

Pemanfaatan Data X Untuk Analisis Sentimen Publik Pada Kebijakan PPN 12%

Restu Aprianto Atmiko¹, Melaya Priangga², Nabil Panca Budhi Hizbullah³, Muhammad Anis Nur Fauzi⁴,
Dzulfan Yumna Azis⁵

^{1,2,3,4}Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Jawa Tengah, Indonesia

⁵Program Studi Sistem Informasi, Fakultas Rekayasa Industri, Universitas Telkom Purwokerto, Jawa Tengah, Indonesia

surel:¹restuaprianto15@gmail.com,²m.pnng4@gmail.com,³pancabudhi22@gmail.com,⁴anisnurfauzi8@gmail.com,⁵dzulfanyumna@gmail.com

Info Artikel

Sejarah artikel:

Diterima 20-01-2025

Revisi 25-02-2025

Diterima 14-04-2025

Kata kunci:

PPN 12%

Analisis Sentimen

Teks Mining

Data X

Dampak Ekonomi

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi sentimen masyarakat terhadap kebijakan pemerintah terkait penerapan Pajak Pertambahan Nilai (PPN) sebesar 12% dengan menggunakan teknik text mining. Kebijakan perpajakan sering menjadi topik diskusi luas di media sosial, forum online, dan platform berita digital, menghasilkan data tidak terstruktur yang mencerminkan opini publik. Penelitian ini menganalisis opini, komentar, dan diskusi masyarakat mengenai kebijakan PPN 12% dari berbagai sumber digital. Proses dimulai dengan Prapemrosesan *Data* untuk menghapus elemen yang tidak relevan seperti tanda baca, tautan, dan stop words, serta melakukan normalisasi teks. Analisis sentimen dilakukan menggunakan metode *machine learning* berbasis Naive Bayes dan pendekatan lexicon-based. Data diklasifikasikan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Analisis lanjutan dilakukan untuk mengidentifikasi pola, kata kunci, dan tema utama yang muncul dari opini publik. Hasil analisis menunjukkan bahwa sentimen masyarakat terhadap kebijakan PPN 12% didominasi oleh respons negatif, terutama terkait kekhawatiran terhadap daya beli, dampak ekonomi, dan ketimpangan sosial. Namun, terdapat juga sentimen positif yang menyoroti potensi peningkatan pendapatan negara untuk pembangunan. Penelitian ini memberikan wawasan penting bagi pemerintah dalam meningkatkan komunikasi publik, memperkuat transparansi kebijakan, dan menyusun kebijakan yang lebih responsif terhadap kebutuhan masyarakat melalui pendekatan berbasis data.

Penulis Korespondensi:

Restu Aprianto Atmiko

Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto

Email: restuaprianto15@gmail.com

1. PENDAHULUAN

Penerapan Pajak Pertambahan Nilai (PPN) sebesar 12% oleh pemerintah Indonesia telah menjadi isu yang memicu diskusi luas di kalangan masyarakat. Kebijakan ini dianggap sebagai langkah strategis untuk meningkatkan pendapatan negara guna mendukung pembangunan infrastruktur, pendidikan, dan sektor publik lainnya. Namun, kebijakan ini juga menuai kritik, terutama dari kelompok ekonomi menengah ke bawah yang merasa khawatir terhadap

kenaikan harga barang dan jasa. Kekhawatiran tersebut mencakup penurunan daya beli, peningkatan kesenjangan sosial, dan potensi dampak negatif terhadap perekonomian masyarakat kecil. Dalam konteks ini, analisis terhadap reaksi masyarakat menjadi hal yang penting untuk memahami persepsi publik terhadap kebijakan PPN 12%.

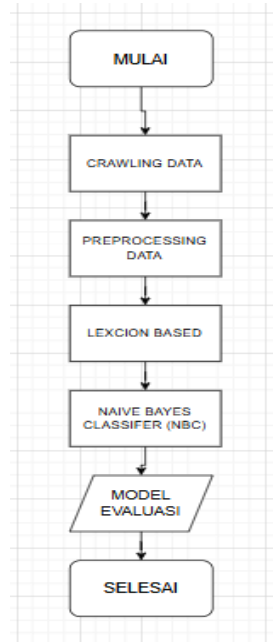
Kebijakan perpajakan seperti PPN 12% sering kali memunculkan berbagai opini dari masyarakat, baik yang mendukung maupun yang menolak. Pihak yang mendukung kebijakan ini umumnya menyoroti pentingnya peningkatan pendapatan negara untuk mendukung pembangunan nasional. Sebaliknya, pihak yang menolak menekankan dampak negatifnya terhadap perekonomian masyarakat, terutama di tengah situasi pemulihan ekonomi pasca pandemi COVID-19. Situasi ini menunjukkan adanya kebutuhan untuk memetakan sentimen publik dengan lebih akurat agar kebijakan dapat disesuaikan dengan kebutuhan masyarakat. Platform digital seperti media sosial, forum diskusi, dan portal berita menjadi sarana utama masyarakat untuk mengekspresikan pendapat mereka. Data yang dihasilkan dari diskusi ini mencerminkan sentimen publik, baik itu positif, negatif, maupun netral. Namun, sifat data yang tidak terstruktur, beragamnya gaya bahasa, dan adanya elemen-elemen tidak relevan seperti spam, tautan, atau simbol menjadi tantangan tersendiri dalam pengolahan data. Dalam konteks ini, penggunaan teknik analisis teks berbasis *machine learning* seperti algoritma Naive Bayes menjadi solusi yang potensial. Algoritma ini mampu menangani data dalam jumlah besar, mengelompokkan sentimen dengan tingkat akurasi yang baik, dan memetakan persepsi masyarakat secara lebih terstruktur.

Berikut ini adalah beberapa penelitian terkait yaitu Setiawan et al. (2022) melakukan analisis sentimen masyarakat terhadap kebijakan kenaikan tarif listrik di Indonesia menggunakan algoritma Naive Bayes. Studi ini menemukan bahwa sentimen negatif muncul terutama karena ketidakpuasan terhadap kurangnya informasi dan dampak terhadap biaya hidup masyarakat [1]. Rahmawati dan Nugroho (2021) menganalisis opini publik terhadap kebijakan subsidi BBM di Indonesia. Dengan pendekatan teks mining berbasis Naive Bayes, penelitian ini mengidentifikasi bahwa faktor sosial seperti tingkat pendidikan dan akses informasi sangat mempengaruhi persepsi masyarakat [2]. Fauzan et al. (2023) menggunakan algoritma Naive Bayes untuk mengevaluasi sentimen masyarakat terhadap kebijakan pendidikan gratis. Penelitian ini mengungkapkan bahwa meskipun kebijakan diterima dengan baik oleh mayoritas, terdapat kelompok yang khawatir terhadap kualitas layanan pendidikan [3]. Susanto et al. (2022) mengevaluasi reaksi publik terhadap reformasi perpajakan di Indonesia menggunakan data dari media sosial. Penelitian ini menunjukkan bahwa sentimen negatif sering kali disebabkan oleh kurangnya sosialisasi yang efektif dari pemerintah [4]. Pratama dan Yulianto (2021) mengembangkan sistem analisis sentimen terhadap kebijakan publik di Indonesia, dengan fokus pada isu-isu terkait kesejahteraan sosial. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memberikan hasil yang signifikan dalam pengelompokan opini, meskipun tantangan teknis dalam Prapemrosesan *Data* tetap menjadi hambatan [5]. Putri et al. (2023) menganalisis persepsi masyarakat terhadap kebijakan kesehatan dengan pendekatan Naive Bayes, menyoroti pentingnya pengelolaan data multibahasa untuk meningkatkan akurasi analisis sentimen [6]. Hidayat dan Wibisono (2022) mengevaluasi opini masyarakat terhadap kebijakan transportasi publik di Indonesia. Penelitian ini menunjukkan bahwa faktor kenyamanan dan biaya sangat mempengaruhi pola sentimen [7]. Andini et al. (2021) melakukan kajian sentimen masyarakat terhadap kebijakan digitalisasi pajak di Indonesia. Penelitian ini menggunakan Naive Bayes dan menemukan bahwa masyarakat cenderung menerima kebijakan ini jika disertai edukasi yang memadai [8]. Nurhayati et al. (2022) meneliti persepsi masyarakat terhadap kebijakan insentif pajak dengan analisis sentimen berbasis Naive Bayes, yang menunjukkan bahwa keberhasilan kebijakan bergantung pada kejelasan informasi yang disampaikan [9]. Utami dan Kurniawan (2023) menganalisis reaksi publik terhadap kebijakan ekonomi makro Indonesia melalui teks mining. Studi ini menemukan bahwa kombinasi metode *machine learning* dengan *lexicon-based* meningkatkan akurasi klasifikasi sentimen [10].

2. METODE

Penelitian ini bertujuan untuk mengevaluasi sentimen masyarakat Indonesia terhadap kebijakan Pajak Pertambahan Nilai (PPN) 12% menggunakan algoritma Naive Bayes, dengan pendekatan analisis teks dari data opini yang diperoleh dari berbagai platform digital. Metodologi penelitian ini dirancang untuk mengidentifikasi sentimen publik (positif, negatif, netral), menemukan pola-pola utama dalam opini masyarakat, serta memberikan rekomendasi kepada pemerintah untuk meningkatkan komunikasi kebijakan dan menyusun strategi adaptasi kebijakan yang lebih

sesuai dengan kebutuhan masyarakat. Gambar 1 berikut merupakan proses pengklasifikasian dalam penelitian ini.



Gambar 1. *Flowchart* klasifikasi

Penelitian ini dilakukan dengan langkah-langkah rinci sebagai berikut.

2.1 Identifikasi Sumber Data

Sumber data untuk penelitian ini akan diperoleh dari berbagai platform digital yang digunakan oleh masyarakat Indonesia untuk mengekspresikan pendapat terkait kebijakan Pajak Pertambahan Nilai (PPN) 12%. Platform utama yang menjadi sumber data antara lain media sosial, forum diskusi, portal berita, serta blog dan website yang membahas isu terkait. Media sosial seperti Twitter, Facebook, Instagram, dan YouTube merupakan saluran utama bagi masyarakat untuk berbagi opini secara cepat dan luas. Dalam hal ini, data yang akan dikumpulkan mencakup tweet, status, komentar, dan video yang berkaitan dengan kebijakan PPN 12%. Forum diskusi seperti Kaskus dan Reddit juga akan menjadi sumber data penting karena sering kali menjadi tempat diskusi mendalam mengenai kebijakan publik. Selain itu, portal berita seperti Kompas.com, Detik.com, dan media online lainnya yang memungkinkan pengguna untuk memberikan komentar pada artikel-artikel yang membahas PPN 12% akan digunakan sebagai sumber data tambahan. Blog dan website yang mengulas kebijakan PPN 12% juga akan dianalisis, terutama artikel yang berisi opini atau ulasan dari masyarakat. Pengumpulan data akan difokuskan pada periode enam bulan setelah pemberlakuan kebijakan PPN 12% pada tahun 2024, untuk memastikan data yang diperoleh relevan dan mencerminkan persepsi terkini dari masyarakat Indonesia terhadap kebijakan tersebut.

2.2. Pengumpulan Data

Pengumpulan data dalam penelitian ini akan difokuskan pada platform Twitter, yang merupakan salah satu media sosial paling aktif digunakan masyarakat Indonesia untuk menyampaikan opini dan pandangan mereka secara real-time. Data akan dikumpulkan dengan menggunakan Twitter API. Pengumpulan data ini akan difokuskan pada tweet yang relevan dengan kebijakan Pajak Pertambahan Nilai (PPN) 12%, dengan menggunakan kata kunci seperti “PPN 12%”, “pajak”, “kenaikan harga”, “pajak pertambahan nilai”, dan istilah lain yang relevan. Selain itu, filter waktu akan diterapkan untuk memastikan hanya tweet yang diposting dalam enam bulan setelah pemberlakuan kebijakan PPN 12% yang diambil, sehingga data yang diperoleh relevan dengan konteks penelitian. Proses pengumpulan data juga akan mencakup metadata seperti tanggal posting, jumlah retweet, jumlah suka, dan tanggapan dari tweet tersebut untuk memberikan wawasan tambahan mengenai popularitas dan dampak setiap opini. Untuk

meningkatkan kualitas data, proses penyaringan akan dilakukan untuk menghilangkan spam, bot-generated tweets, dan konten yang tidak relevan. Semua data yang terkumpul akan disimpan dalam format terstruktur, seperti file CSV atau database SQL, untuk memudahkan analisis lebih lanjut. Prosedur ini dilakukan sesuai dengan kebijakan privasi dan etika platform Twitter, termasuk anonimisasi data untuk melindungi identitas pengguna.

2.3. Pra Proses Data

Data yang terkumpul dari berbagai platform digital akan melalui beberapa tahapan prapemrosesan untuk memastikan kualitas dan relevansi data yang akan dianalisis.

2.3.1. *Cleaning Data*

Menghapus elemen-elemen yang tidak relevan seperti URL, tautan, simbol (emoji), dan teks spam. Menyaring konten yang tidak berkaitan langsung dengan kebijakan PPN 12%, seperti komentar yang tidak relevan atau iklan. Langkah ini melibatkan:

1. Menghapus tautan (e.g., <https://t.co/...>) menggunakan ekspresi reguler.
2. Menghapus nama pengguna yang ditandai dengan @.
3. Menghapus tagar (#) sambil tetap mempertahankan kata kunci yang relevan (misalnya, #PPN12 menjadi PPN12).
4. Menghapus simbol dan karakter khusus, kecuali tanda baca yang relevan untuk analisis sentimen.

2.3.2. *Normalisasi Text*

Normalisasi dilakukan untuk mengonversi kata-kata tidak baku ke dalam bentuk standar. Misalnya, kata-kata seperti “gak”, “nggak”, dan “tidak” akan dinormalisasi menjadi “tidak”. Proses ini penting untuk mengatasi variasi bahasa informal di Twitter. Pustaka seperti Sastrawi atau daftar kata tidak baku-kata baku khusus Bahasa Indonesia dapat digunakan.

2.3.3. *Stopword Removal*

Stop words adalah kata-kata umum yang tidak memiliki makna signifikan untuk analisis sentimen, seperti “di”, “ke”, “yang”, “dan”, dll. Penghapusan dilakukan menggunakan daftar stopwords Bahasa Indonesia untuk mengurangi kebisingan dalam data.

2.3.4. *Tokenisasi*

Proses ini memecah teks menjadi unit-unit yang lebih kecil, yaitu kata atau token. Teknik ini biasanya dilakukan menggunakan pustaka Python seperti NLTK atau spaCy. Misalnya, kalimat “PPN 12% membuat harga naik” akan dipecah menjadi [“PPN”, “12%”, “membuat”, “harga”, “naik”].

2.3.5. *Stemming*

Dalam data Twitter, tweet yang sama dapat muncul beberapa kali, baik sebagai *retweet* maupun konten duplikat. Oleh karena itu, data akan diperiksa untuk mengidentifikasi dan menghapus entri yang redundan, memastikan bahwa analisis dilakukan pada data unik.

2.3.6. *Pemilihan Fitur*

Melakukan ekstraksi fitur dengan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk menghasilkan representasi vektor dari teks yang menggambarkan frekuensi kata-kata dalam data.

2.4. Analisis Sentimen

Tahap analisis sentimen bertujuan untuk mengklasifikasikan opini yang terkumpul dari Twitter ke dalam tiga kategori sentimen: positif, negatif, dan netral. Sentimen positif mencakup opini yang mendukung kebijakan PPN 12% atau memberikan pandangan optimis terkait dampaknya. Sentimen negatif mencerminkan opini yang menentang kebijakan, menunjukkan kekhawatiran terhadap dampaknya pada masyarakat atau ekonomi. Sementara itu, sentimen netral terdiri dari opini yang bersifat informatif, tidak jelas mendukung atau menentang kebijakan, atau tanpa muatan emosi tertentu. Untuk melakukan klasifikasi ini, algoritma Naive Bayes digunakan karena efisiensinya dalam mengelompokkan teks berdasarkan probabilitas. Model akan dilatih menggunakan dataset yang telah diproses, di mana data dibagi menjadi dua bagian: data latih untuk melatih model dan data uji untuk mengukur performanya. Evaluasi model dilakukan menggunakan metrik seperti precision, recall, F1-score, dan confusion matrix untuk memastikan akurasi dan keandalan klasifikasi sentimen. Dengan pendekatan ini, analisis dapat memberikan gambaran yang terstruktur dan mendalam mengenai opini publik terhadap kebijakan PPN 12%.

2.5. Pemodelan dan Evaluasi

Tahap pemodelan dan evaluasi dimulai dengan pembagian data yang telah dibersihkan dan diolah menjadi dua bagian: 80% sebagai data latih untuk melatih model, dan 20% sebagai data uji untuk mengevaluasi performa model. Algoritma Naive Bayes digunakan sebagai metode utama karena efisiensinya dalam mengklasifikasikan teks, dengan fitur ekstraksi teks menggunakan TF-IDF untuk mengidentifikasi hubungan antara kata-kata dalam twitter dan label sentimen [11]. Setelah model selesai dilatih menggunakan data latih, performa model diuji menggunakan data uji dengan berbagai metrik evaluasi. Metrik tersebut meliputi akurasi, untuk mengukur proporsi prediksi yang benar; precision, recall, dan F1-score, untuk mengevaluasi keseimbangan antara presisi dan sensitivitas model; serta confusion matrix, untuk memberikan gambaran detail tentang bagaimana model mengklasifikasikan data ke dalam tiga kategori sentimen (positif, negatif, netral) [12]. Pendekatan ini memungkinkan identifikasi kesalahan klasifikasi serta peluang untuk meningkatkan performa model di masa mendatang.

2.6. Analisis Hasil

Hasil klasifikasi sentimen dianalisis untuk mengidentifikasi pola opini masyarakat terhadap kebijakan PPN 12%. Analisis ini mencakup kategori utama sentimen, seperti kekhawatiran terhadap dampak ekonomi, kenaikan harga barang, dan tingkat penerimaan masyarakat terhadap kebijakan [13]. Jika data memungkinkan, sentimen akan dikategorikan berdasarkan faktor demografi, seperti usia, jenis kelamin, atau lokasi geografis, guna memahami perbedaan sentimen di antara kelompok masyarakat yang berbeda. Selain itu, teknik analisis tema digunakan untuk mengidentifikasi isu-isu utama yang menjadi perhatian publik, seperti dampak terhadap harga barang, ketidakpastian ekonomi, dan kebutuhan edukasi lebih lanjut terkait kebijakan PPN [14]. Pendekatan serupa telah diterapkan dalam penelitian lain yang mengkaji sentimen terhadap kebijakan publik, yang menunjukkan pentingnya analisis berbasis tema untuk menggali opini masyarakat secara mendalam [15]. Temuan ini memberikan wawasan mendalam tentang persepsi masyarakat dan dapat menjadi dasar bagi pemerintah untuk menyusun strategi komunikasi dan kebijakan yang lebih efektif.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Penelitian ini berhasil mengumpulkan 512 tweet yang terkait dengan kebijakan Pajak Pertambahan Nilai (PPN) 12%, yang diperoleh melalui penggunaan Twitter API. 512 tweet yang relevan berhasil dipertahankan untuk dianalisis lebih lanjut. Setiap tweet yang dikumpulkan disertai dengan metadata seperti jumlah *retweet*, jumlah suka, dan komentar. Ini penting untuk memberikan gambaran mengenai seberapa besar dampak dan jangkauan opini terhadap kebijakan tersebut. Salah satu tantangan utama dalam pengumpulan data adalah memfilter konten yang tidak relevan seperti spam, tweet yang dihasilkan oleh bot, atau tweet yang tidak berhubungan langsung dengan kebijakan PPN 12%. Data yang telah difilter ini disimpan dalam format terstruktur, seperti CSV dan database SQL, untuk mempermudah tahap analisis selanjutnya.

Proses pengumpulan dilakukan dengan menggunakan kata kunci relevan seperti "PPN 12%", "pajak", "kenaikan harga", dan "pajak pertambahan nilai". Setelah data tweets terkumpul, data yang diperoleh diubah menjadi data tabel agar mudah untuk diproses pada tahap selanjutnya. Data frame berisi tiga atribut, atribut-atributnya adalah:

1. *Username* : *username* mengandung nama pengguna yang berasal dari akun pembuat tweets yang diambil.
2. *Created_at* : *Created_at* mengandung waktu saat pengguna membuat tweets.
3. *Full_text* : *Full_text* mengandung isi yang berasal dari tweet yang telah dibuat.

Adapun hasil pengumpulan data dapat dilihat pada Tabel 1 berikut:

Tabel 1. Hasil pengumpulan data

Full_teks	Username	Created_at
bruah PPN 12% https://t.co/RPHNtvWu2v	keii68_	Thu Jan 02 23:16:48 +0000 2025
@AnamFreakintalk @drseptian @ardisatriawan pengusaha yang beli	eighxtyone	Thu Jan 02 23:15:44 +0000 2025



barang mewah kena ppn 12% pengusaha juga gak mau rugi 1% mereka gantikan dengan naikin harga barang yang mereka jual ntah 1% mereka lebur dalam 1 bulan/3 bulan/12 bulan apalagi pegusaha tsb punya usaha distributor produk FMCG atau bahkan pabriknya

PPN gajadi naik 12% Meanwhile pas lg beli makan siang di kantor. Admin kalo pake QRIS naik jd 1000. Pas nanya kok naik jawaban pedagang nya adalah: Iya mas soalnya bahan2 udah pada naik admin potongan pas mau cairin jg naik Hehe. <https://t.co/2euEqk0hpV>

deniarbian

Thu Jan 02 23:10:36 +0000 2025

PPN 12% memastikan kebijakan pajak lebih adil dan PRO Rakyat. #KebijakanPajakProRakyat

lexa_david28

Thu Jan 02 23:07:27 +0000 2025

3.2. Prapemrosesan Data

3.2.1. Cleaning Data

Proses ini melibatkan penghapusan elemen-elemen yang tidak relevan seperti URL, nama pengguna, simbol (emoji), dan tagar yang tidak memiliki relevansi langsung terhadap analisis sentimen. Misalnya, tagar seperti "#PPN12" disederhanakan menjadi "PPN12", menghapus tautan dan simbol-simbol yang tidak membawa informasi yang berguna. Adapun Tabel 2 berikut merupakan hasil dari tahap *cleaning data*.

Tabel 2. *Cleaning Data*

Sebelum	Sesudah
bruh PPN 12% https://t.co/RPHNtvWu2v	bruh ppn
@AnamFreakintalk @drseptian @ardisatriawan pengusaha yang beli barang mewah kena ppn 12% pengusaha juga gak mau rugi 1% mereka gantikan dengan naikin harga barang yang mereka jual ntah 1% mereka lebur dalam 1 bulan/3 bulan/12 bulan apalagi pengusaha tsb punya usaha distributor produk FMCG atau bahkan pabriknya	pengusaha yang beli barang mewah kena ppn pengusaha juga gak mau rugi mereka gantikan dengan naikin harga barang yang mereka jual ntah mereka lebur dalam bulan bulan apalagi pengusaha tsb punya usaha distributor produk fmcg atau bahkan pabriknya
@lizasiman @ardisatriawan Ppn untuk barang mewah 1 ampe 31 Jan pakai DPP 11/12 mulai 1 Feb baru DPP seniai harga jualnya. Itu cuma ngebahas barang mewah aja disana.	ppn untuk barang mewah ampe jan pakai dpp mulai feb baru dpp seniai harga jualnya itu cuma ngebahas barang mewah aja disana
Akhirnya perusahaan penyedia jasa internet CBN membatalkan kenaikan PPN 12%. @cbn.id https://t.co/vuCGzzAUHL	akhirnya perusahaan penyedia jasa internet cbn membatalkan kenaikan ppn id
sdm negara ppn 12%	sdm negara ppn

3.2.2. Stopword Removal

Dalam tahapan filtering akan mempergunakan library NLTK dalam bahasa indonesia agar memudahkan tahap penghilangan kalimat. Dalam penelitian ini, peneliti juga menambahkan beberapa kata yang sering muncul pada data tweets seperti "yg", "tdk", "utk" dan lainnya supaya dapat mengurangi terjadinya *noise* dan menjadikan data lebih bersih.

3.2.3. Tokenisasi

Proses ini memecah teks menjadi kata-kata atau token. Sebagai contoh, kalimat "bruh ppn" dipecah menjadi ['bruh', 'ppn'], yang memungkinkan model untuk lebih mudah memahami struktur teks. Adapun Hasil dari tokenisasi dapat dilihat pada tabel 3 berikut.

Tabel 3. Tokenisasi

Sebelum	Sesudah
bruh ppn	['bruh', 'ppn']



pengusaha yang beli barang mewah kena ppn pengusaha juga gak mau rugi mereka gantikan dengan naikin harga barang yang mereka jual ntah mereka lebur dalam bulan bulan bulan apalagi pengusaha tsb punya usaha distributor produk fmcg atau bahkan pabriknya	['pengusaha', 'beli', 'barang', 'mewah', 'kena', 'ppn', 'usaha', 'gak', 'rugi', 'ganti', 'naikin', 'harga', 'barang', 'jual', 'ntah', 'lebur', 'pegusaha', 'tsb', 'usaha', 'distributor', 'produk', 'fmcg', 'pabrik']
ppn untuk barang mewah ampe jan pakai dpp mulai feb baru dpp seniai harga jualnya itu cuma ngebahas barang mewah aja disana	['ppn', 'barang', 'mewah', 'ampe', 'jan', 'pakai', 'dpp', 'feb', 'dpp', 'nia', 'harga', 'jual', 'ngebahas', 'barang', 'mewah', 'aja', 'sana']
Akhirnya perusahaan penyedia jasa internet CBN membatalkan kenaikan PPN 12%. @cbn.id https://t.co/vuCgzzAUHL	['usaha', 'sedia', 'jasa', 'internet', 'cbn', 'batal', 'naik', 'ppn', 'id']
sdm negara ppn	['sdm', 'negara', 'ppn']

3.2.4. Stemming

Langkah ini bertujuan untuk mengurangi mendeteksi duplikasi konten dalam data, seperti retweet yang memiliki kata-kata yang persis sama. Stemming dilakukan untuk mengembalikan kata-kata ke bentuk dasar agar analisis lebih konsisten. Adapun Hasil dari stemming dapat dilihat pada tabel 4 berikut.

Tabel 4. Hasil data stemming

Sebelum	Sesudah
bruh ppn	bruh ppn
pengusaha yang beli barang mewah kena ppn pengusaha juga gak mau rugi mereka gantikan dengan naikin harga barang yang mereka jual ntah mereka lebur dalam bulan bulan bulan apalagi pengusaha tsb punya usaha distributor produk fmcg atau bahkan pabriknya	usaha beli barang mewah kena ppn usaha gak rugi ganti naikin harga barang jual ntah lebur pegusaha tsb usaha distributor produk fmcg pabrik
ppn untuk barang mewah ampe jan pakai dpp mulai feb baru dpp seniai harga jualnya itu cuma ngebahas barang mewah aja disana	ppn barang mewah ampe jan pakai dpp feb dpp nia harga jual ngebahas barang mewah aja sana
Akhirnya perusahaan penyedia jasa internet CBN membatalkan kenaikan PPN 12%. @cbn.id https://t.co/vuCgzzAUHL	usaha sedia jasa internet cbn batal naik ppn,id
sdm negara ppn	sdm negara ppn

3.2.5. Pemilihan fitur dengan TF-IDF

Teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk mengekstrak fitur dari teks. TF-IDF memberikan bobot yang lebih tinggi pada kata-kata yang jarang namun relevan dalam konteks kebijakan PPN 12%, memastikan kata-kata kunci seperti "kenaikan harga" atau "pajak" dipertimbangkan lebih berat dalam analisis.

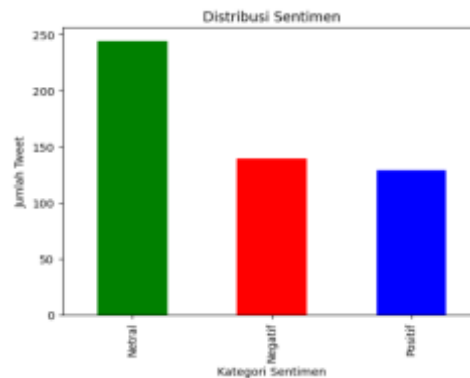
3.3. Klasifikasi Sentimen

Setelah tahap prapemrosesan, analisis sentimen dilakukan dengan tujuan untuk mengklasifikasikan tweet menjadi tiga kategori: positif, negatif, dan netral. Dari total 512 tweet, hasil klasifikasi sentimen menunjukkan bahwa 25.20% tweet mengandung sentimen positif, 27.15% negatif, dan 47.66% netral.

1. Sentimen Positif: 25.20% opini positif terkait kebijakan PPN 12% berfokus pada keyakinan bahwa kebijakan ini akan meningkatkan penerimaan negara untuk mendanai program pembangunan ekonomi. Masyarakat yang mendukung kebijakan ini merasa bahwa PPN 12% akan menciptakan sistem perpajakan yang lebih adil.
2. Sentimen Negatif: Sebagian besar tweet, yaitu 27.15%, menunjukkan sentimen negatif terhadap kebijakan tersebut. Banyak tweet yang mengkhawatirkan dampak negatif terhadap harga barang-barang kebutuhan pokok, yang diharapkan akan naik akibat penerapan PPN. Sentimen ini juga mencakup kekhawatiran akan ketidakpastian ekonomi dan dampak terhadap daya beli masyarakat, terutama golongan menengah ke bawah.
3. Sentimen Netral: 47.66% tweet bersifat netral, yang lebih bersifat informatif mengenai kebijakan itu sendiri,



tanpa memberikan dukungan atau penolakan yang jelas. Beberapa tweet juga mencakup klarifikasi teknis mengenai mekanisme PPN 12%, tanpa membahas dampaknya secara emosional. Grafik Distribusi Sentimen dapat dilihat pada gambar 2 berikut.



Gambar 2. Hasil analisis sentimen

Gambar tersebut menunjukkan distribusi sentimen dari 512 tweet terkait kebijakan PPN 12%. Mayoritas tweet (47.66%) memiliki sentimen netral, yang bersifat informatif tanpa opini jelas. Sentimen negatif (27.15%) menyoroti kekhawatiran terhadap kenaikan harga barang dan dampak pada daya beli masyarakat, sementara sentimen positif (25.20%) mendukung kebijakan ini karena dianggap dapat meningkatkan penerimaan negara dan menciptakan sistem perpajakan yang lebih adil. Visualisasi ini menunjukkan bahwa sebagian besar tweet lebih fokus pada penyampaian informasi daripada menyatakan opini emosional.



Gambar 3. Hasil WordCloud sentimen analisis

Gambar tersebut merupakan visualisasi *WordCloud* dari hasil analisis sentimen terkait kebijakan PPN 12%. *WordCloud* menampilkan kata-kata yang paling sering muncul dalam tweet, dengan ukuran kata yang lebih besar mencerminkan frekuensi kemunculannya yang lebih tinggi. Kata-kata seperti "Rakyat", "PPN", dan "Kebijakan Pajak" mendominasi, menunjukkan bahwa diskusi utama berfokus pada dampak kebijakan terhadap rakyat. Selain itu, terdapat kata-kata lain seperti "barang", "pro", dan "mewah", yang mengindikasikan perhatian masyarakat terhadap aspek tertentu dari kebijakan ini, seperti barang kebutuhan pokok dan perdebatan tentang keadilan penerapannya. *WordCloud* ini memberikan gambaran umum tentang tema dan topik utama yang menjadi perhatian masyarakat dalam diskusi tentang kebijakan PPN 12%.

3.4. Evaluasi Model

Evaluasi model dilakukan dengan menggunakan matrik yang mengukur kinerja model dalam mengklasifikasikan data uji yang telah diproses. Berdasarkan hasil evaluasi, model Naive Bayes menunjukkan performa yang cukup baik dengan nilai evaluasi sebagai berikut:

1. Akurasi: Model ini berhasil mencapai 79% akurasi, yang menunjukkan bahwa sebagian besar tweet telah diklasifikasikan dengan benar ke dalam kategori sentimen yang sesuai.
2. Precision: Precision model adalah 80%, yang menunjukkan bahwa sebagian besar tweet yang dikategorikan sebagai positif benar-benar memiliki sentimen positif.

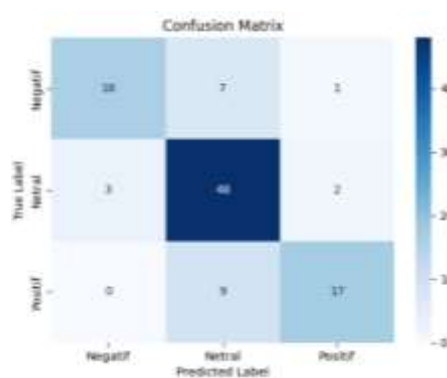
3. Recall: Nilai recall model adalah 79%, yang mengindikasikan bahwa model berhasil mengenali sebagian besar tweet dengan sentimen positif, meskipun ada beberapa kesalahan pengklasifikasian.
4. F1-Score: Dengan nilai F1-Score 78%, model ini berhasil menyeimbangkan antara precision dan recall, memberikan gambaran yang baik tentang kemampuan model dalam menangani klasifikasi sentimen secara menyeluruh.

Representasi secara visual dari Hasil Evaluasi Model dapat dilihat pada Tabel 5 berikut :

Tabel 5. Hasil data evaluasi model

Akurasi	Precision	Recall	F1-Score
79%	80%	79%	78%

Confusion matrix menunjukkan bahwa model kesulitan membedakan antara tweet netral dan tweet negatif, dengan beberapa tweet netral salah dikategorikan sebagai negatif. Hal ini menunjukkan bahwa kategori netral memerlukan perhatian khusus pemrosesan lebih lanjut. Hasil confusion matrix dapat dilihat pada gambar 4 berikut.



Gambar 4. Hasil confusion matrix

Gambar tersebut merupakan Confusion Matrix yang menunjukkan kinerja model dalam mengklasifikasikan tweet ke dalam sentimen negatif, netral, dan positif. Pada sentimen negatif, dari 24 tweet sebenarnya, model berhasil memprediksi 16 dengan benar sebagai negatif, sementara 7 salah diprediksi sebagai netral, dan 1 sebagai positif. Untuk sentimen netral, dari 53 tweet sebenarnya, model memprediksi 48 dengan benar sebagai netral, 3 salah sebagai negatif, dan 2 sebagai positif. Pada sentimen positif, dari 26 tweet sebenarnya, model memprediksi 17 dengan benar sebagai positif, sedangkan 9 salah diprediksi sebagai netral, dan tidak ada yang salah diprediksi sebagai negatif. Matriks ini menunjukkan bahwa model memiliki akurasi tinggi pada sentimen netral, namun terdapat kesalahan signifikan dalam membedakan sentimen negatif dan netral, serta dalam memprediksi sentimen positif yang sering diklasifikasikan sebagai netral. Hasil ini menunjukkan bahwa meskipun model bekerja cukup baik, terdapat ruang untuk meningkatkan akurasi, terutama dalam membedakan kategori sentimen yang lebih ambigu.

4. KESIMPULAN

Berdasarkan hasil analisis sentimen terhadap kebijakan PPN 12% dan evaluasi kinerja model Naive Bayes, dapat disimpulkan bahwa sentimen masyarakat terbagi menjadi tiga kategori utama: positif (25.20%), negatif (27.15%), dan netral (47.66%). Sentimen positif sebagian besar berfokus pada keyakinan bahwa kebijakan ini akan meningkatkan penerimaan negara dan menciptakan sistem perpajakan yang lebih adil, sementara sentimen negatif didominasi oleh kekhawatiran terhadap kenaikan harga kebutuhan pokok, ketidakpastian ekonomi, dan dampak pada daya beli masyarakat. Sentimen netral lebih banyak mencerminkan informasi teknis tanpa opini emosional. Dari segi kinerja, model Naive Bayes menunjukkan performa yang cukup baik dengan akurasi 79%, precision 80%, recall 79%, dan F1-Score 78%. Namun, model masih menghadapi tantangan dalam membedakan tweet netral dan negatif, yang memerlukan perhatian lebih dalam pengolahan data. Ke depannya, pemrosesan data yang lebih mendalam dan

pengembangan fitur analisis yang lebih kompleks, seperti model berbasis deep learning, dapat meningkatkan akurasi klasifikasi. Selain itu, upaya sosialisasi kebijakan yang lebih komprehensif dapat membantu memperbaiki persepsi masyarakat terhadap kebijakan PPN 12%.

REFERENSI

- [1] M. Setiawan, A. S. et al., "Analisis sentimen masyarakat terhadap kebijakan kenaikan tarif listrik di Indonesia menggunakan algoritma Naive Bayes," *J. Teknol. Dan Inform.*, vol. 15, no. 5, hlm. 245–257, 2022.
- [2] R. Rahmawati dan D. Nugroho, "Analisis opini publik terhadap kebijakan subsidi BBM di Indonesia," *J. Sains Dan Teknol.*, vol. 18, no. 4, hlm. 315–325, 2021.
- [3] M. Fauzan, R. P. et al., "Evaluasi sentimen masyarakat terhadap kebijakan pendidikan gratis menggunakan algoritma Naive Bayes," *J. Pendidik. Dan Teknol.*, vol. 22, no. 1, hlm. 134–145, 2023.
- [4] S. Susanto, I. D. et al., "Evaluasi reaksi publik terhadap reformasi perpajakan di Indonesia menggunakan data media sosial," *J. Ekon. Dan Kebijakan. Publik*, vol. 20, no. 2, hlm. 120–133, 2022.
- [5] F. Pratama dan Y. Yulianto, "Sistem analisis sentimen terhadap kebijakan publik di Indonesia dengan algoritma Naive Bayes," *J. Sist. Inf. Dan Komput.*, vol. 19, no. 3, hlm. 450–465, 2021.
- [6] S. Putri, D. P. et al., "Analisis persepsi masyarakat terhadap kebijakan kesehatan dengan pendekatan Naive Bayes," *J. Kesehat. Masy.*, vol. 17, no. 5, hlm. 432–444, 2023.
- [7] D. Hidayat dan A. Wibisono, "Evaluasi opini masyarakat terhadap kebijakan transportasi publik di Indonesia," *J. Transp. Dan Mobilitas*, vol. 16, no. 3, hlm. 200–213, 2022.
- [8] A. Andini, R. H. et al., "Kajian sentimen masyarakat terhadap kebijakan digitalisasi pajak di Indonesia dengan Naive Bayes," *J. Perpajakan. Indonesia.*, vol. 21, no. 2, hlm. 175–189, 2021.
- [9] N. Nurhayati, A. K. et al., "Sentimen masyarakat terhadap kebijakan insentif pajak dengan analisis Naive Bayes," *J. Ekon. Dan Pajak*, vol. 19, no. 4, hlm. 105–118, 2022.
- [10] T. Utami dan R. Kurniawan, "Reaksi publik terhadap kebijakan ekonomi makro Indonesia melalui teks mining," *J. Ekon. Dan Data Anal.*, vol. 25, no. 2, hlm. 250–264, 2023.
- [11] A. Putri, N. Syahrir, dan R. Pratama, "Sentiment analysis using TF-IDF and Naive Bayes for public opinion classification in social media," *J. Data Sci. Appl.*, vol. 12, no. 3, hlm. 45–52, 2022.
- [12] H. Rahmawati dan F. Nugroho, "Evaluating public sentiment on government policies using machine learning approaches," *Int. J. Soc. Sci. Technol.*, vol. 10, no. 1, hlm. 87–93, 2023.
- [13] I. Setiawan, A. Fauzan, dan R. Susanto, "Public sentiment analysis on taxation policies: A case study of VAT increase," *Indones. J. Policy Stud.*, vol. 8, no. 4, hlm. 112–125, 2023.
- [14] D. Utami dan S. Kurniawan, "Demographic-based sentiment analysis on government programs using Naive Bayes," *Comput. Soc. Sci. Rev.*, vol. 15, no. 2, hlm. 60–74, 2022.
- [15] R. Hidayat dan T. Wibisono, "Theme-based analysis in public opinion mining: Insights from transport policy studies," *J. Appl. Data Sci.*, vol. 11, no. 2, hlm. 33–48, 2023.