

# Analisis Sentimen Komentar Tiktok Tentang Kasus TPPO TKI Di Kamboja Menggunakan Klasifikasi Naive Bayes Untuk Mengukur Opini Publik

*Sentiment Analysis Of Tiktok Comments On The TPPO TKI Case In Cambodia Using Naive Bayes Classification To Measure Public Opinion*

Celestina Florecita Mariwy<sup>1</sup>, Chelsy Jemima Grace Watratan<sup>2</sup>, Raimondo Dos Santos<sup>3</sup>, Dessy Natalia Reba<sup>4</sup>, Rio Febrian Sabono<sup>5</sup>, Noudia Inex Pasiakan<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Program Studi S1 Teknik Informatika, Fakultas Teknik, Universitas Papua

E-mail: <sup>1</sup>\*fmariwy@gmail.com, <sup>2</sup>chelsywateratan75@gmail.com,

<sup>3</sup>raimondodossantos5@gmail.com, <sup>4</sup>rebadavina@gmail.com, <sup>5</sup>riosabono11@gmail.com,

<sup>6</sup>inexi207@gmail.com

*\*Corresponding author*

Received 15 May 2025; Revised 27 May 2025; Accepted 30 May 2025

**Abstrak**-Penelitian ini bertujuan untuk memahami opini publik terhadap kasus-kasus Tindak Pidana Perdagangan Orang (TPPO) yang dialami pekerja migran atau Tenaga Kerja Indonesia (TKI) di Kamboja pada tahun 2025. Penelitian ini dilakukan dengan menganalisis sentimen komentar di TikTok. Penelitian ini menggunakan kekuatan algoritma Naive Bayes untuk membaca sentimen tersembunyi di balik komentar dan mengkategorikannya ke dalam spektrum positif, netral, dan negatif, dengan mempertimbangkan semakin populernya media sosial dan prevalensi penipuan pekerjaan online. Dataset ini terdiri dari 7.208 komentar yang dikumpulkan dari platform TikTok dengan menggunakan teknik crawling, yang kemudian diproses dan diberi tag untuk dianalisis. Hasilnya menunjukkan bahwa mayoritas pengguna TikTok cenderung memiliki opini netral terhadap isu TPPO. Model Naive Bayes mencapai tingkat akurasi sebesar 71,06%. Selain meningkatkan pemahaman masyarakat terhadap isu tersebut, penelitian ini diharapkan dapat menjadi amunisi berharga bagi pemerintah Indonesia untuk mengatasi TPPO di masa depan.

Kata kunci: Analisis Sentimen, Naive Bayes, TikTok, TPPO, TKI, Opini Publik

**Abstract**-The goal of this study is to understand public opinion regarding cases of human trafficking experienced by migrant workers, specifically Indonesian migrant workers in Cambodia in 2025. It does so by analyzing the sentiment of comments on TikTok. This research utilizes the power of Naive Bayes algorithm to read the hidden sentiment behind the comments and categorize them into positive, neutral, and negative spectrums, considering the growing popularity of social media and the prevalence of online job scams. The dataset consists of 7,208 comments gathered from the TikTok platform using crawling techniques, preprocessed and tagged for analysis. The results show that the majority of the TikTok users tend to have a neutral opinion on the issue of TPPO. The Naive Bayes model achieved an accuracy rate of 71.06%. In addition to increasing public understanding of the issue, this study is expected to provide valuable ammunition for the Indonesian government to combat TPPO in the future.

**Keywords:** Sentiment Analysis, Naive Bayes, TikTok, TPPO, TKI, Public Opinion

## 1. PENDAHULUAN

Studi ini dilatarbelakangi oleh kemajuan pesat teknologi internet dewasa ini. Akan tetapi, perkembangan teknologi yang begitu cepat juga menimbulkan potensi risiko. Pada dasarnya, setiap orang tentu mendambakan kehidupan yang sejahtera. Sayangnya, fakta berbicara lain. Rendahnya kualitas pendidikan, terutama di Indonesia kemudian berbagai permasalahan yang

dihadapi dalam pendidikan [1] dan ekonomi telah menciptakan celah lebar bagi sindikat kejahatan untuk beraksi. Banyak warga Indonesia yang polos dan kesulitan ekonomi menjadi korban mudah penipuan daring dan tawaran pekerjaan fiktif yang menyesatkan.

Penyalahgunaan teknologi oleh pelaku kejahatan transnasional semakin memperparah situasi ini. Para pelaku mencari korban dengan menyebarkan berita palsu berupa lowongan pekerjaan menggiurkan, menggunakan identitas palsu, dan membuat perjanjian kerja fiktif [2]. Modus operandi ini memanfaatkan kecanggihan media sosial dan platform digital untuk menjangkau calon korban yang rentan secara ekonomi. Situasi ini menggarisbawahi urgensi bagi masyarakat untuk meningkatkan kewaspadaan dan bersikap bijaksana dalam menyikapi informasi yang beredar di media sosial, khususnya tawaran pekerjaan dengan iming-iming yang tidak realistis.

*Human trafficking* atau perdagangan manusia merupakan sebuah tindakan kriminal transnasional yang terjadi baik di dalam maupun melampaui batas wilayah suatu negara. Sebagaimana kegiatan kriminal lainnya, perdagangan manusia merupakan aktivitas ilegal yang bersifat rahasia sehingga sulit dideteksi dan diukur. Kebijakan "tanpa sekat" antar negara, ironisnya turut memicu peningkatan kasus perdagangan manusia. Asia Tenggara pun tercatat sebagai salah satu wilayah dengan kerentanan tertinggi terhadap kejahatan ini. Beberapa negara seperti Myanmar, Kamboja, Filipina, Vietnam, dan Indonesia menjadi negara asal utama bagi korban perdagangan manusia. Sementara itu, Singapura dan Malaysia lebih sering menjadi destinasi atau negara tujuan dari praktik perdagangan manusia tersebut [3].

Kompleksitas permasalahan TPPO ini menunjukkan bahwa penanganannya tidak dapat dilakukan secara sepihak oleh satu negara saja. Diperlukan kerjasama internasional yang komprehensif, baik melalui kerjasama bilateral, multilateral, maupun regional untuk memerangi kejahatan transnasional ini secara efektif. Sebagai contoh, Indonesia dan Kamboja telah menginisiasi kolaborasi bilateral dalam upaya penanggulangan kasus TPPO, terutama yang melibatkan warga negara Indonesia sebagai korban dalam sektor perjudian daring, yang diwujudkan melalui forum seperti Pertemuan Bilateral Indonesia-Kamboja tentang Masalah Imigrasi. Namun demikian, upaya pencegahan yang paling fundamental harus dimulai dari meningkatkan kesadaran, mendidik populasi yang rentan, dan mengurangi permintaan tenaga kerja dan layanan yang diperjualbelikan sangat penting dalam menciptakan masyarakat yang lebih aman[4].

Kasus TPPO TKI di Kamboja pada tahun 2025 ini memicu berbagai tanggapan dari masyarakat. Opini publik ini dapat dimanfaatkan oleh pemerintah sebagai bahan evaluasi di masa depan. Namun, karena pendapat masyarakat belum diketahui secara pasti, analisis sentimen diperlukan untuk memahami pandangan mereka terhadap kasus ini. Analisis sentimen tidak hanya dapat dilakukan melalui survei langsung, tetapi juga melalui media sosial yang mencerminkan perilaku masyarakat Indonesia [5]. Sentimen, dalam definisi KBBI, adalah emosi yang membingkai cara kita memandang suatu persoalan. Analisis sentimen didefinisikan sebagai proses sistematis yang berfokus pada pemahaman, ekstraksi, dan pengolahan data tekstual dengan tujuan mengidentifikasi polaritas sentimen yang terkandung dalam pernyataan atau opini [6].

TikTok menjadi media yang populer bagi masyarakat untuk menyebarkan informasi [7] mengenai kasus ini. Hal ini memerlukan perhatian serius, mengingat banyaknya fitur dan kemudahan komunikasi yang ditawarkan oleh TikTok [8]. TikTok sangat up to date terhadap kasus ini dan berperan penting dalam membentuk opini dan perilaku publik, terutama di kalangan generasi muda, karena merupakan platform berbagi video pendek yang populer.

Penelitian ini mengadopsi algoritma Naïve Bayes (NB), sebuah metode klasifikasi data yang didasarkan pada prinsip probabilitas. Keunggulan Naive Bayes terletak pada tingkat akurasi dan kecepatannya yang tinggi dalam mengolah data berukuran besar. Proses klasifikasi data terdiri dari dua tahap: pelatihan dengan contoh (training model) dan klasifikasi data yang kategorinya belum diketahui[9].

Studi analisis sentimen sebelumnya oleh [10] memanfaatkan algoritma Naive Bayes untuk mengklasifikasikan opini penggemar sepak bola di Twitter terkait antusiasme terhadap Indonesia

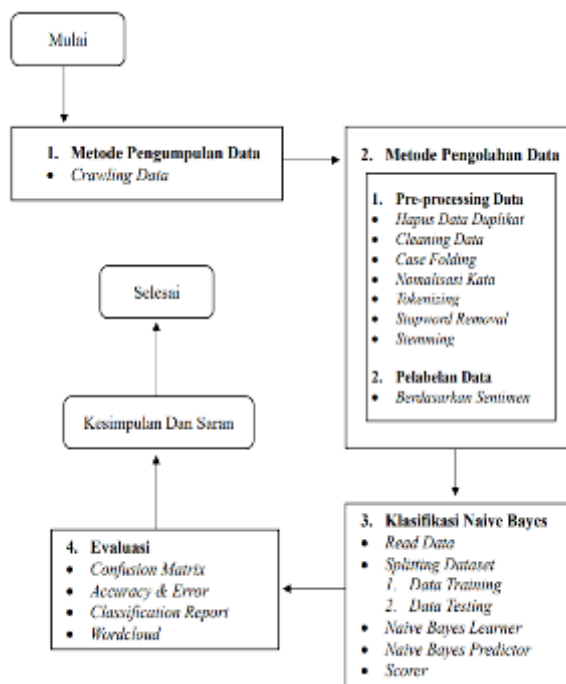
sebagai tuan rumah Piala Dunia U-17. Temuan studi tersebut menunjukkan akurasi yang tinggi, mencapai sekitar 97% pada data uji, dengan nilai presisi, recall, dan F1-score yang mengindikasikan kinerja model yang positif dalam klasifikasi sentimen. Penelitian lain yang membandingkan metode K-NN, Naive Bayes, dan Decision Tree untuk analisis sentimen opini terhadap PT PAL Indonesia, menunjukkan bahwa metode Naive Bayes menghasilkan akurasi tertinggi sebesar 84,08%, diikuti oleh K-NN (83,38%) dan Decision Tree (81,09%) [11]. Dengan demikian, terkonfirmasi bahwa algoritma Naive Bayes sangat efektif dalam membaca dan mengelompokkan sentimen.

Studi ini dilaksanakan dengan tujuan untuk memahami bagaimana opini publik terhadap kasus TPPO yang menimpa TKI di Kamboja pada tahun 2025, yang dianalisis melalui komentar pada video TikTok dan dikelompokkan menjadi sentimen positif, netral, dan negatif. Belum ada penelitian yang mendalam tentang opini publik terkait TPPO di TikTok, khususnya yang dialami TKI di Kamboja pada tahun 2025.

Hasil analisis opini publik ini diharapkan dapat menjadi dasar evaluasi bagi Pemerintah Indonesia dalam upaya mencegah terulangnya kasus serupa di waktu mendatang. Lebih dari itu, riset ini juga ingin mengedukasi masyarakat tentang bahaya TPPO, serta mengajak mereka turut serta mengawasi dan mendukung upaya pemerintah melindungi para pahlawan devisa. Studi ini menyajikan wawasan mengenai sentimen publik terhadap isu TPPO (Tindak Pidana Perdagangan Orang), yang diekstraksi dari platform daring yang dominan di kalangan generasi muda. Temuan analisis ini berpotensi menjadi landasan evaluasi kebijakan pemerintah dan peningkatan kesadaran publik terkait isu tersebut.

## 2. METODE PENELITIAN

Pada riset ini, digunakan metode *Naive Bayes* untuk mengklasifikasikan komentar dari pengguna TikTok mengenai kasus TPPO yang dialami TKI di Kamboja. Proses penelitian ini meliputi beberapa fase, dimulai dengan pengumpulan data, dilanjutkan dengan pengolahan data, penerapan *Naive Bayes*, hingga evaluasi hasil, sebagaimana diilustrasikan pada Gambar 1.



Gambar 1 Fase Penelitian.

## 2.1 Pengumpulan Data

Pemilihan metode ini didasari oleh meningkatnya popularitas media sosial dan maraknya praktik penipuan pekerjaan daring. Data penelitian bersumber dari TikTok, sebuah platform yang sangat digemari oleh Generasi Z, yang mewakili lebih dari 60% basis penggunaannya. Generasi Z didefinisikan sebagai individu yang lahir setelah tahun 1996 [12]). Platform ini memungkinkan pengguna untuk berbagi dan mendiskusikan isu-isu sosial secara interaktif.

Kasus Tindak Pidana Perdagangan Orang (TPPO) Tenaga Kerja Indonesia (TKI) di Kamboja menjadi viral dan banyak diperbincangkan di TikTok, mencerminkan opini publik yang beragam dan up-to-date terhadap isu ini. Pada 22 April 2025, kami melakukan data scraping komentar TikTok dengan teknik *crawling* data menggunakan *TikTok Comment Scraper* dari Apify (<https://apify.com/quacker/tiktok-comments-scraper>). Dari proses ini, berhasil dikumpulkan sebanyak 7.208 komentar yang langsung di-export ke Excel untuk analisis sentimen lebih lanjut.

## 2.2 Pengolahan Data

Pada fase ini, data diolah melalui proses pra-pemrosesan (*pre-processing*) dan pelabelan data. *Pre-processing* merupakan serangkaian tahapan untuk mempersiapkan dan memodifikasi data sebelum digunakan oleh model atau algoritma. Tujuannya adalah untuk meningkatkan mutu data, mereduksi gangguan (*noise*), memperbaiki kontras, menyesuaikan dimensi, serta menghilangkan bagian-bagian yang tidak relevan. Proses ini memastikan bahwa data yang digunakan dalam pelatihan model memiliki kualitas yang baik, sehingga menghasilkan prediksi yang lebih akurat. Pelabelan data dilakukan untuk memberikan label pada setiap data, sehingga dapat diidentifikasi apakah kontennya mengandung sentimen positif atau negatif. Proses ini dijalankan secara otomatis dengan memanfaatkan kamus *Lexicon*, yang memuat kumpulan kata beserta bobotnya untuk membantu menentukan sentimen dalam teks. Nilai polaritas dihitung berdasarkan total bobot seluruh kata dalam komentar, dan hasilnya digunakan sebagai label untuk menentukan apakah komentar tersebut mengandung sentimen positif, netral, atau negatif. Komentar dikategorikan positif apabila nilai polaritasnya lebih besar dari nol, negatif jika kurang dari nol, dan netral jika sama dengan nol.

## 2.3 Klasifikasi Naive Bayes

Algoritma *Naive Bayes*, meskipun berbasis pada asumsi independensi fitur yang disederhanakan (naif), menawarkan pendekatan klasifikasi teks yang komputasionalnya efisien. Dalam konteks analisis sentimen komentar TikTok terkait isu TPPO di Kamboja, *Naive Bayes* digunakan untuk kategorisasi teks. Keberhasilan penerapan metode ini bergantung pada validitas asumsi independensi dan kualitas *pre-processing* data komentar.

## 2.4 Evaluasi

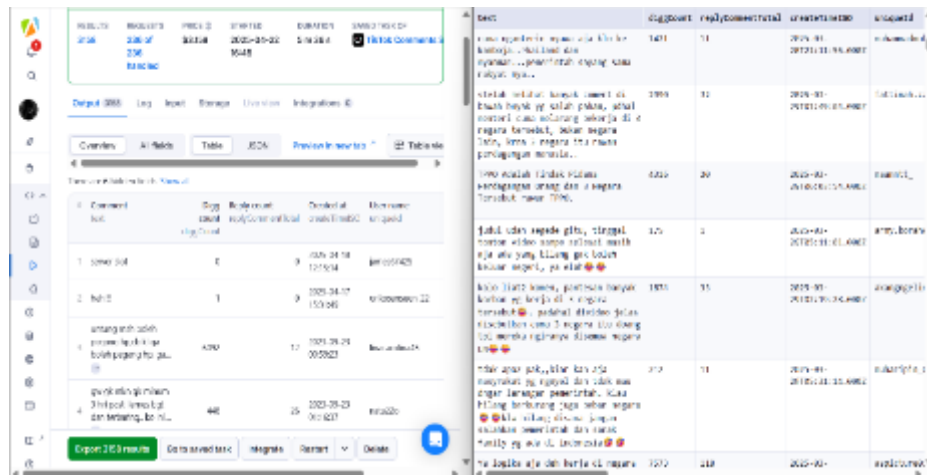
Penilaian performa model melibatkan serangkaian pengujian empiris menggunakan data latih dan data uji, yang bertujuan untuk mengukur berbagai metrik performa kunci. Analisis komparatif antara performa model pada kedua set data ini memungkinkan identifikasi potensi *bias* dan *overfitting*. Secara khusus, evaluasi menyoroti kemampuan model dalam melakukan klasifikasi sentimen secara akurat, baik dalam data yang digunakan untuk melatih model maupun data independen yang merepresentasikan skenario aplikasi dunia nyata.

# 3. HASIL DAN PEMBAHASAN

## 3.1 Pengumpulan Data

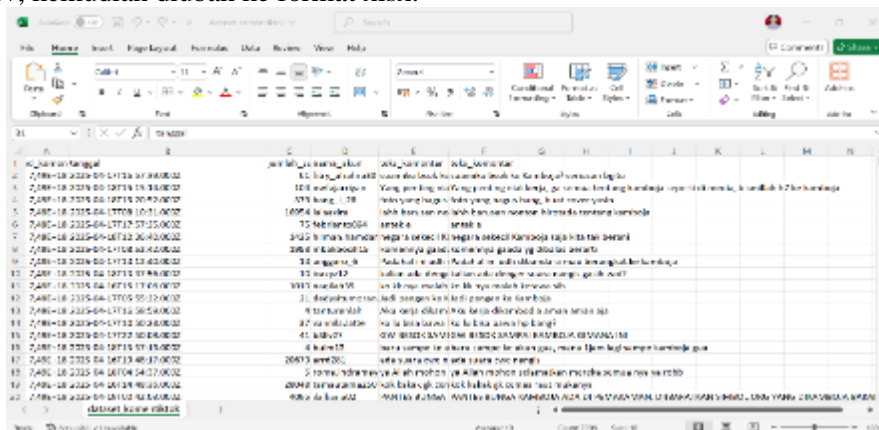
### 3.1.1 Scraping Data Tiktok

Proses pengambilan data dilakukan terhadap 3 video yang berbeda dari beberapa akun yang berbeda, yaitu dari akun @tribunjakartaofficial, @lingkarnews, dan @aphoyka2ngprabu.



Gambar 2 Proses Scraping Komentar.

Gambar 2 menunjukkan proses *scraping* komentar di platform web *scraping* *apify.com* dari 3 video berbeda di akun @tribunjakartaofficial, @lingkarnews, dan @aphoyka2ngprabu. Selanjutnya, semua komentar yang telah di-*scraping* disimpan dalam format csv, kemudian diubah ke format xlsx.



Gambar 3 Hasil Scraping Komentar.

Gambar 3 menunjukkan hasil *scraping* komentar pada postingan Tiktok akun @tribunjakartaofficial, @lingkarnews, dan @aphoyka2ngprabu yang menghasilkan 7208 komentar. Video akun @tribunjakartaofficial menghasilkan 1255 komentar, video akun @lingkarnews menghasilkan 3583 komentar, dan video akun @aphoyka2ngprabu menghasilkan 2190 komentar. Tabel 1 menyajikan tampilan hasil *scraping* komentar dalam bentuk tabel.

Tabel 1 Data Hasil Komentar

Akun Tiktok	Tema Video	Jumlah Data
@tribunjakartaofficial	Soleh Dermawan (24), Pemuda Asal Kota Bekasi Tewas Diduga Jadi Korban Tindak Pidana Perdagangan Orang (TPPO) di Kamboja.	1255 Komentar
@lingkarnews	Pemerintah Larang WNI Bekerja di Thailand Kamboja Dan Myanmar	3583 Komentar

@aphoyka2ngprabu	Oknum Polri Terlibat TPPO Kamboja Jaringan Jual Beli Ginjal Manusia	2190 Komentar
<b>Total</b>	<b>7208</b>	<b>Komentar</b>

### 3.2 Pre-processing Data

Tahap pra-pemrosesan (*pre-processing*) melibatkan *text mining* untuk menganalisis teks dan mengekstrak informasi yang berguna. Melalui penerapan *text mining*, sebuah proses iteratif dua langkah diinisiasi dengan penataan tekstual data mentah, yang kemudian dieksploitasi untuk mengekstrak inteligensi dan wawasan yang relevan. Berikut adalah tahapan-tahapan *pre-processing* data:

#### 3.2.1 Hapus Data Duplikat

Untuk menjamin mutu dan keutuhan data analisis, data duplikat dihilangkan. Komentar yang dianggap duplikat adalah komentar yang teksnya benar-benar identik. Proses ini menghapus komentar dengan teks yang sama persis untuk mencegah redundansi dan bias, terutama dalam analisis sentimen yang memerlukan data representatif dan unik. Tujuannya adalah untuk menghindari distorsi hasil analisis akibat komentar yang berulang, sehingga menghasilkan analisis yang lebih valid dan objektif[13].

##### PROSES HAPUS DATA DUPLIKAT

```
[ ] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7208 entries, 0 to 7207
Data columns (total 1 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   teks_komentar  7208 non-null    object
dtypes: object(1)
memory usage: 56.4+ KB

[ ] df.drop_duplicates(subset = "teks_komentar", keep = 'first', inplace = True)
df.info()
df.head(11)

<class 'pandas.core.frame.DataFrame'>
Index: 6960 entries, 0 to 7207
Data columns (total 1 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   teks_komentar  6960 non-null    object
dtypes: object(1)
memory usage: 188.8+ KB
```

Gambar 4 Hasil Hapus Data Duplikat.

Gambar 4 menunjukkan bahwa sebanyak 68 data duplikat telah dihapus dari 7208 komentar awal, sehingga tersisa 6960 data bersih yang digunakan untuk analisis lebih lanjut demi mendapatkan insight yang lebih akurat dan terpercaya.

#### 3.2.2 Cleaning Data

*Cleaning* data atau pembersihan data adalah proses menghilangkan bagian-bagian data teks yang tidak sesuai atau tidak penting.

	teks_komentar	Cleaning
0	suamiku bsok ke Kamboja? seriusan bgitu	suamiku bsok ke Kamboja? seriusan bgitu
1	Yang penting niat kerja, ga semua tentang kamb...	Yang penting niat kerja ga semua tentang kambo...
2	foto yang bagus bang, buat cover yasin	foto yang bagus bang buat cover yasin
3	lahh barusan nonton hirohada tentang kamboja	lahh barusan nonton hirohada tentang kamboja
4	antek e	antek e
5	negara sekecil Kamboja saja kita tak berani	negara sekecil Kamboja saja kita tak berani
6	komennya gaada yg dibalas berarti	komennya gaada yg dibalas berarti
7	Padahal ini udh dibandara mau berangkat ke kam...	Padahal ini udh dibandara mau berangkat ke kam...
8	kalian ada denger suara nangis gash woi?	kalian ada denger suara nangis gash woi?
9	ko kk nya malah ketawa sih	ko kk nya malah ketawa sih
10	Jadi pengen ke Kamboja	Jadi pengen ke Kamboja

Gambar 5 Hasil *Cleaning* Data.

Gambar 5 menampilkan hasil pembersihan data yang meliputi penghilangan retweet, tanda baca, mention (@), URL, angka yang tidak relevan, serta karakter khusus lainnya yang tidak dibutuhkan untuk analisis lebih lanjut.

### 3.2.3 Case Folding

*Case folding* merupakan teknik konversi teks yang mengubah seluruh karakter menjadi huruf kecil, dengan tujuan mempermudah identifikasi dan pemrosesan informasi.

	teks_komentar	Cleaning	Case Folding
0	suamiku bsok ke Kamboja? seriusan bgitu	suamiku bsok ke Kamboja? seriusan bgitu	suamiku bsok ke kamboja? seriusan bgitu
1	Yang penting niat kerja, ga semua tentang kamb...	Yang penting niat kerja ga semua tentang kambo...	yang penting niat kerja, ga semua tentang kamb...
2	foto yang bagus bang, buat cover yasin	foto yang bagus bang buat cover yasin	foto yang bagus bang, buat cover yasin
3	lahh barusan nonton hirohada tentang kamboja	lahh barusan nonton hirohada tentang kamboja	lahh barusan nonton hirohada tentang kamboja
4	antek e	antek e	antek e
5	negara sekecil Kamboja saja kita tak berani	negara sekecil Kamboja saja kita tak berani	negara sekecil kamboja saja kita tak berani
6	komennya gaada yg dibalas berarti	komennya gaada yg dibalas berarti	komennya gaada yg dibalas berarti
7	Padahal ini udh dibandara mau berangkat ke kam...	Padahal ini udh dibandara mau berangkat ke kam...	padahal ini udh dibandara mau berangkat ke kam...
8	kalian ada denger suara nangis gash woi?	kalian ada denger suara nangis gash woi?	kalian ada denger suara nangis gash woi?
9	ko kk nya malah ketawa sih	ko kk nya malah ketawa sih	ko kk nya malah ketawa sih
10	Jadi pengen ke Kamboja	Jadi pengen ke Kamboja	jadi pengen ke kamboja

Gambar 6 Hasil *Case Folding* Data.

Gambar 6 memperlihatkan implementasi *case folding*, yakni proses reduksi varian huruf kapital menjadi huruf kecil yang diterapkan sebagai langkah *pre-processing* untuk analisis leksikal berikutnya.

### 3.2.4 Normalisasi Kata

Normalisasi adalah proses standarisasi kata-kata dalam teks komentar agar selaras dengan aturan tata bahasa yang diakui oleh KBBI.

	teks_komentar	Cleaning	Case Folding	Normalisasi
0	suamiku bsok ke Kamboja? seriusan bgitu	suamiku bsok ke Kamboja? seriusan bgitu	suamiku bsok ke kamboja? seriusan bgitu	suamiku besok ke kamboja? seriusan begitu
1	Yang penting niat kerja, ga semua tentang kamb...	Yang penting niat kerja ga semua tentang kambo...	yang penting niat kerja, ga semua tentang kamb...	yang penting niat kerja, tidak semua tentang k...
2	foto yang bagus bang, buat cover yasin	foto yang bagus bang buat cover yasin	foto yang bagus bang, buat cover yasin	foto yang bagus bang, buat cover yasin
3	lahh barusan nonton hirohada tentang kamboja	lahh barusan nonton hirohada tentang kamboja	lahh barusan nonton hirohada tentang kamboja	lah barusan menonton hirohada tentang kamboja
4	antek e	antek e	antek e	antek e
5	negara sekecil Kamboja saja kita tak berani	negara sekecil Kamboja saja kita tak berani	negara sekecil kamboja saja kita tak berani	negara sekecil kamboja saja kita tak berani
6	komennya gaada yg dibalas berarti	komennya gaada yg dibalas berarti	komennya gaada yg dibalas berarti	komennya yang dibalas berarti
7	Padahal ini udh dibandara mau berangkat ke kam...	Padahal ini udh dibandara mau berangkat ke kam...	padahal ini udh dibandara mau berangkat ke kam...	padahal ini sudah dibandara mau berangkat ke k...
8	kalian ada denger suara nangis gash woi?	kalian ada denger suara nangis gash woi?	kalian ada denger suara nangis gash woi?	kalian ada dengar suara menangis woi?
9	ko kk nya malah ketawa sih	ko kk nya malah ketawa sih	ko kk nya malah ketawa sih	kok kakak ya malah ketawa sih
10	Jadi pengen ke Kamboja	Jadi pengen ke Kamboja	jadi pengen ke kamboja	jadi pengen ke kamboja

Gambar 7 Hasil Normalisasi Kata.

Gambar 7 menyajikan contoh bagaimana kata-kata informal diubah menjadi bentuk formal sesuai dengan pedoman KBBI.

### 3.2.5 Tokenizing

*Tokenizing* atau tokenisasi merupakan proses memotong teks komentar menjadi kata-kata kecil atau bagian-bagian kecil yang disebut token.

	teks_komentar	Cleaning	Case Folding	Normalisasi	Tokenize
0	suamiku bosk ke Kamboja? seriusan bgtlu	suamiku bosk ke Kamboja? seriusan bgtlu	suamiku bosk ke kamboja? seriusan bgtlu	suamiku bosk ke kamboja? seriusan bgtlu	[suamiku, bosk, ke, kamboja?, seriusan, bgtlu]
1	Yang penting niat kerja, ga semua lentang kambo...	Yang penting niat kerja, ga semua lentang kambo...	yang penting niat kerja, ga semua lentang kambo...	yang penting niat kerja, tidak semua lentang k...	[yang, penting, niat, kerja, tidak, semua, le...
2	foto yang bagus bang, buat cover yasin	foto yang bagus bang buat cover yasin	foto yang bagus bang, buat cover yasin	foto yang bagus bang, buat cover yasin	[foto, yang, bagus, bang, buat, cover, yasin]
3	lahh barusan nonton hiroada tentang kamboja	lahh barusan nonton hiroada tentang kamboja	lahh barusan nonton hiroada tentang kamboja	lah barusan menonton hiroada tentang kamboja	[lah, barusan, menonton, hiroada, tentang, ka...
4	antek e	antek e	antek e	antek e	[antek, e]
5	negara secekl Kamboja saja kita tak berani	negara secekl Kamboja saja kita tak berani	negara secekl kamboja saja kita tak berani	negara secekl kamboja saja kita tak berani	[negara, secekl, kamboja, saja, kita, tak, be...
6	komenya gaada yg dibales berarti	komenya gaada yg dibales berarti	komenya gaada yg dibales berarti	komenya yang dibales berarti	[komenya, yang, dibales, berarti]
7	Padahal ini udh dibandara mau berangkat ke kam...	Padahal ini udh dibandara mau berangkat ke kam...	padahal ini udh dibandara mau berangkat ke kam...	padahal ini sudah dibandara mau berangkat ke k...	[padahal, ini, sudah, dibandara, mau, berangkat...
8	kalian ada denger suara nangis gash wol?	kalian ada denger suara nangis gash wol?	kalian ada denger suara nangis gash wol?	kalian ada denger suara menangis wol?	[kalian, ada, denger, suara, menangis, wol?]
9	ko kik nya malah ketawa sih	ko kik nya malah ketawa sih	ko kik nya malah ketawa sih	kok kakak ya malah ketawa sih	[kok, kakak, ya, malah, ketawa, sih]
10	Jadi pengen ke Kamboja	Jadi pengen ke Kamboja	jadi pengen ke kamboja	jadi pengen ke kamboja	[jadi, pengen, ke, kamboja]

Gambar 8 Hasil *Tokenizing* Data.

Gambar 8 menunjukkan bagaimana aliran teks dipilah dan diurai menjadi komponen-komponen terkecilnya, mulai dari kata-kata hingga simbol, layaknya mozaik yang dirangkai dari pecahan-pecahan bermakna.

### 3.2.6 Stopword Removal

Proses penghapusan *stopwords* dilakukan untuk membersihkan teks dari kata-kata yang kurang informatif, seperti konjungsi dan preposisi, sehingga analisis dapat lebih difokuskan pada kata-kata kunci. Daftar *stopwords* yang digunakan dalam penelitian ini bersumber dari korpus resmi NLTK, yang tersedia melalui *package stopwords* dan dapat diakses melalui fungsi `stopwords.words()`. Implementasi NLTK dalam Python memungkinkan identifikasi dan penghapusan *stopwords* seperti "if", "in", "it", "to", "but", dan lain-lain, sehingga menghasilkan data "StopWords"[14].

	teks_komentar	Cleaning	Case Folding	Normalisasi	Tokenize	Stopword Removal
0	suamiku bosk ke Kamboja? seriusan bgtlu	suamiku bosk ke Kamboja? seriusan bgtlu	suamiku bosk ke kamboja? seriusan bgtlu	suamiku bosk ke kamboja? seriusan bgtlu	[suamiku, bosk, ke, kamboja?, seriusan, bgtlu]	[suamiku, bosk, ke, kamboja?, seriusan, bgtlu]
1	Yang penting niat kerja, ga semua lentang kambo...	Yang penting niat kerja, ga semua lentang kambo...	yang penting niat kerja, ga semua lentang kambo...	yang penting niat kerja, tidak semua lentang k...	[yang, penting, niat, kerja, tidak, semua, le...	Yang penting niat kerja, tidak semua lentang k...
2	foto yang bagus bang, buat cover yasin	foto yang bagus bang buat cover yasin	foto yang bagus bang, buat cover yasin	foto yang bagus bang, buat cover yasin	[foto, yang, bagus, bang, buat, cover, yasin]	foto yang bagus bang, buat cover yasin
3	lahh barusan nonton hiroada tentang kamboja	lahh barusan nonton hiroada tentang kamboja	lahh barusan nonton hiroada tentang kamboja	lah barusan menonton hiroada tentang kamboja	[lah, barusan, menonton, hiroada, tentang, ka...	lah barusan menonton hiroada tentang kamboja
4	antek e	antek e	antek e	antek e	[antek, e]	antek e
5	negara secekl Kamboja saja kita tak berani	negara secekl Kamboja saja kita tak berani	negara secekl kamboja saja kita tak berani	negara secekl kamboja saja kita tak berani	[negara, secekl, kamboja, saja, kita, tak, be...	negara secekl kamboja saja kita tak berani
6	komenya gaada yg dibales berarti	komenya gaada yg dibales berarti	komenya gaada yg dibales berarti	komenya yang dibales berarti	[komenya, yang, dibales, berarti]	komenya yang dibales berarti
7	Padahal ini udh dibandara mau berangkat ke kam...	Padahal ini udh dibandara mau berangkat ke kam...	padahal ini udh dibandara mau berangkat ke kam...	padahal ini sudah dibandara mau berangkat ke k...	[padahal, ini, sudah, dibandara, mau, berangkat...	padahal ini sudah dibandara mau berangkat ke k...
8	kalian ada denger suara nangis gash wol?	kalian ada denger suara nangis gash wol?	kalian ada denger suara nangis gash wol?	kalian ada denger suara menangis wol?	[kalian, ada, denger, suara, menangis, wol?]	kalian ada denger suara menangis wol?
9	ko kik nya malah ketawa sih	ko kik nya malah ketawa sih	ko kik nya malah ketawa sih	kok kakak ya malah ketawa sih	[kok, kakak, ya, malah, ketawa, sih]	kok kakak ya malah ketawa sih
10	Jadi pengen ke Kamboja	Jadi pengen ke Kamboja	jadi pengen ke kamboja	jadi pengen ke kamboja	[jadi, pengen, ke, kamboja]	jadi pengen ke kamboja

Gambar 9 Hasil *Stopword Removal* Data.

Gambar 9 mengilustrasikan hasil penyuntingan dengan menyoroti dampak penghapusan kata-kata yang tidak penting.

### 3.2.7 Stemming

Proses *Stemming* data bertujuan untuk mereduksi kata ke bentuk dasarnya dengan menghilangkan awalan dan akhiran. Contohnya, "mendukung" dan "dukungan" keduanya diubah menjadi "dukung". Jadi, yang dihapus dari kata "mendukung" adalah awalan "men", dan dari kata "dukungan" adalah akhiran "an".



	teks_komentar	Cleaning	Case Folding	Normalisasi	Tokenize	Stopword Removal	Stemming
0	suamiku besok ke Kamboja? seriusan bghu	suamiku besok ke Kamboja seriusan bghu	suamiku besok ke kamboja? seriusan bghu	suamiku besok ke kamboja? seriusan bghu	[suamiku, besok, ke, kamboja?, seriusan, bghu]	[suamiku, besok, kamboja?, seriusan]	suami besok kamboja serius
1	'Yang penting niat kerja, ga semua berangkat kamo...	'Yang penting niat kerja, ga semua berangkat kamo...	yang penting niat kerja, ga semua berangkat kamo...	yang penting niat kerja, tidak semua berangkat k...	[yang, penting, niat, kerja, tidak, semua, te...	[niat, kerja, kamboja, media, bismillah, h7...	niat kerja kamboja media bismillah h7 kamboja
2	foto yang bagus bang, buai cover yasin	foto yang bagus bang, buai cover yasin	foto yang bagus bang, buai cover yasin	foto yang bagus bang, buai cover yasin	[foto, yang, bagus, bang, buai, cover, yasin]	[foto, bagus, bang, cover, yasin]	foto bagus bang cover yasin
3	lehh barusan nonton hirotada berangkat kamboja	lehh barusan nonton hirotada berangkat kamboja	lehh barusan nonton hirotada berangkat kamboja	lah barusan menonton hirotada berangkat kamboja	[ah, barusan, menonton, hirotada, berangkat, ka...	[barusan, menonton, hirotada, kamboja]	barusan tonton hirotada kamboja
4	antek e	antek e	antek e	antek e	[antek, e]	[antek, e]	antek e
5	negara sekeci Kamboja saja kita tak berani	negara sekeci Kamboja saja kita tak berani	negara sekeci kamboja saja kita tak berani	negara sekeci kamboja saja kita tak berani	[negara, sekeci, kamboja, saja, kita, tak, be...	[negara, kamboja, berani]	negara kamboja berani
6	komennya gaada yg dibales berarti	komennya gaada yg dibales berarti	komennya gaada yg dibales berarti	komennya yang dibales berarti	[komennya, yang, dibales, berarti]	[komennya, dibales]	komennya balas
7	Patahal ini udh dibandara mau berangkat ke kam...	Patahal ini udh dibandara mau berangkat ke kam...	padahal ini udh dibandara mau berangkat ke kam...	padahal ini sudah dibandara mau berangkat ke k...	[padahal, ini, sudah, dibandara, mau, berangkat...	[dibandara, berangkat, kamboja]	bandara berangkat kamboja
8	kalian ada denger suara menang woi?	kalian ada denger suara menang woi?	kalian ada denger suara menang woi?	kalian ada denger suara menang woi?	[kalian, ada, denger, suara, menang, woi?]	[denger, suara, menang, woi?]	dengar suara menang woi
9	ko k nya malah ketawa sih	ko k nya malah ketawa sih	ko k nya malah ketawa sih	kok kakak ya malah ketawa sih	[kok, kakak, ya, malah, ketawa, sih]	[kakak, ya, ketawa, sih]	kakak ya ketawa sih
10	Jadi pengen ke Kamboja	Jadi pengen ke Kamboja	jadi pengen ke kamboja	jadi pengen ke kamboja	[jadi, pengen, ke, kamboja]	[pengen, kamboja]	pengin kamboja

Gambar 10 Hasil *Stemming* Data.

Gambar 10 menunjukkan bagaimana kata-kata dengan akar kata yang sama dapat dikenali.

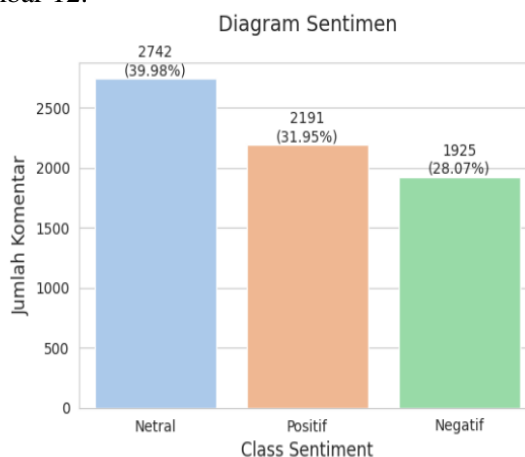
### 3.3 Pelabelan Data

Setelah tahapan *pre-processing*, dataset akan dilakukan pelabelan data yang nantinya akan digunakan sebagai data latih. Proses ini akan mengelompokkan data berdasarkan sentimen positif, netral, dan negatif. Dalam proses pelabelan data ini akan dilakukan pelabelan data yang akan digunakan yaitu sebanyak 6960 data set.

	Stemming	Score	Sentiment
0	suami besok kamboja serius	-1	Negatif
1	niat kerja kamboja media bismillah h7 kamboja	2	Positif
2	foto bagus bang cover yasin	2	Positif
3	barusan tonton hirotada kamboja	-1	Negatif
4	antek e	-1	Negatif
5	negara kamboja berani	0	Netral
6	komennya balas	0	Netral
7	bandara berangkat kamboja	-1	Negatif
8	dengar suara menang woi	1	Positif
9	kakak ya ketawa sih	2	Positif
10	pengin kamboja	0	Netral

Gambar 11 Hasil Pelabelan Data.

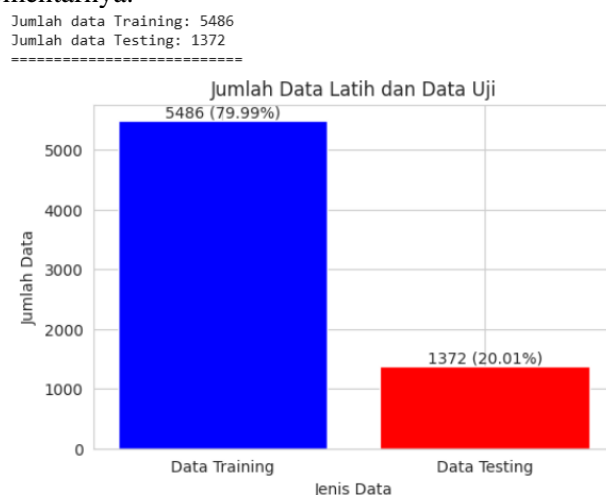
Gambar 11 memperlihatkan hasil pelabelan data, yang kemudian divisualisasikan dalam bentuk diagram pada Gambar 12.

Gambar 12 *Visualisasi* Hasil Pelabelan Data.

Gambar 12 menunjukkan bahwa dari 6960 data komentar yang ada, didapatkan bahwa sebanyak 2191 komentar atau sekitar 31,95% bersentimen positif, 2742 komentar atau sekitar 39,98% bersentimen netral, dan 1925 komentar atau sekitar 28,07% bersentimen negatif. Hasil ini menunjukkan bahwa sebagian besar pengguna Tiktok lebih mengekspresikan pandangan netral mereka terhadap kasus TPPO di Kamboja.

### 3.4 Implementasi Naive Bayes Classifier

Untuk mengklasifikasikan sentimen, Naive Bayes menggunakan pendekatan berbasis probabilitas. Ia mencoba menebak frekuensi kemunculan kata-kata spesifik dalam setiap kategori sentimen: positif, netral, dan negatif. Tebakan ini, yang berupa probabilitas, kemudian dimanfaatkan untuk menentukan apakah sebuah komentar cenderung positif, netral, atau negatif. Dataset pada penelitian ini diperoleh dari komentar pengguna TikTok tentang kasus TPPO di Kamboja yang telah melalui proses pembersihan dan transformasi data. Dalam penelitian ini, dataset dibagi menjadi dua bagian, yaitu 80% untuk data pelatihan dan 20% untuk data pengujian. Gambar 13 menunjukkan hasil prediksi data pengujian menggunakan metode *Naive Bayes Classifier* untuk teks komentarnya.



Gambar 13 Visualisasi Jumlah Data.

Gambar 13 menunjukkan bahwa jumlah data *training* (data latih) adalah sebanyak 5486 atau sekitar 79,99% dan jumlah data *testing* (data uji) adalah sebanyak 1372 atau sekitar 20,01%.

### 3.5 Evaluasi Hasil

#### 3.5.1 Confusion Matrix

*Confusion matrix* metrik evaluasi yang digunakan untuk menunjukkan jumlah prediksi yang benar dan salah dari model klasifikasi saat memproses data yang telah diberi label. Gambar 14 menampilkan hasil *confusion matrix* dengan pembagian data 80:20 (80% data yang dilatih dan diuji pada 20% sisanya).

**Naive Bayes Confusion Matrix**

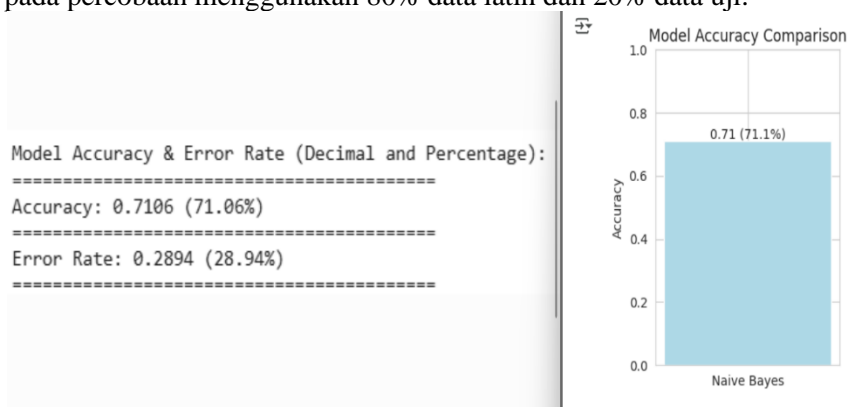
Actual	Negatif	294	73	18
	Netral	92	346	111
	Positif	24	79	335
		Negatif	Netral	Positif
		Predicted		

Gambar 14 Visualisasi Confusion Matrix Rasio 80:20

Gambar 14 menunjukkan bahwa hasil *confusion matrix*-nya adalah *True* Positif 335, *True* Netral 346, dan *True* Negatif 294.

### 3.5.2 Accuracy dan Error

Akurasi adalah ukuran seberapa sering model membuat prediksi yang tepat, biasanya dihitung dengan membagi jumlah prediksi yang benar dengan total prediksi yang dibuat. *Error* adalah kebalikan dari akurasi, yaitu jumlah prediksi yang salah dari semua prediksi yang dilakukan model. Biasanya dihitung dengan membandingkan jumlah prediksi yang salah dengan total prediksi. Kedua hal ini membantu kita memahami seberapa baik kinerja model dalam mengklasifikasi data menggunakan metode *Naive Bayes*. Gambar 15 menunjukkan hasil akurasi dan *error* pada percobaan menggunakan 80% data latih dan 20% data uji.



Gambar 15 Hasil Accuracy &amp; Error.

Gambar 15 menunjukkan bahwa hasil akurasi menggunakan model *Naive Bayes* mencapai 71,06% dan nilai *error*-nya adalah 28,94%. Jika dibandingkan dengan penelitian serupa yang juga menggunakan algoritma *Naive Bayes* pada Analisis Sentimen Tanggapan Masyarakat Terhadap Cyberbullying di Media Sosial Menggunakan Algoritma *Naive Bayes* yang memperoleh akurasi sebesar 77,29% [15]. Sehingga Akurasi model Anda (71,06%) sedikit lebih rendah dibandingkan penelitian serupa (77,29%). Perbedaan sekitar 6,23% menunjukkan bahwa performa model penelitian ini berada di level sedang.

### 3.5.3 Classification Report

*Classification Report* adalah alat yang menunjukkan seberapa baik sebuah model dalam mengklasifikasi data. Di dalamnya, berisi metrik seperti *precision* (ketepatan), *recall* (ingat), *F1-score*, dan *support* atau jumlah data pada setiap sentimen positif, netral dan negatif. Dan dari hasil *classification report* ini, kita bisa melihat performa model *Naive Bayes* secara detail dan

menentukan kelemahan atau kekuatan dalam klasifikasi. Gambar 16 menunjukkan hasil dari *Classification Report* model *Naive Bayes*.

### Classification Report for Naive Bayes:

	precision	recall	f1-score	support
Negatif	0.717	0.764	0.740	385.000
Netral	0.695	0.630	0.661	549.000
Positif	0.722	0.765	0.743	438.000
accuracy	0.711	0.711	0.711	0.711
macro avg	0.711	0.720	0.714	1372.000
weighted avg	0.710	0.711	0.709	1372.000

Gambar 16 Hasil *Classification Report*.

Gambar 16 menunjukkan bahwa hasil classification report dari model *Naive Bayes* berhasil mengidentifikasi sentimen positif dengan baik. Dari semua komentar yang diprediksi positif oleh model, nilai *presisi* adalah 72,02% benar-benar positif, nilai *recall*-nya mengidentifikasi 76,05% dan untuk ukuran keseimbangan antara *presisi* dan *recall* untuk sentimen positif adalah 74,03%. Namun performanya menurun pada sentimen netral, karena dari semua komentar yang diprediksi netral oleh model, nilai *presisi* adalah 69,05% benar-benar positif, nilai *recall*-nya mengidentifikasi 63% dan untuk ukuran keseimbangan antara *presisi* dan *recall* untuk sentimen positif adalah 66,01%. Dan untuk sentimen negatif, dari semua komentar yang diprediksi netral oleh model, nilai *presisi* adalah 71,07% benar-benar positif, nilai *recall*-nya mengidentifikasi 76,04% dan untuk ukuran keseimbangan antara *presisi* dan *recall* untuk sentimen positif adalah 74%.

### 3.5.4 Wordcloud

*Wordcloud* adalah tampilan kata-kata hasil dari analisis sentimen positif, netral dan negatif akan di visualisasikan melalui gambar. Gambar 17 adalah tampilan *Wordcloud* sentimen positif, Gambar 18 adalah tampilan *Wordcloud* sentimen netral dan Gambar 19 adalah tampilan *Wordcloud* sentimen negatif.



Gambar 17 Tampilan *Wordcloud* untuk Sentimen Positif.





614

serta perpanjangan durasi pengumpulan data, diharapkan dapat meningkatkan tingkat akurasi sekaligus merefleksikan tren sentimen selama periode waktu tertentu.

#### DAFTAR PUSTAKA

- [1] F. N. A. Kurniawati, "Meninjau Permasalahan Rendahnya Kualitas Pendidikan Di Indonesia Dan Solusi," *Acad. Educ. J.*, vol. 13, no. 1, pp. 1–13, 2022, doi: 10.47200/aoej.v13i1.765.
- [2] N. Valerisella, N. Z. Sliviani, and A. Situmeang, "Cooperation Frameworks of Indonesia and Cambodia Against Transnational Organized Crime in Online Gambling Human Trafficking," *Yust. Tirtayasa J. Tugas Akhir*, vol. 5, no. 1, p. 1, Mar. 2025, doi: 10.51825/yta.v5i1.29901.
- [3] F. Tarisa Sabrina and H. Prasetyo, "Media Hukum Indonesia (MHI) Published by Yayasan Daarul Huda Krueng Mane Upaya Perlindungan Hukum dan Pemulangan Dalam Kasus Tindak Pidana Perdagangan Orang di Myanmar (Legal Protection And Return Efforts of Human Trafficking People In Myanmar Cases)," vol. 2, no. 3, p. 494, doi: 10.5281/zenodo.12547373.
- [4] T. Saffanah, Rasito, D. I. A. Putra, and Yuliatin, "Strengthening Law Enforcement and Public Awareness to Combat Human Trafficking in Indonesia: Challenges and Solutions," *Al-Ahkam J. Ilmu Syari'ah dan Huk.*, vol. 9, no. 2, pp. 91–112, 2024, doi: 10.22515/alahkam.v9i2.9686.
- [5] A. Septiani *et al.*, "ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA NAÏVE BAYES TERHADAP REGULASI TIKTOK SHOP PADA MEDIA SOSIAL X (TWITTER)," 2024.
- [6] R. L. Atimi and Enda Esyudha Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *J. Sains dan Inform.*, vol. 8, no. 1, pp. 88–96, 2022, doi: 10.34128/jsi.v8i1.419.
- [7] I. Carolin, G. D. Victoria, S. Dina, and M. Nastain, "Pengaruh Penggunaan New Media Tiktok Terhadap Pembentukan Konsep Diri Generasi Muda Indonesia 2022," *JKOMDIS J. Ilmu Komun. Dan Media Sos.*, vol. 3, no. 1, pp. 35–40, Jan. 2023, doi: 10.47233/jkomdis.v3i1.509.
- [8] N. Apriliani, N. Suarna, and W. Prihartono, "ANALISIS SENTIMEN REVIEW PENGGUNAAN TIKTOK MELALUI PENDEKATAN ALGORITMA NAÏVE BAYES," 2023.
- [9] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 452–458, 2023, doi: 10.36040/jati.v7i1.6303.
- [10] M. Kholilullah and U. Hayati, "ANALISIS SENTIMEN PENGGUNA TWITTER(X) TENTANG PIALA DUNIA USIA 17 MENGGUNAKAN METODE NAIVE BAYES," 2024. [Online]. Available: <https://apify.com/quacker/twitter-scraper>
- [11] F. S. Pattuha and H. Hendry, "Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 506, 2022, doi: 10.30865/jurikom.v9i2.4016.
- [12] D. A. Parapat and A. A. Azhar, "Gudang Jurnal Multidisiplin Ilmu Pemanfaatan Media Sosial Tiktok Sebagai Sumber Informasi Berita Bagi Mahasiswa FIS UIN Sumatera Utara Medan," vol. 2, pp. 46–56, 2024.
- [13] F. N. Hidayat and S. Sugiyono, "Analisis Sentimen Masyarakat Terhadap Perekrutan Pppk Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *J. Sains dan Teknol.*, vol. 5, no. 2, pp. 665–672, 2023, doi: 10.55338/saintek.v5i2.1359.
- [14] B. Hakim, "Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 2, pp. 16–

- 22, 2021, doi: 10.30813/jbase.v4i2.3000.
- [15] H. S. Rifai, S. Febrianti, and I. Santoso, “Analisis Sentimen Tanggapan Masyarakat Terhadap Cyberbullying Di Media Sosial Menggunakan Algoritma Naïve Bayes ( Nb ),” *J. IKRAITH-INFORMATIKA*, vol. 7, no. 2, pp. 183–196, 2023.