67%
Decission Tree	93,44%	93,62%	93,44%	93,43%

Setelah perbandingan *SMOTE* algoritma *Naive Bayes* menunjukan peningkatan dalam skor presisi, *recall*, serta *F1-score*, namun pada skor akurasi mengalami penurunan. Kemudian algoritma *Decission Tree* menunjukkan peningkatan dalam skor *recall* serta *F1-score*, namun pada skor akurasi serta presisi mengalami penurunan.

Penelitian ini menunjukkan hasil bahwa akurasi model Decision Tree cenderung(93,44%) lebih tinggi dibandingkan dengan menggunakan model Naive Bayes (92,68%). Namun demikian model *Naive Bayes* lebih efisien dalam menangani dataset yang lebih besar. Dari seluruh tweet yang dianalisis sebanyak 94,54% mengandung sentimen negatif terkait kekhawatiran tentang dampak ekonomi dan peningkatan beban pajak sebagaimana sejalan dengan penelitian sebelumnya yang membahas kebijakan kenaikan pajak dapat memengaruhi ekonomi perekonomian [9]-[10]. Pemilihan model *Naive Bayes* dan *Decision Tree* pada penelitian ini menunjukkan hasil akurasi lebih tinggi dibandingkan dengan penelitian sebelumnya yang menggunakan metode KNN yaitu sebesar 66,67% [12] dan penelitian yang membahas tentang kebijakan kenaikan pajak menggunakan metode Bert dengan hasil akurasi sebesar 83% [11]. Penelitian ini juga mengungguli penelitian sebelumnya menggunakan algoritma decission tree dengan nilai akurasi 81,34% [7]. Penelitian ini memberikan wawasan bagi pembuat kebijakan mengenai persepsi publik terkait Pajak Pertambahan Nilai 12% dan menawarkan strategi komunikasi yang lebih efektif.

4. KESIMPULAN

Penelitian ini menunjukkan hasil bahwa model Decision Tree memiliki tingkat akurasi yang lebih tinggi (93,44%) dibandingkan dengan Naive Bayes (92,68%) dalam melakukan analisis sentimen terkait kebijakan PPN 12%. Sebagian besar opini publik yang dianalisis menunjukkan sentimen negatif yaitu sebesar 94,54% yang didominasi oleh kekhawatiran terhadap dampak ekonomi dan peningkatan beban pajak, sementara sentimen positif hanya mencakup 5,46%, dengan fokus pada potensi peningkatan pendapatan negara dan pembangunan. Sentimen netral tidak ditemukan dalam data yang dianalisis. Meski model Decision Tree lebih baik dalam mengenali pola yang kompleks, namun Naive Bayes menunjukkan efisiensi dalam menangani sentimen yang lebih seimbang dan memiliki kecepatan pengolahan data yang lebih tinggi. Penelitian ini memberikan wawasan berharga bagi para pembuat kebijakan untuk lebih memahami persepsi masyarakat serta mengembangkan strategi komunikasi yang lebih efektif guna meningkatkan dukungan publik terhadap kebijakan perpajakan. Sebagai rekomendasi penelitian selanjutnya diusulkan untuk menggunakan metode deep learning untuk mendapatkan analisis sentimen yang lebih mendalam dan hasil yang lebih akurat.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

DAFTAR PUSTAKA

[1] S. F. Zis, N. Effendi, and E. R. Roem, "Perubahan Perilaku Komunikasi Generasi Milenial dan Generasi Z di Era Digital," *Satwika Kaji. Ilmu Budaya dan Perubahan Sos.*, vol. 5, no. 1, pp. 69–87, 2021, doi: 10.22219/satwika.v5i1.15550.

[2] A. Islamiati and N. Amalia, "Pengaruh Media Sosial X dalam Cerita Alternate Universe pada Minat Baca Generasi Z," *Edukatif J. Ilmu Pendidik.*, vol. 6, no. 4, pp. 3926–3934, 2024, doi: 10.31004/edukatif.v6i4.7423.

[3] R. N. Muhammad, L. W. S, and B. Tanggahma, "Pengaruh Media Sosial Pada Persepsi Publik Terhadap Sistem Peradilan : Analisis Sentimen di Twitter," vol. 7, no. 1, pp. 507–516, 2024, doi: https://doi.org/10.31933/unesrev.v7i1.2327.

[4] V. A. Sulistiani and M. Hamka, "Analisis Sentimen Pengguna Media Sosial Terhadap Identitas Kependudukan Digital Menggunakan Metode Support Vector Machine (SVM)," *J. Inf. Syst. ...*, vol. 9, no. 4, pp. 2185–2195, 2024, doi: http://orcid.org/0000-0002-0984-3608.

[5] F. Syarief, "Pemanfaatan Media Sosial Dalam Proses Pembentukan Opini Publik," *J. Komun.*, vol. 8, no. 3, pp. 262–266, 2017, doi: https://doi.org/10.31294/jkom.v8i3.3092.

[6] J. Akuntansi, D. A. N. Keuangan, N. Fauziah, M. Alkautsar, Y. Suryaman, and F. F. Roji, "Pelabelan VADER Dalam Menganalisis Persepsi Masyarakat Terhadap Kenaikan Tarif PPN di Indonesia," vol. 12, no. 2, pp. 228–238, 2024, doi:10.29103/jak.v12i2.16796.

[7] I. D. Hardyatman and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Rencana Kenaikan PPN 12 % Di Indonesia Pada Media Sosial X Menggunakan Metode Decision Tree," vol. 6, no. 2, pp. 1128–1136, 2025, doi: 10.47065/josh.v6i2.6573.

[8] W. A. Anggraeni, F. Fahru Roji, and M. Alkautsar, "Analisis Sentimen Publik terhadap Kebijakan Insentif Perpajakan Dengan Pendekatan VADER (Valence Aware Dictionary And Sentiment Reasoner)," *J. Proaksi*, vol. 10, no. 4, pp. 465–477, 2023, doi: 10.32534/jpk.v10i4.4732.

[9] A. Aulia, S. Maisaroh, A. F. Ananta, and W. Pangestoeti, "Dampak Kenaikan PPN 12 % terhadap Pendapatan Negara dan Kesejahteraan Masyarakat," no. 42, 2025, doi: https://doi.org/10.62383/.

[10] I. M. Putri, "Kenaikan Ppn 12% Dan Dampaknya Terhadap Eknomi," *J. Ilm. Manajemen, Ekon. Akunt.*, vol. 8, no. 2, pp. 934–944, 2024, doi: 10.31955/mea.v8i2.4077.

[11] F. Imawan, D. F. Shiddieq, F. F. Roji, Analisis Sentimen Publik di X Terhadap Rencana Kenaikan PPN 12% Menggunakan Bert Engineering, Journal of Computer Engineering, System and Science, vol. 10, no. 1, pp. 136–148, 2025, doi: 10.24114/cess.v10i1.62392.

[12] T. Ngongo *et al.*, "Kendaraan Bermotor Dengan Metode KNN," vol. 8, no. 6, pp. 11946–11949, 2024, doi: https://doi.org/10.36040/jati.v8i6.11685.

[13] P. S. Zakaria, R. Julianto, and R. S. Bernada, "Implementasi Naive Bayes Menggunakan Python dalam Klasifikasi Data," *BIIKMA Bul. Ilm. Ilmu Komput. dan Multimed.*, vol. 1, no. 2, pp. 126–131, 2023.

[14] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.

[15] 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, doi: 10.34148/teknika.v10i1.311.

[16] A. Kartika Sari, Akhmad Irsyad, Dinda Nur Aini, Islamiyah, and Stephanie Elfriede Ginting, "Analisis Sentimen Twitter Menggunakan Machine Learning untuk Identifikasi Konten Negatif," *Adopsi Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 64–73, 2024, doi: 10.30872/atasi.v3i1.1373.

[17] N. Nurdin, L. Jama, T. Z. Magnus, R. Priskila, and V. H. Pranatawijaya, "Analisis Sentimen Dampak Artificial Intelligence (AI) Untuk Pendidikan Pada X Menggunakan Naïve Bayes," *J. Inform. Upgris*, vol. 10, no. 1, pp. 15–19, 2024, doi: 10.26877/jiu.v10i1.18867.

[18] I. Indriati and A. Ridok, "Sentiment Analysis for Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (Nwknn)," *J. Enviromental Eng. Sustain. Technol.*, vol. 3, no. 1, pp. 23–32, 2016, doi: 10.21776/ub.jeest.2016.003.01.4.

[19] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 452–458, 2023, doi: 10.36040/jati.v7i1.6303.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[20] C. A. Sugianto and F. R. Maulana, "Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama)," *Techno.Com*, vol. 18, no. 4, pp. 321–331, 2019, doi: 10.33633/tc.v18i4.2587.

[21] Mohamad Ripai, Umi Hayati, Wita Widyawati, Heliyanti Susana, and Fathurrohman, "Klasifikasi Surat Pemberitahuan Pajak Daerah Menggunakan Metode Regresi Logistik Biner Untuk Mengetahui Patuh Dan Tidak Patuh Dalam Pembayaran Pajak Daerah," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 6, no. 1, pp. 27–33, 2022, doi: 10.32485/kopertip.v6i1.128.

[22] R. Ridwan, E. H. Hermaliani, and M. Ernawati, "Penerapan: Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian," *Comput. Sci.*, vol. 4, no. 1, pp. 80–88, 2024, doi: 10.33005/scan.v15i1.1850

[23] A. A. Arifiyanti and E. D. Wahyuni, "Smote: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining," *SCAN - J. Teknol. Inf. dan Komun.*, vol. 15, no. 1, 2020, doi: 10.33005/scan.v15i1.1850.

[24] S. Nor, M. A. Muslim, and M. Aswin, "Pengenalan Pola Dasar Angka berdasarkan Gerakan Tangan menggunakan Machine Learning," vol. 10, no. 3, pp. 595–608, 2022, doi: https://doi.org/10.26760/elkomika.v10i3.595.