

# Analisis Sentimen Publik Terhadap Kenaikan Pajak Pertambahan Nilai (PPN) Sebesar 12% Menggunakan *Naïve Bayes Classifier*

Lu'luil Jannah\*, Sriani

Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: <sup>1,\*</sup>luluiljannah153@gmail.com, <sup>2</sup>sriani@uinsu.ac.id

Email Penulis Korespondensi: luluiljannah153@gmail.com\*

Submitted: 17/10/2025; Accepted: 19/12/2025; Published: 31/12/2025

**Abstrak**—Perkembangan teknologi digital membuka peluang yang semakin luas bagi masyarakat untuk menyampaikan pendapat secara bebas melalui berbagai *platform* media sosial. Salah satu isu yang banyak menarik perhatian publik adalah kebijakan pemerintah terkait kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12%. Ragam opini yang muncul dari masyarakat dapat dimanfaatkan sebagai bahan pertimbangan dalam proses evaluasi maupun pengambilan keputusan. Penelitian ini bertujuan untuk mengidentifikasi sentimen publik terhadap kebijakan kenaikan PPN dengan memanfaatkan algoritma *Naïve Bayes Classifier*. Tahapan penelitian meliputi pengumpulan data dari media sosial, pra-pemrosesan teks seperti *case folding*, tokenisasi, *stopword removal*, dan *stemming* serta pengubahan data teks ke bentuk numerik menggunakan metode TF-IDF. Dari total 600 data yang berhasil dihimpun dari media social X, dengan 80% digunakan sebagai data pelatihan dan 20% data sebagai data pengujian. Hasil penelitian menunjukkan adanya 51 *tweet* yang bernada positif, 352 bernada netral, dan 197 bernada negatif. Model *Naïve Bayes* menghasilkan performa klasifikasi yang cukup baik dengan akurasi 81,36%, presisi rata-rata 88%, *recall* 79%, dan *F1-Score* 82%. Temuan ini membuktikan bahwa *Naïve Bayes* merupakan algoritma yang efektif dan layak diandalkan untuk mengklasifikasikan opini publik secara cepat dan sistematis. Dengan demikian, model ini berpotensi menjadi alat pendukung dalam menganalisis persepsi masyarakat terhadap kebijakan pemerintah secara berbasis data.

**Kata Kunci:** Sentimen; Pajak; *Naïve Bayes*; Medsos; TF-IDF.

**Abstract**— The development of digital technology has opened up wider opportunities for the public to freely express their opinions through various social media platforms. One issue that has attracted a lot of public attention is the government's policy to increase the Value Added Tax (VAT) to 12%. The variety of opinions that have emerged from the public can be used as material for consideration in the evaluation and decision-making processes. This study aims to identify public sentiment towards the VAT increase policy by utilizing the *Naïve Bayes Classifier* algorithm. The research stages include data collection from social media, text pre-processing such as case folding, tokenization, *stopword removal*, and *stemming*, as well as converting text data into numerical form using the TF-IDF method. Of the total 600 data successfully collected from social media X, 80% data were used as training data 20% data as test data. The results showed that there were 51 positive tweets, 352 neutral tweets, and 197 negative tweets. The *Naïve Bayes* model produced fairly good classification performance with an accuracy of 81.36%, an average precision of 88%, a recall of 79%, and an *F1-Score* of 82%. These findings prove that *Naïve Bayes* is an effective and reliable algorithm for classifying public opinion quickly and systematically. Thus, this model has the potential to be a supporting tool in analyzing public perception of policies.

**Keywords:** Sentiment; Tax; *Naïve Bayes*; Socmed; TF-IDF.

## 1. PENDAHULUAN

Di era digital, media sosial, forum daring, dan portal berita menjadi sumber utama informasi dan tempat masyarakat menyampaikan opini publik[1], [2]. Salah satu isu yang ramai diperbincangkan adalah kenaikan Pajak Pertambahan Nilai (PPN). Sebagai pajak atas transaksi barang dan jasa, PPN menjadi sumber penerimaan negara, namun kebijakan kenaikannya memunculkan pro dan kontra di ruang publik[3]. Sebagian masyarakat menilai kenaikan PPN penting untuk mendukung pendapatan negara pasca pandemi[4], sementara yang lain merasa terbebani karena berpotensi meningkatkan harga dan menurunkan daya beli, terutama bagi kelompok berpenghasilan menengah ke bawah. Perbedaan pandangan ini tampak jelas melalui unggahan dan diskusi di media sosial[5], [6].

Untuk memahami pola opini publik, digunakan analisis sentimen sebagai pendekatan komputasional yang bertujuan mengidentifikasi kecenderungan sikap masyarakat berdasarkan data teks[7],[8]. Analisis sentimen, atau opinion mining, bekerja dengan mengekstraksi dan mengklasifikasikan teks ke dalam kategori positif, negatif, atau netral. Metode ini merupakan bagian dari text classification yang erat kaitannya dengan natural language processing (NLP) dan text mining [9], [10]. Pendekatan ini mampu menggambarkan pandangan, emosi, serta sikap pengguna terhadap suatu topik, baik yang bersifat faktual maupun nonfaktual [11]. Analisis sentimen telah digunakan dalam berbagai bidang, seperti evaluasi layanan, penilaian kualitas produk, dan analisis persepsi publik terhadap kebijakan pemerintah. Dengan kemampuannya menghasilkan informasi secara sistematis, metode ini menjadi instrumen penting dalam menyediakan wawasan berbasis data untuk mendukung pengambilan keputusan yang lebih tepat dan strategis [12], [13].

Pada penelitian ini digunakan algoritma *Naïve Bayes*, yaitu metode klasifikasi probabilistik yang banyak digunakan dalam pengolahan teks karena bersifat sederhana, cepat, dan memiliki tingkat akurasi yang baik. *Naïve Bayes Classifier* (NBC) bekerja berdasarkan Teorema Bayes dengan asumsi bahwa setiap fitur saling independen dalam memengaruhi kelas[14]. Meskipun berbasis asumsi yang sederhana, algoritma ini terbukti efektif dalam

berbagai tugas klasifikasi, terutama pada analisis sentimen, karena mampu memproses data dengan efisiensi tinggi dan menghasilkan prediksi yang akurat[15]. Selain itu, *Naïve Bayes* mampu membangun model prediktif yang digunakan untuk mengelompokkan data dengan label yang belum diketahui, sehingga menjadikannya salah satu metode yang paling populer dalam pengolahan teks[16].

Penelitian yang dilakukan oleh Agustina, C.A. dan Citra, D.H. (2022) berjudul “Implementasi Algoritma *Naïve Bayes* untuk Analisis Sentimen Ulasan Shopee pada Google Play Store” Penelitian ini menganalisis opini pengguna terhadap aplikasi Shopee berdasarkan 969 ulasan bersih hasil web scraping dari Google Play Store. Menggunakan algoritma *Naïve Bayes*, metode *Hold-Out* menghasilkan akurasi 83%, sedangkan *10-fold cross validation* mencapai 82%. Nilai *precision*, *recall*, dan *f1-score* masing-masing sebesar 83%, 100%, dan 91%, yang menunjukkan sentimen pengguna cenderung positif. Hasil ini membuktikan bahwa *Naïve Bayes* efektif dalam mengklasifikasikan opini berbasis teks dan bermanfaat untuk memahami kebutuhan serta meningkatkan kualitas layanan aplikasi [17]. Penelitian lain yang dilakukan oleh Rohim dan Pratama (2022) berjudul “Analisis Sentimen Publik di Media Sosial Instagram atas Kinerja Presiden Joko Widodo” menggunakan *Naïve Bayes Classifier* (NBC) untuk mengklasifikasikan opini publik menjadi sentimen positif, negatif, dan netral. Hasil penelitian menunjukkan kinerja algoritma yang baik dengan nilai akurasi 83%, *precision* 81%, dan *recall* 84%, yang membuktikan bahwa NBC efektif dalam menganalisis persepsi masyarakat terhadap kinerja politik melalui media sosial [18].

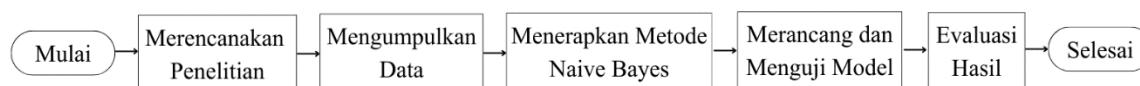
Pada penelitian sebelumnya, algoritma *Naïve Bayes* digunakan untuk menganalisis sentimen terhadap aplikasi dan kinerja politik, sedangkan pada penelitian ini algoritma yang sama diterapkan untuk menganalisis sentimen publik terhadap kebijakan kenaikan Pajak Pertambahan Nilai (PPN) sebesar 12%. Dengan demikian, penelitian ini berfokus pada isu ekonomi dan kebijakan fiskal, yang menjadi perbedaan utama dibandingkan konteks penelitian sebelumnya. Maka rumusan masalah yang diangkat dalam penelitian ini adalah bagaimana menghasilkan model klasifikasi menggunakan *Naïve Bayes Classifier* (NBC) dalam analisis sentimen kenaikan Pajak Pertambahan Nilai senilai 12% dengan menggunakan data dari media social X.

Tujuan penelitian ini adalah menganalisis pandangan publik terhadap kenaikan PPN 12% menggunakan algoritma *Naïve Bayes Classifier* untuk mengidentifikasi sentimen positif, negatif, atau netral serta mengevaluasi akurasi model. Selain menggambarkan persepsi masyarakat terhadap kebijakan fiskal, hasil penelitian diharapkan menjadi masukan bagi pemerintah dalam menilai efektivitas kebijakan dan menyusun strategi komunikasi yang lebih tepat. Dengan demikian, analisis sentimen berfungsi tidak hanya sebagai metode klasifikasi, tetapi juga sebagai alat pendukung pengambilan keputusan berbasis data.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui beberapa tahapan sistematis, dimulai dari perencanaan penelitian, pengumpulan data, dan pengolahan data untuk menjamin kualitas data yang digunakan. Selanjutnya, penelitian ini menerapkan algoritma *Naïve Bayes* sebagai metode utama dalam proses klasifikasi sentimen, yang dipilih karena kesederhanaan dan kemampuannya dalam menangani data teks. Tahap berikutnya mencakup perancangan, pengujian, serta evaluasi model berbasis *Naïve Bayes* untuk menilai tingkat akurasi dan efektivitas hasil klasifikasi. Melalui rangkaian tahapan tersebut, penelitian ini diharapkan menghasilkan analisis sentimen yang valid serta bermanfaat sebagai dasar pengambilan keputusan berbasis data. Tahapan penelitian dapat dilihat pada gambar berikut.



**Gambar 1.** Alur Penelitian

### 2.2 Perencanaan Penelitian

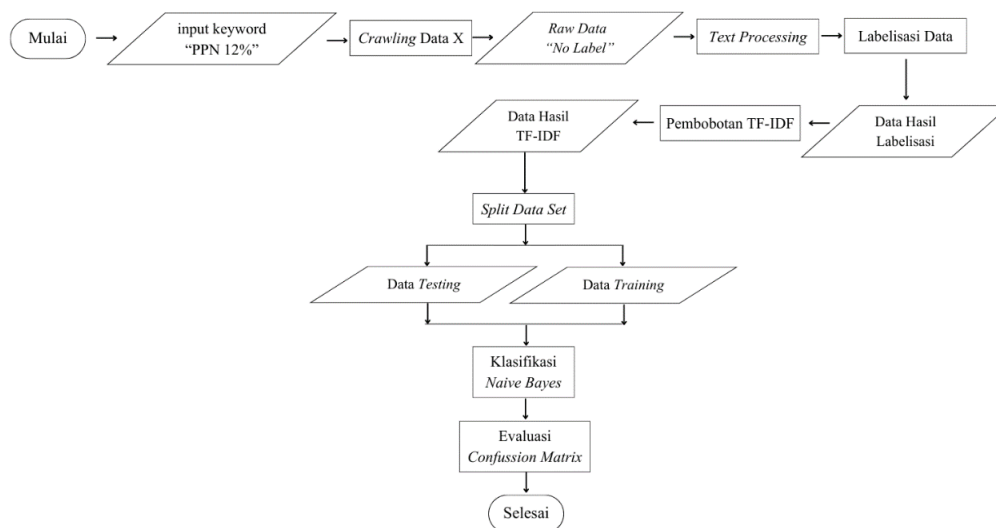
Dalam pelaksanaan penelitian, perencanaan memiliki peran yang sangat penting sebagai dasar dalam menentukan arah dan langkah penelitian. Berdasarkan kerangka penelitian yang telah disusun, dilakukan identifikasi terhadap permasalahan yang berkaitan dengan analisis sentimen publik mengenai kenaikan Pajak Pertambahan Nilai (PPN) pada platform X, yang menimbulkan berbagai macam opini dari pengguna. Pada platform tersebut, terdapat banyak cuitan dan komentar yang membahas isu kenaikan PPN, di mana masyarakat menyampaikan pandangan, tanggapan, maupun asumsi mereka terhadap kebijakan tersebut. Dengan demikian, diperlukan analisis sentimen untuk mengkategorikan pendapat masyarakat menjadi positif, negatif, atau netral dengan memanfaatkan algoritma *Naïve Bayes Classifier*, sehingga arah dan kecenderungan persepsi publik terhadap kebijakan kenaikan PPN dapat diketahui secara lebih objektif dan terukur.

### 2.3 Pengumpulan Data

Penelitian ini memanfaatkan metode *keyword web crawling* untuk mengumpulkan sebanyak 600 data publik dari media sosial X menggunakan kata kunci “PPN 12%”, yang berisi berbagai opini masyarakat mengenai kebijakan kenaikan Pajak Pertambahan Nilai (PPN). Pengambilan data dilakukan secara otomatis dengan bantuan library *tweet-harvest* yang terhubung melalui API X, serta diintegrasikan menggunakan *Python* dan *Node.js*. *Node.js* berperan sebagai jembatan komunikasi antara program *Python* dan API X, sehingga proses permintaan dan penerimaan data dapat berjalan lebih cepat dan efisien. Pendekatan ini memungkinkan pengambilan data secara real-time, terstruktur, dan akurat, sehingga data yang dihasilkan tidak hanya relevan tetapi juga mampu merepresentasikan persepsi publik secara lebih komprehensif terhadap isu kenaikan PPN di media sosial.

## 2.4 Pengolahan dan Penerapan Naïve Bayes Classifier

Proses pengolahan data dimulai dengan tahap pra-pemrosesan yang mencakup beberapa langkah seperti *case folding*, pembersihan data, normalisasi, penghapusan kata-kata tidak penting (*stopword removal*), dan stemming. Setelah itu, data diberi label menggunakan metode berbasis leksikon (*lexicon-based*) dan diubah menjadi bentuk numerik melalui TF-IDF. Selanjutnya, data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Proses klasifikasi dilakukan menggunakan algoritma *Naïve Bayes*, dan hasilnya dievaluasi dengan menggunakan *confusion matrix* untuk mengukur akurasi dan kinerja model. Pendekatan ini memungkinkan analisis sentimen dilakukan dengan efisien serta memberikan wawasan yang dapat mendukung pengambilan keputusan berbasis data.



**Gambar 2.** Flowchart Algoritma Naïve Bayes

Berikut adalah rumus yang digunakan dalam algoritma naive bayes.

$$P(\text{Term} | \text{Kelas}) = \frac{\text{Frekuensi Term di Kelas} + 1}{\text{Total Kata di kelas} + \text{Jumlah Kata unik}} \quad (1)$$

## 2.5 Perancangan dan Pengujian Model

Data mentah (*raw data*) yang diperoleh kemudian melalui tahap *preprocessing* guna membersihkan dan menyiapkan teks dengan beberapa teknik seperti *case folding*, *cleaning*, *normalization*, *stopword removal*, dan *stemming*. Setelah data bersih, dilakukan proses pelabelan sentimen menggunakan metode *lexicon based*, diikuti dengan pembobotan kata menggunakan *TF-IDF*. Setelah data diberi bobot, langkah selanjutnya adalah membagi data menjadi dua bagian, yakni data pelatihan (*training*) dan data pengujian (*testing*). Kemudian, dilakukan proses klasifikasi menggunakan algoritma *Naïve Bayes* untuk memprediksi sentimen pada data pengujian. Pada tahap akhir, kinerja model dievaluasi dengan *confusion matrix* untuk menghitung tingkat akurasi dari metode klasifikasi yang diterapkan, sehingga diperoleh hasil akhir berupa tingkat ketepatan sistem secara keseluruhan.

## 2.6 Evaluasi Akhir

Tahap penutup dari penelitian ini merupakan proses evaluasi menyeluruh terhadap hasil analisis sentimen yang telah dilakukan. Pada tahap ini, hasil klasifikasi yang dihasilkan oleh algoritma *Naïve Bayes Classifier* dinilai menggunakan berbagai metrik kinerja, seperti akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai tersebut kemudian dianalisis untuk menggambarkan arah kecenderungan opini publik terhadap kebijakan kenaikan Pajak Pertambahan Nilai (PPN) sebesar 12%, baik yang menunjukkan sentimen positif, negatif, maupun netral. Dengan demikian, proses evaluasi akhir ini tidak hanya berfungsi untuk menilai sejauh mana efektivitas sistem, tetapi juga

memberikan insight yang dapat dimanfaatkan sebagai dasar pertimbangan dalam pengambilan keputusan dan evaluasi kebijakan publik.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Analisis Data

Dalam penelitian ini, data diperoleh dari platform media sosial X, yang saat ini juga dikenal dengan nama X. Sebanyak 600 *tweet* dikumpulkan sebagai sumber data utama, menggunakan kata kunci yang relevan dengan topik penelitian, yaitu isu kenaikan PPN 12%. *Tweet-tweet* tersebut mencerminkan beragam opini masyarakat, baik yang bersifat positif, negatif, maupun netral. Dengan jumlah data yang cukup signifikan, diharapkan dapat memberikan representasi yang akurat dan bervariasi mengenai persepsi publik terhadap kebijakan perpajakan ini. Data lebih lanjut dapat dilihat pada tabel berikut.

**Tabel 1.** Hasil *Crawling* Data

No	<i>created_at</i>	<i>full_text</i>
1	Wed Jul 02 10:23:50 +0000 2025	Dirjen Pajak Bimo Wijayanto sebut tarif PPN Indonesia 11% masih rendah dibanding rata-rata OECD &
2	Wed Jul 02 07:44:13 +0000 2025	@Dimietryaaa Hai Kak. Apakah maksudnya adalah PPN atas penjualan aset yang merupakan aktiva yang menurut tujuan semula tidak untuk diperjualbelikan oleh Pengusaha Kena Pajak? Jika iya silakan menggunakan DPP 11/12 dari harga jual tarif 12% dan kode 09 ya.
3	Wed Jul 02 07:07:42 +0000 2025	Bukannya udah direncanakan PPN mah? Tinggal ganti 11/12 sama angka lain kan. mau 13/12 mau 15/12 mau 20/12. Kan gak ada aturan yang ngelarang. <a href="https://t.co/vImjvOmdgC">https://t.co/vImjvOmdgC</a>
...	...	...
...	...	...
598	Fri Apr 25 14:03:58 +0000 2025	Parti at my Habitat : Aries = bayar 250 jt + ppn 12% Taurus = free &
599	Fri Apr 25 11:29:27 +0000 2025	@n4mericano_ kiarin efek ppn 12%
600	Fri Apr 25 11:03:19 +0000 2025	@locsigur riswit kenapa PPN 12%

#### 3.2 Representasi Data

Tabel berikut menampilkan 12 sampel komentar yang diambil secara acak dari 600 *tweet* yang dikumpulkan melalui proses *crawling* di X. Sampel ini mewakili beragam opini masyarakat terkait kenaikan PPN 12% dan bertujuan memberikan gambaran awal mengenai karakteristik data, termasuk variasi bahasa, struktur kalimat, dan potensi sentimen sebelum dilakukan pelabelan dan analisis lebih lanjut.

**Tabel 2.** Representasi Data

No	<i>full_text</i>
1	Dirjen Pajak Bimo Wijayanto sebut tarif PPN Indonesia 11% masih rendah dibanding rata-rata OECD &
2	Bukannya udah direncanakan PPN mah? Tinggal ganti 11/12 sama angka lain kan. mau 13/12 mau 15/12 mau 20/12. Kan gak ada aturan yang ngelarang. <a href="https://t.co/vImjvOmdgC">https://t.co/vImjvOmdgC</a>
3	Menurut Menkeu Sri Mulyani target pajak 2025 gak tercapai karena batalnya kenaikan PPN dari 11% ke 12% plus tekanan eksternal seperti harga komoditas yang jatuh. Akibatnya penerimaan pajak turun dan ada potensi kehilangan Rp71 triliun.
4	@tiyananger @txtfromkeuangan gede banget PPN 12%? bukannya 1 1%?
5	@BigAlphaID Sinyal PPN 12% akan segera dilaksanakan...
6	Aneh bgt batalnya ppn 12% jadi defisit. Gak ngaca ?
7	Batalnya PPN 12% itu irelevan. Perbandingannya dgn 3 tahun sebelumnya dimana PPN sama2 11%.
8	Sialan lah Coretax smpe jam segini belum bisa kelapor PPN gua. Cacat! @DitjenPajakRI . . . . #coretax #djp <a href="https://t.co/Txhxr7ydT3">https://t.co/Txhxr7ydT3</a>
9	@remenabrick oiyh ppn 12% ..
10	@onyourm_oki @anugrah_qonita betul semua buku bebas PPN bisa beli buku apa saja di Gramedia termasuk buku impor cek struk nilai bebas PPN-nya berapa semoga membantu
11	@n4mericano_ kiarin efek ppn 12%
12	@locsigur riswit kenapa PPN 12%

### 3.3 Preprocessing Data

Pada penelitian ini, dilakukan proses preprocessing teks dengan tujuan mengubah data mentah menjadi data yang siap digunakan untuk analisis sentimen. Tahapan yang dilakukan meliputi *cleaning* (menghapus karakter khusus, angka, serta tautan), *case folding* (mengonversi seluruh teks menjadi huruf kecil), *tokenizing* (memecah kalimat menjadi potongan kata), normalisasi (menyeragamkan atau menyederhanakan kata tidak baku), *stopword removal* (menghilangkan kata-kata umum yang tidak memiliki makna penting), serta *stemming* (mengembalikan kata ke bentuk dasarnya). Seluruh proses ini bertujuan untuk menghasilkan data teks yang bersih, konsisten, dan terstruktur agar dapat mendukung analisis secara lebih optimal.

#### 1. *Cleaning*

*Cleaning* data merupakan langkah awal dalam proses *preprocessing* yang bertujuan menghilangkan elemen teks yang tidak relevan, seperti simbol, angka, dan tanda baca, emotikon, URL, mention, serta hashtag yang tidak memiliki makna penting. Proses ini berfungsi untuk mengurangi noise pada data sehingga menghasilkan teks yang lebih bersih dan berkualitas. Hasil dari tahap pembersihan data tersebut ditampilkan pada tabel berikut.

**Tabel 3.** Tahap *Cleaning Data*

<i>Cleaning Data</i>
<p>dirjen pajak bimo wijayanto sebut tarif ppn indonesia masih rendah dibanding rata rata oecd            ukannya udah direncanakan ppn mah tinggal ganti sama angka lain kan mau mau mau kan gak ada aturan yang ngelarang            menurut menkeu sri mulyani target pajak gak tercapai karena batalnya kenaikan ppn ke plus tekanan eksternal seperti harga komoditas yang jatuh akibatnya penerimaan pajak turun dan ada potensi kehilangan triliun gede banget ppn bukannya            sinyal ppn akan segera dilaksanakan            aneh bgt batalnya ppn jadi defisit gak ngaca            batalnya ppn itu irrelevant perbandingannya dgn tahun sebelumnya dimana ppn sama            sialan lah coretax smpe jam segini belum bisa kelapor ppn gua cacat            oiya ppn            betul semua buku bebas ppn bisa beli buku apa saja di gramedia termasuk buku impor cek struk nilai bebas ppn nya berapa semoga membantu            kiarin efek ppn            riswit kenapa ppn</p>

#### 2. *Casefolding*

*Casefolding* adalah tahap *preprocessing* teks yang mengubah seluruh huruf menjadi huruf kecil untuk menyamakan bentuk kata. Langkah ini mencegah sistem membedakan kata yang sama karena perbedaan kapitalisasi, seperti “PPN”, “Ppn”, dan “ppn”, sehingga membantu model analisis teks bekerja lebih akurat dan konsisten

**Tabel 4.** Tahap *Casefolding*

<i>Case Folding</i>
<p>dirjen pajak bimo wijayanto sebut tarif ppn indonesia masih rendah dibanding rata rata oecd            ukannya udah direncanakan ppn mah tinggal ganti sama angka lain kan mau mau mau kan gak ada aturan yang ngelarang            menurut menkeu sri mulyani target pajak gak tercapai karena batalnya kenaikan ppn ke plus tekanan eksternal seperti harga komoditas yang jatuh akibatnya penerimaan pajak turun dan ada potensi kehilangan triliun gede banget ppn bukannya            sinyal ppn akan segera dilaksanakan            aneh bgt batalnya ppn jadi defisit gak ngaca            batalnya ppn itu irrelevant perbandingannya dgn tahun sebelumnya dimana ppn sama            sialan lah coretax smpe jam segini belum bisa kelapor ppn gua cacat            oiya ppn            betul semua buku bebas ppn bisa beli buku apa saja di gramedia termasuk buku impor cek struk nilai bebas ppn nya berapa semoga membantu            kiarin efek ppn            riswit kenapa ppn</p>

#### 3. *Tokenizing*



*Tokenisasi* adalah tahap *preprocessing* yang memecah teks atau kalimat menjadi unit-unit lebih kecil, disebut token, biasanya berupa kata. Proses ini memungkinkan sistem menganalisis setiap elemen teks secara terpisah untuk memahami struktur dan makna kalimat dengan lebih baik. Misalnya, kalimat “PPN 12% mulai diberlakukan” diubah menjadi token ['ppn', '12%', 'mulai', 'diberlakukan'] untuk diproses pada analisis sentimen atau klasifikasi teks. Hasil *tokenisasi* dapat dilihat pada tabel berikut.

**Tabel 5.** Tahap *Tokenizing*

<i>Tokenizing</i>
['dirjen', 'pajak', 'bimo', 'wijayanto', 'sebut', 'tarif', 'ppn', 'indonesia', 'masih', 'rendah', 'dibanding', 'rata', 'rata', 'oecd']
'ukannya', 'udah', 'direncanakan', 'ppn', 'mah', 'tinggal', 'ganti', 'sama', 'angka', 'lain', 'kan', 'mau', 'mau', 'mau', 'kan', 'gak', 'ada', 'aturan', 'yang', 'ngelarang']
['menurut', 'menkeu', 'sri', 'mulyani', 'target', 'pajak', 'gak', 'tercapai', 'karena', 'batalnya', 'kenaikan', 'ppn', 'ke', 'plus', 'tekanan', 'eksternal', 'seperti', 'harga', 'komoditas', 'yang', 'jatuh', 'akibatnya', 'penerimaan', 'pajak', 'turun', 'dan', 'ada', 'potensi', 'kehilangan', 'triliun']
['gede', 'banget', 'ppn', 'bukannya']
['sinyal', 'ppn', 'akan', 'segera', 'dilaksanakan']
['aneh', 'bgt', 'batalnya', 'ppn', 'jadi', 'defisit', 'gak', 'ngaca']
['batalnya', 'ppn', 'itu', 'irelevan', 'perbandingannya', 'dgn', 'tahun', 'sebelumnya', 'dimana', 'ppn', 'sama']
['sialan', 'lah', 'coretax', 'smpe', 'jam', 'segini', 'belum', 'bisa', 'kelapor', 'ppn', 'gua', 'cacat']
['oiyh', 'ppn']
['betul', 'semua', 'buku', 'bebas', 'ppn', 'bisa', 'beli', 'buku', 'apa', 'saja', 'di', 'gramedia', 'termasuk', 'buku', 'impor', 'cek', 'struk', 'nilai', 'bebas', 'ppn', 'nya', 'berapa', 'semoga', 'membantu']
['kiarin', 'efek', 'ppn']
['riswit', 'kenapa', 'ppn']

#### 4. Normalisasi

*Normalisasi* adalah tahap *preprocessing* teks yang mengubah kata-kata tidak baku, slang, atau bentuk tidak standar menjadi bentuk sesuai kaidah bahasa Indonesia. Proses ini membantu sistem mengenali kata dengan makna sama secara konsisten, misalnya “nggak”, “gak”, atau “ga” menjadi “tidak”, dan “udah” menjadi “sudah”. Dengan normalisasi, teks menjadi lebih seragam, memudahkan tahap selanjutnya seperti stemming, vektorisasi, dan klasifikasi. Hasil normalisasi dapat dilihat pada tabel berikut.

**Tabel 6.** Tahap Normalisasi

<i>Normalisasi</i>
['direktur', 'pajak', 'bimo', 'wijayanto', 'mengatakan', 'tarif', 'ppn', 'indonesia', 'masih', 'rendah', 'dibandingkan', 'rata-rata', 'oecd']
['bukankah', 'sudah', 'direncanakan', 'ppn', 'tinggal', 'ganti', 'dengan', 'angka', 'lain', 'ingin', 'ingin', 'ingin', 'tidak', 'ada', 'aturan', 'melarang']
['menurut', 'menteri', 'sri', 'mulyani', 'target', 'pajak', 'tidak', 'tercapai', 'karena', 'batal', 'kenaikan', 'ppn', 'ke', 'dan', 'tekanan', 'eksternal', 'seperti', 'harga', 'komoditas', ' ', 'jatuh', ' ', 'penerimaan', 'pajak', 'turun', ' ', 'ada', 'potensi', 'kehilangan', 'triliun']
['besar', 'sangat', 'ppn', 'bukan']
['sinyal', 'ppn', 'akan', 'segera', 'dilaksanakan']
['aneh', 'banget', 'batal', 'ppn', 'menjadi', 'defisit', 'tidak', 'bercermin']
['batal', 'ppn', 'relevan', 'perbandingan', 'dengan', 'tahun', 'sebelum', 'ppn', 'sama']
['sialan', 'coretax', 'sampai', 'jam', 'sekarang', 'belum', 'bisa', 'melapor', 'ppn', 'saya', 'cacat']
['iya', 'ppn']
['benar', 'semua', 'buku', 'bebas', 'ppn', 'dapat', 'membeli', 'buku', 'gramedia', 'termasuk', 'buku', 'impor', 'periksa', 'struk', 'nilai', 'bebas', 'ppn', 'semoga', 'membantu']
['kirain', 'efek', 'ppn']
['riswit', 'kenapa', 'ppn']

#### 5. Stopword

*Stopword* merupakan kata-kata umum yang sering muncul dalam teks namun tidak memiliki pengaruh besar terhadap makna kalimat, seperti “yang”, “dan”, “di”, “adalah”, atau “itu”. Pada tahap *preprocessing* teks, kata-kata tersebut dihapus melalui proses *stopword removal* agar sistem dapat lebih berfokus pada kata-kata yang memiliki makna penting. Dengan demikian, analisis seperti klasifikasi sentimen atau text clustering dapat

dilakukan dengan tingkat akurasi dan efisiensi yang lebih baik. Hasil dari proses penghapusan *stopword* ditampilkan pada tabel berikut.

**Tabel 7.** Tahap *Stopword*

<i>Stopword Removal</i>
['direktur', 'pajak', 'bimo', 'wijayanto', 'mengatakan', 'tarif', 'ppn', 'indonesia', 'rendah', 'rata-rata']
['direncanakan', 'ppn', 'tinggal', 'ganti', 'angka', 'aturan', 'melarang']
['menteri', 'sri', 'mulyani', 'target', 'pajak', 'tercapai', 'batal', 'kenaikan', 'ppn', 'tekanan', 'eksternal', 'harga', 'komoditas', 'jatuh', 'penerimaan', 'pajak', 'turun', 'potensi', 'kehilangan', 'triliun']
['besar', 'ppn']
['sinyal', 'ppn', 'dilaksanakan']
['aneh', 'banget', 'batal', 'ppn', 'defisit', 'bercermin']
['batal', 'ppn', 'relevan', 'perbandingan', 'tahun', 'sebelum', 'ppn']
['sialan', 'coretax', 'jam', 'melapor', 'ppn', 'cacat']
['ppn']
['benar', 'buku', 'bebas', 'ppn', 'membeli', 'buku', 'gramedia', 'termasuk', 'buku', 'impor', 'periksa', 'struk', 'nilai', 'bebas', 'ppn']
['kirain', 'efek', 'ppn']
['riswit', 'kenapa', 'ppn']

## 6. Stemming

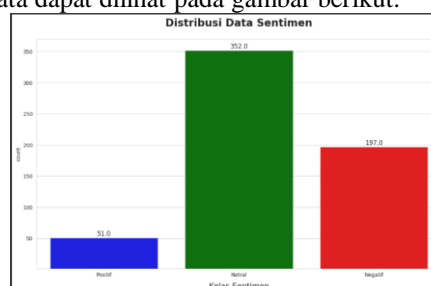
*Stemming* adalah proses mengubah kata turunan menjadi bentuk dasar untuk menyatukan makna yang sama, misalnya “menurunkan” atau “penurunan” menjadi “turun”. Tahap ini meningkatkan konsistensi data dan efisiensi analisis teks, serta umum digunakan dalam NLP seperti klasifikasi teks, analisis sentimen, dan pencarian informasi. Hasil *stemming* dapat dilihat pada tabel berikut.

**Tabel 8.** Tahap *Stemming*

<i>Stemming</i>
['direktur', 'pajak', 'bimo', 'wijayanto', 'kata', 'tarif', 'ppn', 'indonesia', 'rendah', 'rata']
['rencana', 'ppn', 'tinggal', 'ganti', 'angka', 'atur', 'larang']
['menteri', 'sri', 'mulyani', 'target', 'pajak', 'capai', 'batal', 'naik', 'ppn', 'tekan', 'eksternal', 'harga', 'komoditas', 'jatuh', 'terima', 'pajak', 'turun', 'potensi', 'hilang', 'triliun']
['besar', 'ppn']
['sinyal', 'ppn', 'laksana']
['aneh', 'banget', 'batal', 'ppn', 'defisit', 'cermin']
['batal', 'ppn', 'relevan', 'banding', 'tahun', 'sebelum', 'ppn']
['sial', 'coretax', 'jam', 'lapor', 'ppn', 'cacat']
['ppn']
['benar', 'buku', 'bebas', 'ppn', 'beli', 'buku', 'gramedia', 'masuk', 'buku', 'impor', 'periksa', 'struk', 'nilai', 'bebas', 'ppn']
['kira', 'efek', 'ppn']
['riswit', 'kenapa', 'ppn']

## 3.4 Pelabelan Data

Proses pelabelan dilakukan secara semi-otomatis untuk mengklasifikasikan *tweet* berdasarkan sentimen. Dari 600 *tweet*, terdapat 51 positif, 352 netral, dan 197 negatif, menunjukkan bahwa sebagian besar tanggapan masyarakat terhadap kenaikan PPN 12% bersifat netral, dengan sebagian lain mengekspresikan dukungan atau ketidaksetujuan. Hasil Pelabelan data dapat dilihat pada gambar berikut.



**Gambar 3.** Hasil Pelabelan

Hasil analisis menunjukkan bahwa sentimen netral menjadi yang paling dominan karena banyak tweet hanya berisi informasi mengenai kenaikan PPN 12% tanpa menyertakan opini. Mayoritas pengguna sekadar membagikan berita atau menyampaikan fakta, sehingga model mengklasifikasikannya sebagai netral. Pola ini menggambarkan bahwa respons publik masih moderat dan belum menunjukkan kecenderungan kuat terhadap sentimen positif maupun negatif.

### 3.5 TF – IDF

Setelah proses preprocessing selesai dilakukan, tahap selanjutnya adalah perhitungan *TF-IDF* untuk menentukan tingkat kepentingan setiap kata dalam sebuah dokumen terhadap keseluruhan korpus. Nilai *TF* menunjukkan seberapa sering suatu kata muncul dalam dokumen, sedangkan *IDF* menggambarkan seberapa jarang kata tersebut muncul di seluruh kumpulan data. Gabungan kedua nilai ini memberikan bobot yang lebih tinggi pada kata-kata yang informatif dan menurunkan bobot pada kata-kata yang bersifat umum. Vektor *TF-IDF* yang dihasilkan kemudian digunakan sebagai input pada model klasifikasi sentimen, sehingga pola dan makna teks dapat teridentifikasi dengan lebih akurat. Hasil perhitungan nilai *TF-IDF* ditampilkan pada tabel berikut.

**Tabel 9.** Hasil *TF-IDF*

TERM	TF-IDF									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
direktur	0,174	0	0	0	0	0	0	0	0	0
pajak	0,1439	0	0,1439	0	0	0	0	0	0	0
bimo	0,174	0	0	0	0	0	0	0	0	0
wijayanto	0,174	0	0	0	0	0	0	0	0	0
kata	0,174	0	0	0	0	0	0	0	0	0
tarif	0,174	0	0	0	0	0	0	0	0	0
ppn	0,0927	0,1325	0,0464	0,4637	0,3091	0,1546	0,265	0,1546	0,9274	0,1237
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
impor	0	0	0	0	0	0	0	0	0	0,116
periksa	0	0	0	0	0	0	0	0	0	0,116
struk	0	0	0	0	0	0	0	0	0	0,116
nilai	0	0	0	0	0	0	0	0	0	0,116

### 3.6 Klasifikasi Algoritma Naïve Bayes

Sebelum klasifikasi dengan *Naïve Bayes*, data dibagi menjadi 80% untuk latih dan 20% untuk uji, sehingga model dapat belajar dari sebagian besar data dan diuji pada data baru untuk menghindari *overfitting*. Dalam penelitian ini, dari 12 sampel, 10 digunakan sebagai data latih dan 2 sebagai data uji, memberikan gambaran awal tentang performa model dalam memprediksi sentimen. Hasil pembagian data dapat dilihat pada tabel berikut.

**Tabel 10.** *Split Data*

Data	Dokumen	Full Text	Sentimen
Data Latih	D1	['direktur', 'pajak', 'bimo', 'wijayanto', 'kata', 'tarif', 'ppn', 'indonesia', 'rendah', 'rata']	Positif
	D2	['rencana', 'ppn', 'tinggal', 'ganti', 'angka', 'atur', 'larang']	Netral
	D3	['menteri', 'sri', 'mulyani', 'target', 'pajak', 'capai', 'batal', 'naik', 'ppn', 'tekan', 'eksternal', 'harga', 'komoditas', 'jatuh', 'terima', 'pajak', 'turun', 'potensi', 'hilang', 'triliun']	Negatif
	D4	['besar', 'ppn']	Negatif
	D5	['sinyal', 'ppn', 'laksana']	Positif
	D6	['aneh', 'banget', 'batal', 'ppn', 'defisit', 'cermin']	Negatif
	D7	['batal', 'ppn', 'relevan', 'banding', 'tahun', 'sebelum', 'ppn']	Negatif
	D8	['sial', 'coretax', 'jam', 'lapor', 'ppn', 'cacat']	Negatif
	D9	['ppn']	Netral



Data	Dokumen	Full Text	Sentimen
	D10	['benar', 'buku', 'bebas', 'ppn', 'beli', 'buku', 'gramedia', 'masuk', 'buku', 'impor', 'periksa', 'struk', 'nilai', 'bebas', 'ppn']	Positif
Data	D11	['kira', 'efek', 'ppn']	?
Uji	D12	['riswit', 'kenapa', 'ppn']	?

Langkah berikutnya adalah menghitung probabilitas prior dan probabilitas kondisional. Probabilitas prior mencerminkan peluang awal suatu kelas berdasarkan jumlah data dalam masing-masing kategori, sementara probabilitas kondisional menilai kemungkinan kemunculan kata tertentu di setiap kelas. Kedua nilai ini menjadi dasar bagi algoritma Naive Bayes dalam melakukan klasifikasi sentimen.

#### 1. Probabilitas Prior

Probabilitas prior dihitung dari jumlah dokumen dalam setiap kelas dan menunjukkan peluang awal sebuah dokumen termasuk ke dalam kelas tertentu sebelum mempertimbangkan kata atau fitur di dalamnya, mencerminkan proporsi masing-masing kelas dalam keseluruhan data.

Total dokumen  $N = 10$

Jumlah dokumen per kelas:

a. Negatif = 5 (D3,D4,D6,D7,D8)

b. Netral = 2 (D2, D9)

c. Positif = 3 (D1, D5,D10)

#### Probabilitas prior:

$$P(\text{Negatif}) = \frac{5}{10} = 0,5$$

$$P(\text{Netral}) = \frac{2}{10} = 0,2$$

$$P(\text{Positif}) = \frac{3}{10} = 0,3$$

#### 2. Probabilitas Kondisional

Probabilitas kondisional adalah peluang munculnya suatu kata tertentu dengan syarat kata tersebut berada dalam kelas tertentu, digunakan untuk menilai seberapa besar kemungkinan kata muncul dalam dokumen dari kelas tersebut.

Jumlah kata unik = 77

Total kata di kelas :

Negatif = 41(empat puluh satu) kata

Netral = 8 (delapan) kata

Positif = 28 (dua puluh delapan) kata

Setelah menghitung jumlah kata pada tiap kelas (positif, netral, negatif), data uji kemudian diklasifikasikan menggunakan probabilitas prior dan kondisional untuk menentukan sentimennya.

a. ['kira', 'efek', 'ppn']

$$P(\text{kira} | \text{negatif}) = \frac{0 + 1}{41 + 77} = \frac{1}{118} = 0,008474576$$

$$P(\text{kira} | \text{netral}) = \frac{0 + 1}{8 + 77} = \frac{1}{85} = 0,011764706$$

$$P(\text{kira} | \text{positif}) = \frac{0 + 1}{28 + 77} = \frac{1}{105} = 0,00952381$$

$$P(\text{efek} | \text{negatif}) = \frac{0 + 1}{41 + 77} = \frac{1}{118} = 0,008474576$$

$$P(\text{efek} | \text{netral}) = \frac{0 + 1}{8 + 77} = \frac{1}{85} = 0,011764706$$

$$P(\text{efek} | \text{positif}) = \frac{0 + 1}{28 + 77} = \frac{1}{105} = 0,00952381$$

$$P(\text{ppn} | \text{negatif}) = \frac{6 + 1}{41 + 77} = \frac{7}{118} = 0,059322034$$

$$P(\text{ppn} | \text{netral}) = \frac{2 + 1}{8 + 77} = \frac{3}{85} = 0,035294118$$

$$P(\text{ppn} | \text{positif}) = \frac{4 + 1}{28 + 77} = \frac{5}{105} = 0,047619048$$

b. ['riswit', 'kenapa', 'ppn']

$$P(\text{riswit} | \text{negatif}) = \frac{0 + 1}{41 + 77} = \frac{1}{118} = 0,008474576$$

$$P(\text{riswit} | \text{netral}) = \frac{0 + 1}{8 + 77} = \frac{1}{85} = 0,011764706$$

$$P(\text{riswit} | \text{positif}) = \frac{0 + 1}{28 + 77} = \frac{1}{105} = 0,00952381$$

$$P(\text{kenapa} | \text{negatif}) = \frac{0 + 1}{41 + 77} = \frac{1}{118} = 0,008474576$$

$$P(\text{kenapa} | \text{netral}) = \frac{0 + 1}{8 + 77} = \frac{1}{85} = 0,011764706$$

$$P(\text{kenapa} | \text{positif}) = \frac{0 + 1}{28 + 77} = \frac{1}{105} = 0,00952381$$

$$P(\text{ppn} | \text{negatif}) = \frac{6 + 1}{41 + 77} = \frac{7}{118} = 0,059322034$$

$$P(\text{ppn} | \text{netral}) = \frac{2 + 1}{8 + 77} = \frac{3}{85} = 0,035294118$$

$$P(\text{ppn} | \text{positif}) = \frac{4 + 1}{28 + 77} = \frac{5}{105} = 0,047619048$$

3. Prediksi

a. Data Uji 1

$$P(\text{negatif}) = 0,5 * 0,008474576 * 0,008474576 * 0,059322034 = 0,0000021302$$

$$P(\text{netral}) = 0,2 * 0,008474576 * 0,008474576 * 0,059322034 = 0,0000009770$$

$$P(\text{positif}) = 0,3 * 0,00952381 * 0,00952381 * 0,047619048 = 0,0000012958$$

Berdasarkan hasil perhitungan probabilitas dari ketiga kelas (negatif, netral, dan positif), diketahui bahwa kelas negatif memiliki nilai probabilitas tertinggi. Oleh karena itu, data uji 1 diklasifikasikan ke dalam kelas negatif.

b. Data Uji 2

$$P(\text{negatif}) = 0,5 * 0,008474576 * 0,008474576 * 0,059322034 = 0,0000021302$$

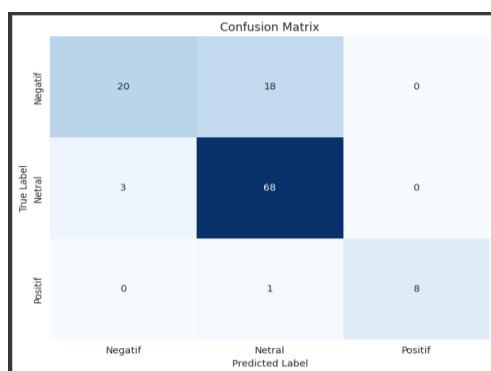
$$P(\text{netral}) = 0,2 * 0,008474576 * 0,008474576 * 0,059322034 = 0,0000009770$$

$$P(\text{positif}) = 0,3 * 0,00952381 * 0,00952381 * 0,047619048 = 0,0000012958$$

Berdasarkan hasil perhitungan probabilitas dari ketiga kelas (negatif, netral, dan positif), diketahui bahwa kelas negatif memiliki nilai probabilitas tertinggi. Oleh karena itu, data uji 2 diklasifikasikan ke dalam kelas negatif.

### 3.7 Confusion Matrix

*Confusion matrix* digunakan untuk menilai kinerja algoritma *Naïve Bayes* dengan menampilkan jumlah prediksi yang benar maupun yang salah untuk setiap kelas, seperti positif, netral, dan negatif. Selain itu, matriks ini berfungsi sebagai dasar untuk menghitung metrik evaluasi utama, termasuk akurasi, presisi, *recall*, dan *F1-score*, sehingga memberikan gambaran menyeluruh tentang kemampuan model dalam mengklasifikasikan data. Hasil dari *confusion matrix* tidak hanya menunjukkan seberapa tepat model dalam memprediksi kelas, tetapi juga membantu mengidentifikasi area di mana model mungkin mengalami kesalahan atau bias, sehingga dapat digunakan sebagai acuan untuk perbaikan model di tahap selanjutnya. dapat dilihat pada gambar berikut.



**Gambar 4.** *Confusion Matrix*

### 3.8 Akurasi

Hasil akurasi algoritma *Naive Bayes* mencapai 81,36%, menunjukkan model mampu mengklasifikasikan sekitar 81 dari 100 data uji dengan benar. Meski akurasi tinggi, metrik tambahan seperti *precision*, *recall*, dan *F1-score* tetap diperlukan untuk menilai kinerja model secara merata di seluruh kelas sentimen, terutama jika terdapat ketidakseimbangan data. Akurasi *Naive Bayes* pada penelitian ini sebesar 81,36%, sedikit lebih rendah dibandingkan penelitian sebelumnya yang berada di angka 82–83%. Meski begitu, hasil ini masih sebanding dan menunjukkan bahwa *Naive Bayes* tetap efektif digunakan pada data sentimen terkait kenaikan PPN. Dapat dilihat pada gambar berikut.

	precision	recall	f1-score	support
Negatif	0.87	0.53	0.66	38
Netral	0.78	0.96	0.86	71
Positif	1.00	0.89	0.94	9
accuracy			0.81	118
macro avg	0.88	0.79	0.82	118
weighted avg	0.83	0.81	0.80	118

**Gambar 5.** Akurasi

Akurasi *Naive Bayes* sebesar 81,36% muncul karena algoritma ini bekerja sangat baik pada data teks dengan pola kata yang berulang dan distribusi fitur yang cukup jelas, sehingga model dapat mengenali kecenderungan sentimen secara konsisten. Hasil ini menunjukkan bahwa model cukup dapat diandalkan untuk membaca respons publik terhadap kebijakan PPN 12%. Bagi pembuat kebijakan, temuan ini berarti analisis sentimen yang dihasilkan dapat dijadikan dasar awal untuk memahami persepsi Masyarakat apakah dominan mendukung, menolak, atau netral sehingga dapat membantu pemerintah menyesuaikan strategi komunikasi, meredam resistensi, atau mengevaluasi efektivitas penerapan PPN 12%.

## 4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma *Naive Bayes Classifier* efektif untuk menganalisis sentimen masyarakat terhadap kenaikan PPN 12%. Proses analisis dilakukan melalui tahapan preprocessing, mulai dari pembersihan data, case folding, tokenisasi, normalisasi kata, penghilangan stopword, hingga stemming sebelum teks diubah menjadi representasi numerik melalui TF-IDF. Dari 600 tweet, diperoleh 51 sentimen positif, 352 netral, dan 197 negatif, menunjukkan dominasi opini netral dan lebih banyaknya respons negatif dibanding positif. Hal ini mengindikasikan bahwa sebagian besar pengguna X bersikap informatif tanpa keberpihakan yang jelas, sementara sebagian lainnya menunjukkan ketidaksetujuan terhadap kebijakan tersebut. Model dilatih dengan pembagian data 80% training dan 20% testing untuk memaksimalkan pembelajaran sekaligus menjaga generalisasi. Evaluasi menunjukkan akurasi 81,36%, presisi rata-rata 88%, recall 79%, dan F1-score 82%, sehingga *Naive Bayes* dinilai cukup andal dalam mengklasifikasikan opini publik. Temuan ini menegaskan bahwa analisis sentimen berbasis *Naive Bayes* dapat menjadi alat penting bagi pemerintah dan pemangku kebijakan untuk memahami persepsi publik dan merumuskan strategi komunikasi yang lebih tepat. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, data hanya berasal dari satu platform yaitu X, sehingga belum mewakili opini dari ekosistem media sosial yang lebih luas. Kedua, dominasi sentimen netral berpotensi mengurangi kekayaan variasi emosi dalam data. Ketiga, proses pelabelan semi-otomatis berbasis lexicon dapat menimbulkan bias, terutama pada kalimat yang ambigu atau bersifat sarkastik. Oleh karena itu, penelitian selanjutnya disarankan memperluas sumber data serta membandingkan performa *Naive Bayes* dengan algoritma lain seperti SVM, Random Forest, LSTM, maupun BERT untuk memperoleh hasil klasifikasi yang lebih optimal dan komprehensif.

## REFERENCES

- [1] N. K. Firda Diana Kartika Rani, Wasesa Rengga Mukti, Atipa Muji, "Peran Media Online Dalam Sebuah Komunikasi," *Merdeka Indones. J. Int.*, vol. 4, no. 1, pp. 67–74, 2024, [Online]. Available: <https://merdekaindonesia.com/index.php/MerdekaIndonesiaJournalInternati/article/view/121/62>
- [2] A. A. R. Rahma, H. Ardianti, and K. Firman, "PERAN MEDIA SOSIAL DALAM DINAMIKA SOSIAL MASYARAKAT KONTEMPORER Ayu," *J. Komun. Digit. dan Penyiaran Islam*, vol. 4, no. 1, pp. 9–15, 2024.
- [3] A. R. Aisya, T. H. N. Purwati, and S. A. Kusrin, "Kajian Hukum Atas Subjek Dan Objek Pajak Penghasilan DalamPrespektif Keadilan Dan Kepatuhan Wajib Pajak," *J. Huk.*, vol. 6, no. 2, pp. 148–160, 2025.
- [4] W. Agun, L. Datrini, and A. Amlayasa, "Kepatuhan Wajib Pajak Dalam Memenuhi Kewajiban," *Wicaksana, JurnalLingkungan Pembang.*, vol. 6, no. 1, p. 9, 2022.
- [5] L. L. Latifah and S. Rosyadi, "Analisis Kebijakan Fiskal Pasca Pandemi Covid-19 Studi Kasus Indonesia," *El-Mal J. Kaji. Ekon. Bisnis Islam*, vol. 5, no. 3, pp. 1787–1801, 2024, doi: 10.47467/elmal.v5i3.799.
- [6] P. A. Ramadhani *et al.*, "Jurnal Administrasi Karya Dharma Volume 4 Nomor 2 ( 2025 ) September 2025 Analisis

- Dampak Kebijakan Kenaikan Tarif Pajak Pertambahan Nilai ( PPN ) Menggunakan Pendekatan Inkremental PENDAHULUAN Negara yang menempati peringkat keempat di dunia dengan popul,” *J. Adm. Karya Dharma*, vol. 4, no. September, pp. 12–26, 2025.
- [7] T. Y. Pahtoni and H. Jati, “Analisis Sentimen Data X Terkait Chatgpt Menggunakan Orange Data Mining,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 2, pp. 329–336, 2024, doi: 10.25126/jtiik.20241127276.
- [8] R. Mursyid and A. D. Indriyanti, “Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: X),” *J. Informatics Comput. Sci.*, vol. 6, no. 02, pp. 371–383, 2024, doi: 10.26740/jinacs.v6n02.p371-383.
- [9] W. Alfiyani, D. Abdul Fatah, and F. Irhamni, “Penerapan Algoritma *Naïve Bayes* Untuk Analisis Sentimen Pada Media Sosial X Terhadap Performa Tim Nasional Sepak Bola Indonesia Di Era Kepemimpinan Shin Tae-Yong,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 3, pp. 3969–3977, 2025, doi: 10.36040/jati.v9i3.13451.
- [10] A. W. V. Hutabarat, N. L. S. S. Adnyani, and K. Suryadi, “Analisis Sentimen Data Ulasan Pengguna MyPertamina di X dengan Metode Text Mining,” *J. Rekayasa Sist. Ind.*, vol. 13, no. 1, pp. 145–154, 2024, doi: 10.26593/jrsi.v13i1.6958.145-154.
- [11] N. Raisa and N. Riza, “Sentimen Analisis Terhadap Opini Masyarakat Mengenai Drama Korea Pada X Menggunakan Metode *Naïve Bayes*,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1312–1320, 2023, doi: 10.36040/jati.v7i2.6765.
- [12] B. Irawan, A. Bahtiar, N. Bayes, G. P. Store, and N. Bayes, “Penerapan Metode Naive Bayes Pada Analisis Sentimen Aplikasi,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 759–766, 2024.
- [13] Sriani, Suhardi, and La Saufa Yardha, “Analisis Sentimen Pengguna Aplikasi Mobile JknMenggunakan Algoritma Naive BayesClassifier Dan C4.5,” *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 555–563, 2024, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [14] S. Safaatulloh, R. Hariyanto, and M. Udin, “Penerapan Metode *Naïve Bayes* Classifier Untuk Kerusakan Perangkat Edc Yang Ada Di Spbu,” *J. Mnemon.*, vol. 7, no. 2, pp. 207–212, 2024, doi: 10.36040/mnemonic.v7i2.10687.
- [15] F. Matheos Sarimole and Kudrat, “Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada X Menggunakan Algoritma Naive Bayes Dan Support Vector Machine,” *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 1–8, 2024, [Online]. Available: <https://doi.org/10.55338/saintek.v5i1.2702>
- [16] N. Riyanah and Fatmawati, “Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu (Implementation of Algorithms *Naïve Bayes* for Classification Recipients Help Letter Description Not Able),” *J. Teknol. Inf. dan Multimed.*, vol. 2, no. 4, pp. 206–213, 2021.
- [17] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, “Implementasi Algoritma Naive 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, doi: 10.57152/malcom.v2i1.195.
- [18] A. N. Rohim and A. R. Pratama, “Analisis Sentimen Publik di Media Sosial Instagram atas Kinerja Presiden Joko Widodo,” 2022.