

## Metode *Naïve Bayes* untuk Analisis Sentimen Kenaikan PPN 12% di Media Sosial Tiktok

Lela Budiarti <sup>a</sup>, Alwendi <sup>b</sup>, Muhamad Rafi Akbar <sup>c</sup>

<sup>a</sup> Universitas Muhammadiyah Tapanuli Selatan

<sup>b</sup> Universitas Graha Nusantara Padang Sidimpuan

<sup>c</sup> STMIK Indragiri, Pekanbaru

### INFORMASI ARTIKEL

#### Sejarah Artikel:

Diterima Redaksi: 20 Oktober 2025

Revisi Akhir: 1 Januari 2026

Diterbitkan Online: 6 Januari 2026

### KATA KUNCI

Analisis Sentimen, *Naïve Bayes Classifier*,  
Pajak, Pajak Pertambahan Nilai (PPN),  
Tiktok

### KORESPONDENSI

Lela Budiarti  
Teknologi Informasi  
lelabudiarti1@gmail.com

### ABSTRAK

Pajak merupakan salah satu sumber utama pendapatan negara, termasuk di Indonesia. Pajak telah menjadi sumber penerimaan terbesar bagi negara sejak dahulu dan berperan penting dalam mendukung pembangunan nasional. Salah satu sumber pendapatan terbesar berasal dari pajak, termasuk Pajak Pertambahan Nilai (PPN). Kenaikan PPN dapat menimbulkan beberapa masalah, termasuk potensi berbelanja di luar negeri seiring dengan kenaikan harga barang dan jasa di Indonesia. Dengan kenaikan PPN ini, maka harga barang akan mengalami kenaikan dan penjualan akan mengalami penurunan sehingga berdampak pada turunnya sektor barang dan jasa. Rencana kenaikan PPN ini kemudian viral di perbincangan di media sosial salah satunya yaitu Tiktok. Banyak masyarakat memperdebatkan rencana kenaikan sehingga terjadinya pro dan kontra terhadap rencana kenaikan PPN ini. Oleh sebab itu penelitian ini bertujuan untuk melakukan klasifikasi komentar dari Tiktok mengenai PPN 12% dengan menggunakan metode *naive bayes classifier*. Beberapa tahap yang dilakukan yaitu *crawling* data, pelabelan data, *preprocessing*, ekstraksi fitur, klasifikasi dengan *naive bayes classifier* dan pengujian dengan *confusion matrix*. Dataset yang didapat merupakan hasil *crawling* data dari komentar yang ada di media sosial Tiktok mulai tanggal 1 Desember sampai dengan 15 Januari 2025. Dari 900 data komentar, hasil pengujian mendapatkan akurasi 96%, presisi 98%, recall 96%, dan fi-score 97%.

**DOI:** <https://10.46961/jommit.v9i2.1815>

### 1. PENDAHULUAN

*Knowledge Discovery in Database* (KDD) atau yang dikenal dengan *Data Mining* adalah proses pengumpulan data yang bertujuan untuk menemukan pola, pengetahuan, dan juga informasi. Namun, pasti ada algoritma atau teknik yang digunakan untuk menemukan pola tersebut. Output yang dapat dihasilkan dari proses *Data Mining* digunakan sebagai pilihan untuk pengambilan keputusan [1]. Tujuan dari *Knowledge Discovery in*

*Database* (KDD) dan *Data Mining* adalah untuk menerapkan metode saintifik pada *data mining*. *Data Mining* (DM) adalah inti dari proses KDD, melibatkan kesimpulan dari algoritma yang mengeksplorasi data, mengembangkan model dan menemukan pola yang sebelumnya tidak diketahui. Tujuan KDD dan *data mining* adalah untuk menggali informasi tersembunyi dari sebuah basis data yang sangat besar [2].

*Text Mining* adalah proses inovasi akan informasi atau isu terkini yang sebelumnya tidak terungkap menggunakan mekanisme dan menganalisa data pada jumlah besar. Dalam

menganalisa keseluruhan atau sebagian *unstructured text*, *text mining* membuat asosiasi satu bagian *text* dengan lainnya berdasarkan aturan tertentu. Hasil yang diharapkan adalah kata baru yang tidak terungkap jelas sebelumnya [3].

Klasifikasi adalah teknik untuk mendapatkan fungsi untuk membedakan jenis kategori atau kelas data. Banyak teknik yang tersedia dalam bidang *data mining* yang dapat digunakan untuk mengolah sejumlah besar data menjadi informasi bermanfaat [4]. Tujuan klasifikasi adalah untuk memprediksi atau memperkirakan kelas dari data baru yang belum memiliki merek. Untuk mencapai tujuan ini, perlu dibuat cara untuk membedakan kelas data dengan metode tertentu [5]. Salah satu metode yaitu menggunakan *Naïve Bayes* adalah metode klasifikasi statistik yang digunakan untuk memprediksi kemungkinan keanggotaan kelas tertentu, menghitung kemungkinan untuk suatu hipotesis, dan menghitung kemungkinan kelas dari setiap kelompok atribut yang ada, serta menentukan kelas mana yang paling optimal [6].

Pada penelitian Joshua Muliawan dan Erick Dazki di dapat hasil penelitian analisis sentimen pemindahan Ibu Kota Negara Indonesia menggunakan tiga algoritma: *naïve bayes*, *knn*, dan *random forest* dan didapatkan nilai akurasi metode Algoritma *Naïve Bayes Classifier* yang didapat nilai keakuratan sebesar 65.26%, Algoritma *K-Nearest Neighbor* sebesar 58.25%, serta Algoritma *Random Forest* sebesar 45.05% [7]. Penelitian selanjutnya dilakukan oleh Subarkah dkk, membahas tentang *Sentiment Analysis On Reviews Of Wom-en's Tops On Shopee Marketplace Using Naïve Bayes Algorithm* dapat disimpulkan hasil pengujian menggunakan *naïve bayes* menunjukkan nilai akurasi sebesar 89% [8].

Terdapat pula penelitian lain yang dilakukan oleh Rafi dkk, mengenai Metode *Support Vector Machine* dan *Naïve Bayes* untuk Analisis Santimen Ibu Kota Nusantara, yaitu *Naïve Bayes*, dan *Support Vector Machine*. Dari pengujian ini, algoritma *Naïve Bayes* memiliki skor akurasi 92% dan algoritma *Support Vector Machine* mendapatkan skor akurasi sebesar 98% [9]. Penelitian lainnya oleh Sarimole dan Septian juga meneliti tentang Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 pada Twitter dengan Metode *Naïve Bayes* dan *Support Vector Machine*. Didapati hasil pengujian model SVM menunjukkan akurasi sebesar 91.61%, sedangkan model *Naïve Bayes* menunjukkan akurasi sebesar 98.80%, yang berarti model *Naïve Bayes* lebih akurat dibanding model *Support Vector Machine* [10].

Melalui pelaksanaan penelitian ini diharapkan dapat mengungkap pemahaman tentang respons masyarakat terhadap kenaikan PPN 12% yang ada di Indonesia apakah cenderung bersifat positif, negatif, atau netral. Selain itu, tujuan penelitian ini adalah untuk menemukan kata-kata yang sering muncul.

## 2. TINJAUAN PUSTAKA

### 2.1. Knowledge Discovery in Databases

KDD merupakan metode yang digunakan untuk mendapat pengetahuan yang berasal dari *database* yang ada. Hasil pengetahuan yang diterima dapat dimanfaatkan untuk basis pengetahuan (*knowledge base*) yang dipergunakan dalam keperluan mengambil keputusan. *Knowledge Discovery in Database* (KDD) adalah suatu teknik pembentukan pola atau rule dalam informasi. Informasi yang dihasilkan didapatkan dari suatu

data yang besar atau dikenal dengan tambang data yang disimpan dalam basis data yang awalnya belum diketahui dan menghasilkan suatu data yang potensial bermanfaat [11].

### 2.2. Natural Language Processing

*Natural Language Processing* (NLP) adalah bidang kecerdasan buatan yang mempelajari interaksi manusia dan komputer menggunakan bahasa alami. Model komputer seperti ini berguna untuk memudahkan komunikasi antara manusia dan komputer yang mencari informasi sehingga keduanya dapat berkomunikasi dalam bahasa alami [12].

### 2.3. Text Mining

*Text Mining* adalah proses inovasi akan informasi atau isu terkini yang sebelumnya tidak terungkap menggunakan mekanisme dan menganalisa data pada jumlah besar. Dalam menganalisa keseluruhan atau sebagian *unstructured text*, *text mining* membuat asosiasi satu bagian *text* dengan lainnya berdasarkan aturan tertentu. Hasil yang diharapkan adalah kata baru yang tidak terungkap jelas sebelumnya [13].

### 2.4. Pre-Processing

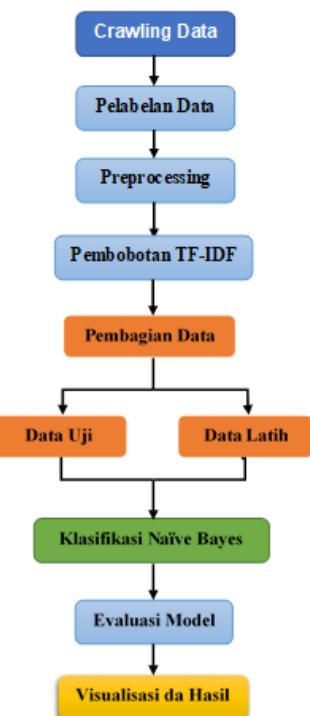
Pengolahan data (*pre-processing*) adalah langkah mempersiapkan data mentah sebelum akan melakukan proses selanjutnya. Tahapan ini memiliki tujuan untuk dapat menyeragamkan kata dan mengurangi volume kata dari sekumpulan dataset serta memudahkan penelitian karena akan membuat data lebih terstruktur. Proses pengolahan data ini akan melewati 6 (enam) tahapan, yaitu: *Cleaning*, *Case folding*, *Normalization*, *Stemming*, *Tokenization*, dan *Stopword Removal* [14].

### 2.5. Naïve Bayes Classifier

*Naïve Bayes Classifier* adalah metoda klasifikasi berdasarkan teori Bayes. Metode ini berdasarkan probabilitas yang dipresentasikan oleh ilmuwan Inggris Thomas Bayes. Peluang dari prediksi masa depan berdasarkan pengalaman di masa sebelumnya. Asumsi yang kuat (*Naïve*) terhadap independensi parameter merupakan ciri-ciri dari metode *Naïve Bayes*. Salah satu kelebihan dari *naïve*, tidak membutuhkan dataset yang besar, mudah dipahami dan masih banyak kelebihan – kelebihan teori *Naïve Bayes* [15].

## 3. KONSEP PERANCANGAN

Metodologi yang diterapkan dalam penelitian ini mengikuti tahapan yang umum dalam *data mining* yang dikenal sebagai proses *Knowledge Discovery in Database* (KDD). Dalam melakukan penelitian ini terdapat beberapa tahapan proses yang perlu dilakukan. Tujuan dibuat tahapan proses tersebut agar penelitian ini dapat berjalan sesuai dengan yang direncanakan. Tahapan proses penelitian dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

### 3.1. Tahap Pengambilan Data

Proses *crawling* data komentar dari media sosial Tiktok dari unggahan pengguna Tiktok dengan menggunakan *Scraper* dengan kata kunci “kenaikan ppn 12%” dilakukan menggunakan bahasa pemrograman *Python* yang selanjutnya disimpan dalam format .csv, dari data yang didapat akan dilanjutkan ke tahap *preprocessing*.

### 3.2. Pelabelan Data

*Labelling* data dilakukan program *lexicon based* dengan menggunakan kamus sentimen. Tujuan dari pelabelan data untuk mengklasifikasikan data berdasarkan karakteristik dari isi komentar atau kalimat. Pelabelan ini dibagi menjadi tiga klasifikasi, yaitu positif, negatif, dan netral.

### 3.3. Preprocessing

- Dalam tahapan *preprocessing* dalam penelitian ini adalah :
1. *Case folding* adalah proses menyamakan huruf pada teks menjadi huruf kecil dimana tidak setiap teks akan konsisten dalam menggunakan huruf kapital.
  2. *Cleansing* merupakan proses menganalisa kualitas data dengan cara mengubah, memodifikasi, atau menghapus data-data yang dianggap tidak lengkap.
  3. Normalisasi adalah proses menghilangkan tanda baca, angka, simbol, link URL dan *username* di dalam teks.
  4. *Stopword Removal* adalah langkah dimana frasa yang tidak penting dalam penambangan teks untuk divisi apa pun dihilangkan.
  5. *Stemming* adalah tahapan yang digunakan dalam pemotongan awal atau akhir kata dengan memperhatikan awalan umum dan sufiks, yang dapat ditemukan dalam kata.
  6. Langkah selanjutnya yaitu tokenisasi dimana pada tahapan ini akan menghilangkan tanda baca yang tidak diperlukan dan

memotong teks menjadi kata, simbol, karakter atau tanda baca, sehingga menjadi token yang dapat dianalisis.

### 3.4. Tahap TF-IDF

TF-IDF adalah proses pembobotan pada masing-masing kata. Pembobotan TF-IDF dinilai penting, hal ini dikarenakan apabila suatu kata lebih sering muncul dalam suatu dokumen maka nilai kontribusinya akan semakin besar akan tetapi jika hal tersebut terjadi pada beberapa dokumen maka kontribusi yang dimiliki akan lebih kecil. TF-IDF menggunakan rumus menghitung nilai bobot dokumen :

$$W_{at} = TF_{at} \times IDF_{ft} \quad (1)$$

Keterangan:

$W_{at}$  = Nilai dokumen ke-d pada kata ke-t

$Tf_{at}$  = Jumlah kata yang dicari dalam dokumen

$IDF_{ft} = Inverse\ document\ frequency\ (\log\left(\frac{D}{df}\right))$

$D$  = Jumlah dokumen

$df$  = Jumlah dokumen yang mengandung kata

### 3.5. Tahap Analisa Sentimen

Pada tahapan ini penggunaan metode untuk melakukan analisis sentimen dilakukan. Metode yang digunakan adalah *Naïve Bayes*. Penggunaan metode *naïve bayes* selain untuk melihat opini public terhadap kenaikan ppn 12 % adalah untuk menguji seberapa akurat metode *naïve bayes* dalam melakukan analisis sentimen publik dengan topik tersebut.

### 3.6. Tahap Evaluasi Model

Setelah proses klasifikasi yang dilakukan di atas maka tahap selanjutnya yaitu tahap evaluasi klasifikasi di mana pada tahap ini akan dilakukan pengujian menggunakan *confusion matrix* dengan matrik ukuran 3x3.

### 3.7. Tahap Visualisasi

Setelah setiap tahap dan proses dilakukan maka selanjutnya adalah tahap visualisasi. Pada penelitian ini, untuk tahap visualisasi dilakukan dengan menggunakan *library Matplotlib* dan *Wordcloud*. *Output* dari visualisasi ini adalah berupa gambar histogram yang juga akan menampilkan hasil akurasi persentase dari polaritas setiap sentimen yang dihasilkan. Sedangkan untuk visualisasi *wordcloud* menampilkan kata yang sering muncul pada setiap sentimennya.

## 4. HASIL DAN PEMBAHASAN

Pada penelitian analisis sentimen ini, data komentar diambil dari media sosial Tiktok melalui *web scraping* menggunakan Apify.com, dengan total 900 data. Metode klasifikasi yang digunakan adalah *Naive Bayes Classifier* dengan fokus pada topik kenaikan PPN 12% di Indonesia. Tujuannya adalah untuk menganalisis perbandingan keberhasilan dalam hal akurasi, presisi, dan *recall*. Proses mencakup beberapa langkah *pre-processing*, validasi, dan evaluasi, serta penandaan sentimen untuk data komentar yang dikategorikan menjadi sentimen

positif, negatif, dan netral. Gambar 2 menggambarkan proses pengambilan data komentar Tiktok.

The screenshot shows the 'TikTok Comments Scraper - Run' interface. It displays the number of results (200), requests (87), and time taken (2020-12-20 15:17). Below this, there's an 'Output' section with tabs for Log, Input, Storage, Live view, and Triggered Integrations. The main area shows a table of scraped data with columns: Comment ID, text, replyCount, replyCommentID, Created at, User name, Video URL, User ID, Comment ID, and clean\_teks. The table contains several rows of data, each representing a comment from a TikTok post.

Gambar 2. Crawling Data dengan APIFY

#### 4.1. Pelabelan Data

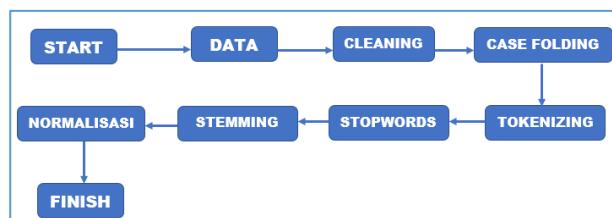
Data perlu dilabeli dan dikategorikan ke dalam 3 kelas sentimen, yakni kelas sentimen positif, negatif, dan netral berdasarkan kata-kata yang terdapat dalam komentar Tiktok dimana *polarity* 1 diberi label positif, *polarity* -1 diberi label negatif, dan *polarity* 0 diberi label netral. Contoh data yang telah dilabeli dapat dilihat pada Gambar 3.

	createTimeISO	uniqueId	text	label	clean_teks
0	2024-12-19T05:10:00.000Z	paulo_statusus	12% guna mewujudkan asas keadian dan kesej... ...	Neutral	guna mewujudkan asas keadian dan kesej...
1	2024-12-19T09:53:50.000Z	o.rando	Mbaak tolong yg paling utama tu bahan pokok gk... ...	Negative	mbaak tolong yg paling utama tu bahan pokok gk...
2	2024-12-19T09:06:39.000Z	icatara88	ooo makanya pada gelisan yaa takut gak bisa ng... ...	Negative	ooo makanya pada gelisan yaa takut gak bisa ng...
3	2024-12-20T12:52:38.000Z	yana_pocom3	gua kaum 58% setuju PPN naik gua sebagai Raky... ...	Negative	gua kaum 58% setuju PPN naik gua sebagai Raky...
4	2024-12-19T05:47:20.000Z	wawan.hemawan777	hanya di indonesia rakyat memilih pajak naik... ...	Negative	hanya di indonesia rakyat memilih pajak naik...
...	...	...	...	...	...
890	2025-01-01T18:37:02.000Z	kaniaamanahkhyari	...	Neutral	...
891	2024-12-26T09:03:50.000Z	malauaraf	...	Neutral	...
893	2025-01-01T16:52:57.000Z	merlangapakaya	...	Neutral	...
896	2024-12-26T05:56:23.000Z	anak_bunda2009	...	Neutral	...
899	2024-12-31T22:23:59.000Z	rzofficial555	@genindra mndra tolong kasih paham	Neutral	[mndra, tolong, kasih, paham]

Gambar 3. Pelabelan Data

#### 4.2. Preprocessing Data

Tahapan *preprocessing* data merupakan langkah penting dalam proses analisis sentimen. Tahap ini dilakukan untuk memastikan data tersusun secara terstruktur sehingga dapat digunakan dalam analisis sentimen. Sesuai dengan yang dijelaskan dalam metodologi penelitian, proses *pre-processing* data mencakup beberapa langkah yaitu *Cleaning*, *Case Folding*, *Normalisasi*, *Stopword Removal*, *Stemming*, dan *Tokenisasi* dapat dilihat pada gambar 4.



Gambar 4. Preprocessing

Hasil *preprocessing* sesudah dilakukannya proses *stemming* dan *tokenized* dapat dilihat pada Gambar 5.

	createTimeISO	uniqueId	text	label	clean_teks
0	2024-12-19T05:10:00.000Z	paulo_statusus	12% guna mewujudkan asas keadian dan kesej... ...	Neutral	[wujud, asas, aditi, sejaher]
1	2024-12-19T09:53:50.000Z	o.rando	Mbaak tolong yg paling utama tu bahan pokok gk... ...	Negative	[mbaak, tolong, utama, bahan, pokok, ppn, baran...]
2	2024-12-19T09:06:39.000Z	icatara88	ooo makanya pada gelisan yaa takut gak bisa ng... ...	Negative	[ooo, gelisan, takut, ngeris, gaya, hidup, bu...]
3	2024-12-20T12:52:38.000Z	yana_pocom3	gua kaum 58% setuju PPN naik gua sebagai Raky... ...	Negative	[guau, tuju, ppn, rakyat, makek, cantum, bagus...]
4	2024-12-19T05:47:20.000Z	wawan.hemawan777	hanya di indonesia rakyat memilih pajak naik... ...	Negative	[Indonesia, rakyat, pajak, naik]
...	...	...	...	...	...
890	2025-01-01T18:37:02.000Z	kaniaamanahkhyari	...	Neutral	...
891	2024-12-26T09:03:50.000Z	malauaraf	...	Neutral	...
893	2025-01-01T16:52:57.000Z	merlangapakaya	...	Neutral	...
896	2024-12-26T05:56:23.000Z	anak_bunda2009	...	Neutral	...
899	2024-12-31T22:23:59.000Z	rzofficial555	@genindra mndra tolong kasih paham	Neutral	[mndra, tolong, kasih, paham]

Gambar 5. Hasil Preprocessing

#### 4.3. Pembobotan TF-IDF

Langkah pembobotan TF-IDF yaitu mekanisme yang mengganti data teks ke bentuk data numerik untuk menghitung jenis setiap kata atau fitur. Untuk simulasi Pembobotan TF-IDF adalah dapat dilihat pada Gambar 6.

```
from sklearn.feature_extraction.text import TfidfVectorizer
tf_idf = TfidfVectorizer(ngram_range=(1,2))
tf_idf.fit(X)
```

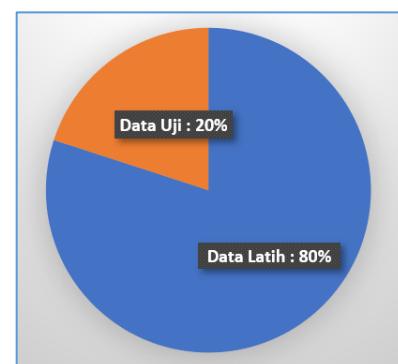
```
TfidfVectorizer
TfidfVectorizer(ngram_range=(1, 2))
```

Gambar 6. Pembobotan TF-IDF

Kode pada Gambar 6 dimaksudkan untuk mengubah teks menjadi representasi TF-IDF dan melakukan transformasi pada data latih dan data uji yang akan diuji.

#### 4.4. Pemisahan Data Latih dan Data Uji

Setelah diperoleh data maka data tersebut dipecah menjadi 2 kelompok yaitu untuk data uji dan data latihan. Pada langkah analisis ini banyaknya data uji 20% sedangkan banyaknya data latih 80% dari total data. Perbandingan data latih dan data uji dapat dilihat pada Gambar 7.



Gambar 7. Perbandingan Data Latih dan Data Uji

#### 4.5. Klasifikasi dengan Naïve Bayes

Pada klasifikasi *Naïve Bayes* (NB) digunakan algoritma *MultinomialNB*. Algoritma ini berfungsi untuk melakukan klasifikasi Naïve Bayes dan mampu mengelola data dalam bentuk teks. Setelah dilakukan pengujian presisi berdasarkan perhitungan akurasi yang telah dilakukan, diperoleh hasil bahwa model mampu melakukan prediksi dengan tingkat akurasi sebesar 96,36%. Hasil keputusan dapat dilihat pada Gambar 8.

```
Jumlah prediksi benar : 663  
Jumlah prediksi salah : 25  
Akurasi pengujian : 96.36627906976744 %
```

Gambar 8. Akurasi *Naïve Bayes*

#### 4.6. Evaluasi *Confusion Matrix*

*Confusion matrix* adalah alat ukur dalam bentuk matriks yang digunakan untuk menentukan tingkat akuratan klasifikasi berbagai kelas berdasarkan algoritma yang diterapkan. Kesimpulan hasil akurasi dari metode *naïve bayes* dapat dilihat pada Gambar 9.

Classification report:				
	precision	recall	f1-score	support
Negatif	0.98	0.96	0.97	240
Netral	0.94	0.98	0.96	275
Positif	0.98	0.95	0.96	173
accuracy			0.96	688
macro avg	0.97	0.96	0.96	688
weighted avg	0.96	0.96	0.96	688

Gambar 9. Hasil Evaluasi *Naïve Bayes*

#### **4.7. Visualisasi WordCloud**

*WordCloud* adalah representasi visual dari sekumpulan kata yang sering muncul dalam sebuah teks. Hasil *visualisasi* kata yang sering muncul pada dataset dapat dilihat pada Gambar 10.



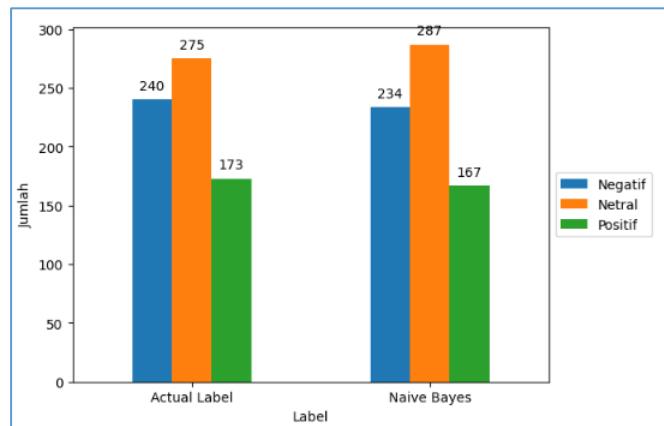
Gambar 10. Hasil Visualisasi dengan *WordCloud*

Kata-kata yang lebih sering muncul akan ditampilkan dalam ukuran yang lebih besar dan warna yang lebih mencolok sedangkan kata-kata yang jarang muncul akan memiliki ukuran yang lebih kecil. Berdasarkan *word cloud* yang ditampilkan: Kata yang paling dominan (berukuran besar) adalah "pajak," "barang," "ppn," "mewah," dan "rakyat." Ini menunjukkan bahwa topik utama dalam teks berkaitan dengan isu pajak (PPN), barang, dan kemewahan yang mungkin sedang menjadi perhatian masyarakat.

Kata lain yang cukup menonjol meliputi "harga," "persen," "kena," dan "negara," yang menunjukkan adanya diskusi terkait kebijakan harga dan dampaknya terhadap masyarakat. Beberapa kata seperti "gaji," "beli," "kerja," dan "butuh" menunjukkan adanya kekhawatiran publik mengenai pengeluaran dan kebutuhan sehari-hari. Kata-kata seperti "internet," "beras," dan "minyak" juga muncul, yang mengindikasikan bahwa topik terkait kebutuhan pokok juga banyak diperbincangkan.

#### **4.8. Hasil Data dalam *Histogram***

Pada tahap ini menampilkan hasil gambar dengan bentuk histogram dapat dilihat pada Gambar 11.



Gambar 11. Hasil Histogram

Pada gambar 11 menampilkan histogram dari total 900 data komentar media sosial Tiktok yang menunjukkan bahwa persentase opini masyarakat untuk kelas pada sentimen positif sebesar 240 komentar data aktual dan 234 komentar data *Naïve Bayes*, untuk sentimen negatif sebesar 173 komentar data aktual dan 167 komentar data *Naïve Bayes*, sedangkan untuk sentimen netral sebesar 275 komentar data aktual dan 287 komentar data *Naïve Bayes*. Jadi hasilnya mayoritas opini masyarakat di media sosial Tiktok bernilai netral.

## 5. KESIMPULAN DAN SARAN

Metode *Naïve Bayes* yang telah diimplementasikan menggunakan bahasa pemrograman *Python* berhasil diterapkan untuk menganalisis sentimen masyarakat di media sosial Tiktok terhadap kenaikan PPN 12% dengan mengklasifikasikan sebanyak 900 data komentar ke kelas sentimen positif, negatif, dan netral. Teknik pembobotan TF-IDF digunakan untuk meningkatkan akurasi analisis dan hasilnya, menunjukkan bahwa sentimen masyarakat di tiktok cenderung Netral.

Metode klasifikasi Naïve Bayes terbukti cukup akurat dalam mengklasifikasikan sentimen masyarakat di Tiktok. Hal ini terlihat dari tingkat akurasi klasifikasi yang dihasilkan yaitu 96%. Kata – kata populer yang muncul dalam topik kenaikan PPN 12% di Tiktok antara lain adalah "pajak," "barang," "ppn," "mewah," dan "rakyat", yang tercermin dalam *wordcloud* dari analisis sentimen.

## DAFTAR PUSTAKA

- [1] Q. A'yuniyah and M. Reza, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di SMA Negeri 15 Pekanbaru," *Indonesian Journal of Informatic Research and Software Engineering.*, vol. 3, no. 1, pp. 39–45, 2023, doi: 10.57152/ijirse.v3i1.484.

[2] D. Oktavia, Y. R. Ramadahan, and M. Minarto, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1,

- pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.
- [3] M. A. Djamarudin, A. Triayudi, and E. Mardiani, “Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan Metode Naïve Bayes Classifier,” *JTIK: Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 2, pp. 161–166, 2022, doi: 10.35870/jtik.v6i2.398.
- [4] N. Lestari, O. S. Riza, and R. Ardinal, “Implementation Of Text Mining And Pattern Discovery With Naive Bayes Algorithm For Classification Of Text Documents,” *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 14, no. 1, pp. 88–102, 2023, doi: <https://doi.org/10.31849/digitalzone.v14i1.13596>
- [5] H. Paul, A. S. Wiguna, and H. Santoso, “Penerapan Algoritma Support Vector Machine Dan Naive Bayes Untuk Klasifikasi Jenis Mobil Terlaris Berdasarkan Produksi Di Indonesia,” *JATI : Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 1, pp. 39–44, 2023, doi: 10.36040/jati.v7i1.5555.
- [6] F. Amaliah and I. K. Dwi Nuryana, “Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter,” *JINACS: Journal of Informatics and Computer Science*, vol. 3, no. 03, pp. 384–393, 2022, doi: 10.26740/jinacs.v3n03.p384-393.
- [7] J. Muliawan, and E. Dazki, “Sentiment Analysis of Indonesia’S Capital City Relocation Using Three Algorithms: Naïve Bayes, KNN, and Random Forest,” *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 5, pp. 1227–1236, 2023, doi: 10.52436/1.jutif.2023.4.5.1436.
- [8] P. Subarkah, P. W. Rahayu, I. Darmayanti, and R. Riyanto, “Sentiment Analysis on Reviews of Women’S Tops on Shopee Marketplace Using Naive Bayes Algorithm,” *JITK: Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, vol. 9, no. 1, pp. 126–133, 2023, doi: 10.33480/jitk.v9i1.4179.
- [9] M. R. Akbar, S. Defit, and Sumijan, “Metode Support Vector Machine dan Naïve Bayes untuk Analisis Santimen Ibu Kota Nusantara,” *Jurnal KomtekInfo*, vol. 11, no. 4, pp. 323–331, 2024, doi: <https://doi.org/10.35134/komtekinfo.v11i4.579>.
- [10] F. M. Sarimole, and W. Septian, “Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine,” *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 890–899, 2024, [Online]. Available: <http://ejournal.sisfokomtek.org/index.php/saintek/article/view/1359>
- [11] R. Fauziah, and A. I. Purnamasari, “Implementasi Algoritma K-Means pada Kasus Kekerasan Anak dan Perempuan Berdasarkan Usia,” *Hello World: Jurnal Ilmu Komputer*, vol. 2, no. 1, pp. 34–41, 2023, doi: 10.56211/helloworld.v2i1.232.
- [12] M. Furqan, S. Sriani, and M. N. Shidqi, “Chatbot Telegram Menggunakan Natural Language Processing,” *WJIT: Walisongo Journal of Information Technology*, vol. 5, no. 1, pp. 15–26, 2023, doi: 10.21580/wjit.2023.5.1.14793.
- [13] J. Florensius Sianipar, Y. R. Ramadhan, and I. Jaelani, “Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naïve Bayes,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 360–367, 2023, doi: 10.30865/klik.v4i1.1033.
- [14] R. B. Dahlian, and D. Sitanggang, “Sentiment Analysis of Digital Television Migration on Twitter Using Naïve Bayes Multinomial Comparison, Support Vector Machines, and Logistic Regression Algorithms,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 2, pp. 280–288, 2023, doi: 10.32736/sisfokom.v12i2.1668.
- [15] D. Hindarto, R. E. Indrajit, and E. Dazki, “Perbandingan Kinerja Akurasi Klasifikasi K-NN, NB dan DT pada APK Android,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 1, pp. 486–503, 2022, doi: 10.35957/jatisi.v9i1.1542.