

# Analisis Sentimen terhadap Fenomena Artis Berjualan di Tiktok Shop pada Media Sosial X Menggunakan Model Naïve Bayes

Desak Made Dwi Utami Putra<sup>1</sup>, I Wayan Dharma Suryawan<sup>2</sup>, Sri Andayani<sup>3</sup>, Erni Yunia Rusadi<sup>4</sup>

<sup>1,2,4</sup>Program Studi Informatika, Institut Bisnis dan Teknologi Indonesia

<sup>3</sup>Program Studi Sistem Informasi, Universitas Katolik Misi Charitas

Jalan Tukad Pakerisan No. 97, Panjer Denpasar

Email: [desak.utami@instiki.ac.id](mailto:desak.utami@instiki.ac.id)<sup>1</sup>, [wayan.dharma@instiki.ac.id](mailto:wayan.dharma@instiki.ac.id)<sup>2</sup>, [andayani\\_s@ukmc.ac.id](mailto:andayani_s@ukmc.ac.id)<sup>3</sup>

**Abstrak** - Aplikasi TikTok sangat digemari oleh masyarakat sebagai sarana hiburan dan sumber pendapatan. Salah satu fitur keunggulan TikTok Shop, memberikan peluang bagi para pelaku usaha untuk mempromosikan produk mereka. Namun, beragam permasalahan muncul akibat fitur ini, karena masyarakat harus bersaing dengan artis. Untuk memahami pandangan masyarakat terhadap fenomena ini, dilakukan analisis sentimen dengan penarikan data menggunakan media sosial X (Twitter) untuk mengumpulkan *tweet* artis yang berjualan di TikTok Shop. Data awal yang terkumpul berjumlah 1.189 *tweet*, yang kemudian diproses dengan bantuan *tools Google Colab* melalui tahapan *pre-processing text* menjadi 770 data (*tweet*) bersih, kemudian melakukan pelabelan data, pemisahan data, dan penerapan metode *Naïve Bayes Classifier* menggunakan bahasa pemrograman *Python*. Hasil penelitian menggunakan pelabelan ahli bahasa lebih unggul dan akurat dari pelabelan *TextBlob* (mesin) dengan total data 494 *tweet* netral, 253 *tweet* positif, dan 23 *tweet* negatif dengan akurasi tertinggi dari *Multinomial Naïve Bayes* sebesar 88.31%. Berdasarkan hasil analisis sentimen, disimpulkan bahwa masyarakat tidak terpengaruh oleh fenomena artis berjualan di TikTok Shop, yang ditunjukkan oleh jumlah *tweet* netral yang lebih tinggi dibandingkan dengan *tweet* positif dan *tweet* negatif.

**Kata Kunci**—Analisis Sentimen, Tiktok, X, Naïve Bayes Classifier

**Abstract**— *TikTok is a popular application used for entertainment and income generation. One of its key features, TikTok Shop, offers opportunities for entrepreneurs to promote their products. However, issues arise as the public must compete with celebrities who also sell on the platform. To understand public responses, a sentiment analysis was conducted on tweets from the social media platform X (Twitter) discussing celebrities selling on TikTok Shop. From 1,189 collected tweets, text pre-processing was performed, resulting in 770 clean tweets. The data was then labeled, split, and analyzed using the Naïve Bayes Classifier method in Google Colab with Python programming. Manual labeling by language experts showed higher accuracy compared to automatic labeling using TextBlob. The results showed 494 neutral tweets, 253 positive tweets, and 23 negative tweets, with the highest accuracy of 88.31% achieved by the Multinomial Naïve Bayes method. In conclusion, the majority of the public holds a neutral stance toward the phenomenon of celebrities selling on TikTok Shop.*

**Key Word**— *Sentiment Analysis, Tiktok, X, Naïve Bayes Classifier*

## I. PENDAHULUAN

TikTok, yang diluncurkan oleh ByteDance pada 29 September 2016, telah berkembang menjadi platform video pendek populer dengan fitur unggul seperti TikTok Shop yang memungkinkan transaksi langsung dalam aplikasi [1]. Fitur ini membuka peluang besar bagi pelaku usaha untuk memasarkan produk mereka secara kreatif. Berdasarkan data dari Business of Apps, jumlah pengguna bulanan TikTok mencapai 1,67 miliar pada tahun 2023, menunjukkan peningkatan popularitas secara global [2]. Namun, muncul kontroversi ketika para selebritas turut memanfaatkan TikTok Shop untuk berjualan, memicu persaingan yang tidak seimbang dengan pedagang kecil. Banyak pengguna media sosial, khususnya di X (Twitter), menyoroti ketimpangan ini, karena artis dianggap menggunakan popularitas mereka untuk menarik perhatian dan menawarkan harga murah serta diskon besar [3].

Fenomena ini memicu perdebatan publik mengenai keadilan dalam berbisnis di platform digital. Beberapa masyarakat menyuarakan keberatan atas dominasi artis dalam TikTok Shop yang dianggap merugikan pedagang kecil, sementara sebagian lainnya mendukung para selebritas sebagai bentuk usaha memanfaatkan ketenaran untuk menambah penghasilan [4]. Untuk memahami respons masyarakat terhadap isu ini, dilakukan analisis sentimen terhadap *tweet* di X menggunakan metode Naïve Bayes Classifier. Metode ini dipilih karena memiliki akurasi tinggi, struktur sederhana, dan efisien dalam menganalisis opini publik apakah cenderung positif, negatif, atau netral terhadap partisipasi artis dalam TikTok Shop [5].

## II. LANDASAN TEORI

### A. Aplikasi TikTok

TikTok merupakan aplikasi platform video pendek populer yang sedang digandrungi oleh semua kalangan. Penggunaanya dapat mencari video, membuat konten bahkan mengobrol dengan pengguna lain melalui fitur live yang disediakan oleh tiktok. Tujuan utama aplikasi TikTok ini adalah pengguna dapat menikmati berbagai konten yang dapat menghibur dan menyenangkan para penggunanya.

Aplikasi TikTok adalah aplikasi yang memberikan *special effects* unik dan menarik yang dapat digunakan oleh penggunanya dengan mudah sehingga dapat membuat video pendek dengan hasil yang keren serta dapat dipamerkan kepada teman teman atau pengguna lainnya [1]. Aplikasi video pendek TikTok ini memiliki fitur musik yang banyak sehingga penggunanya dapat membuat video dengan berbagai kreatifitas tarian, gaya bebas, dan masih banyak lagi. Sehingga dapat mendorong kreativitas penggunanya menjadi *content creator* [6].

### B. Media Sosial X (Twitter)

Twitter adalah satu dari media sosial yang mempunyai fitur untuk penggunanya mengirim dan menerima informasi dalam bentuk teks dan video [7]. Awalnya, Twitter pada saat itu user hanya bisa mengirim dan membaca informasi berbasis teks sampai 140 karakter lalu di tahun 2017 di upgrade menjadi 280 karakter yang sekarang disebut sebagai 'tweet' [8]. Tweet mengandung berupa opini dari penggunaannya terhadap suatu kejadian yang telah dialami. Pada masa sekarang twitter dapat digunakan untuk mengemukakan suatu pendapat tentang suatu hal seperti kegiatan social, memberikan kabar tentang kondisi lalu lintas ataupun berita cuaca dan bencana alam serta dapat memperingati tentang fenomena yang dihadapi [9].

### C. Algoritma Naïve Bayes

Text mining adalah teknik yang digunakan untuk memenuhi kebutuhan dalam proses klasifikasi dokumen dengan konten apapun. Salah satu teknik mengubah sebuah informasi yang merupakan sikap seseorang atau suatu kejadian dengan membagi polaritas dari sebuah teks adalah analisa sentimen. Pembagian tersebut untuk mengetahui teks tersebut bersifat positif, negatif atau netral [10]. Metode klasifikasi data Naive Bayes memiliki teori kinerja yang sejenis dengan Decision Tree dan Neural Network. Naive Bayes yang dikemukakan oleh ilmuwan Inggris Thomas Bayes untuk merupakan salah satu metode klasifikasi data yang memakai perhitungan probabilitas dan statistic [5]. Algoritma Naive Bayes dipilih karena metode klasifikasi dengan formula sederhana dan mudah diterapkan serta memiliki akurasi yang cukup tinggi dibandingkan dengan metode lain [11]. Naive Bayes juga diyakini merupakan metode untuk melakukan pemisahan data terstruktur yang lebih unggul daripada metode pemisahan data terstruktur lainnya dalam hal akurasi dan komputasi. Pada penelitian oleh Aldisa dan Maulana membandingkan antara algoritma Naive Bayes, Decision Tree, dan SVM menunjukkan bahwa presisi Naive Bayes menempati urutan terbaik.

### D. Penelitian Terdahulu

Penelitian terdahulu menunjukkan bahwa analisis sentimen berbasis media sosial, khususnya X (Twitter), banyak digunakan untuk memahami opini publik terhadap isu-isu aktual. Rizkina dan Hasan (2023) menganalisis komentar netizen terhadap pembubaran konser NCT 127 menggunakan metode Naïve Bayes, dengan hasil akurasi sebesar 82,01%. Sementara itu, Safira dan Hasan (2023) meneliti sentimen masyarakat terkait layanan *paylater*,

dengan hasil menunjukkan mayoritas pengguna bersentimen negatif. Naïve Bayes terbukti memberikan akurasi lebih tinggi (91%) dibandingkan metode TextBlob (61%). Kedua penelitian ini memperlihatkan keandalan algoritma Naïve Bayes dalam mengklasifikasi opini pengguna Twitter terhadap fenomena social [12].

Penelitian lain oleh Syah dan Witanti (2022) mengkaji sentimen masyarakat terhadap vaksinasi COVID-19 dengan algoritma SVM. Dari 14.099 tweet, data disaring menjadi 4.708, dengan hasil mayoritas sentimen positif sebesar 83,6% dan akurasi model mencapai 89%. Ruhyana (2019) juga menggunakan Naïve Bayes untuk menganalisis opini terhadap kebijakan ganjil-genap kendaraan, menghasilkan akurasi 86,67% [13]. Hal ini menegaskan bahwa baik SVM maupun Naïve Bayes dapat digunakan secara efektif dalam memahami tanggapan publik terhadap kebijakan pemerintah melalui *media social* [14].

Adapun penelitian oleh Rizqiah dan Kadafi (2022) memanfaatkan algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) untuk menilai opini terhadap vaksin booster. Hasilnya, Naïve Bayes mencapai akurasi lebih tinggi (90,19%) dibanding KNN (82,89%). Terakhir, penelitian oleh Rusadi (2023) secara langsung relevan dengan penelitian ini, karena menganalisis sentimen terhadap artis yang berjualan di TikTok Shop menggunakan Naïve Bayes. Dengan data yang dikumpulkan secara spesifik, studi ini bertujuan menjadi masukan bagi perumusan kebijakan pemerintah dalam mengatur aktivitas perdagangan daring oleh publik figur. Semua penelitian tersebut menunjukkan efektivitas metode klasifikasi dalam menilai opini publik digital terhadap isu-isu yang berkembang [15].

## III. METODE

### A. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di kampus INSTIKI Indonesia yang terletak di jalan Tukad Pakerisan No. 97 Denpasar Selatan, Bali. Waktu penelitian ini dimulai dari bulan September sampai dengan Desember 2023.

### B. Tahapan Penelitian

Adapun tahapan yang dilakukan pada penelitian ini adalah :

1. Pengumpulan data
2. Pre-Processing Text
  - a) Cleaning
  - b) Case Folding
  - c) Tokenizing
  - d) Normalisasi kata
  - e) Stopword Removal / Filtering
  - f) Stemming
  - g) Pelabelan Data
  - h) Splitting Data
  - i) Perhitungan Klasifikasi Naïve Bayes Classifier
  - j) Evaluasi Model

### C. Pengumpulan Data

Proses pengumpulan data dilakukan pada tanggal 01 Januari 2023 – 31 Desember 2023 dengan kata kunci yang sangat spesifik. Pemilihan kata kunci perlu diperhatikan untuk meminimalisir adanya data yang tidak relevan dengan permasalahan. Jumlah data hasil dari proses crawling adalah sebanyak 1189 data dengan menggunakan 6 kata kunci yaitu Artis Live Tiktok Shop, Artis Jualan Sembako, Artis Jualan Tiktok, Artis Tiktok Shop, Artis UMKM Tiktok, dan Produk Artis Tiktok.

### D. Pre-Processing Text

Setelah mengumpulkan data melalui proses *crawling* pada media sosial X (*Twitter*), data tersebut belum diolah dan masih berupa struktur yang tidak teratur. Terdapat duplikasi data dan perulangan serta terdapat simbol atau kata-kata yang tidak dibutuhkan dalam data tersebut. Pada tahap *pre-processing text* yaitu membersihkan data-data agar lebih terstruktur agar menjadi data yang berkualitas.

#### 1) Cleaning

Proses pre-processing dimulai dengan tahap cleaning, yaitu menghapus elemen-elemen yang tidak diperlukan dalam tweet seperti retweet, mention, hashtag, URL, data duplikat, angka, emoji, dan tanda baca (misalnya titik, koma, dan tanda tanya). Langkah ini penting untuk menghilangkan noise atau informasi yang tidak relevan agar data lebih bersih, akurat, dan konsisten sebelum dianalisis lebih lanjut. Dengan data yang telah dibersihkan, algoritma analisis sentimen dapat bekerja secara lebih optimal dalam mengklasifikasikan opini pengguna.

#### 2) Case Folding

Proses Case Folding adalah proses mengubah huruf besar (A- Z) menjadi huruf kecil (a-z) untuk konsistensi data dan keseragaman. Tujuannya analisis teks tanpa terpengaruh oleh variasi huruf.

#### 3) Tokenizing

Proses tokenizing adalah proses memecah kalimat menjadi kalimat tunggal atau biasa disebut token. Pemisahan kalimat ini menggunakan tanda koma.

#### 4) Normalisasi kata

Normalisasi Kata adalah proses perubahan atau pembenahan kata yang tidak tepat (tidak baku) menjadi kata yang tepat (baku) sesuai Kamus Besar Bahasa Indonesia (KBBI). Tujuan normalisasi untuk memperbaiki istilah bahasa gaul (slang) dalam dataset agar memudahkan dalam proses penelitian. Perbaikan kata pada proses ini menggunakan bantuan kamus kata baku.

#### 5) Stopword Removal / Filtering

Proses stopwords/filtering adalah menghapus kata-kata yang sering muncul dalam teks tetapi tidak memberikan makna penting seperti kata “dan”, “pada”, “atau”, dan lainnya. Tujuan dari proses ini, untuk lebih fokus pada kata-

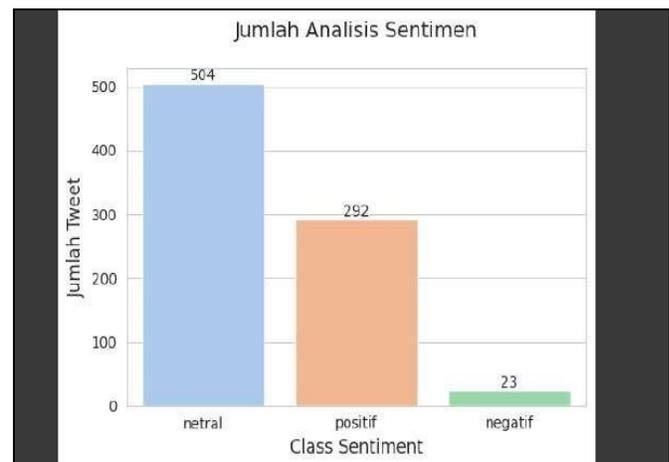
kata yang yang lebih relevan.

#### 6) Stemming

Proses stemming adalah penyederhanaan kata-kata dalam teks dengan mengubah kata-kata menjadi bentuk dasarnya. Contoh dari proses ini yaitu kata “kejahatan” menjadi “jahat”. Tujuan proses ini yaitu mempermudah analisis teks serta pemrosesan bahasa alami sehingga informasi yang didapat lebih efisien.

#### 7) Pelabelan Data

Dalam proses pelabelan data penulis menggunakan ahli bahasa Indonesia dan objek TextBlob. Proses pelabelan data dapat dilihat outputnya pada Gambar 1.



Gambar 1 Bar Plot Total Pelabelan

Jadi, setelah selesai melabelkan *tweet* atau data menggunakan *library TextBlob*, tahap selanjutnya membuat bar plot atau grafik.

#### 8) Splitting Data

Dalam proses penelitian ini, terdapat suatu metode pembagian data yang diterapkan, yaitu dengan rasio 80:20. Artinya 80% dari keseluruhan data digunakan sebagai data latih, sementara 20% sisanya disisihkan sebagai data uji. Penggunaan rasio ini bertujuan untuk memberikan proporsi yang seimbang antara data yang digunakan untuk melatih model (data latih) dan data yang digunakan untuk menguji kinerja model tersebut (data uji).

#### 9) Perhitungan Klasifikasi Naïve Bayes Classifier

Perhitungan klasifikasi Naïve Bayes Classifier menggunakan tiga jenis Naïve Bayes (Gaussian, Multinomial, dan Bernoulli) untuk mendapatkan hasil yang optimal. Dalam proses ini akan melakukan perbandingan Naïve Bayes data labelling TxtBlob dan juga data labelling ahli bahasa, agar mengetahui performa paling baik di antara dua pelabelan tersebut.

#### 10) Evaluasi Model

Setelah perhitungan pada data uji, langkah berikutnya adalah menguji serta mengevaluasi model menggunakan confusion matrix untuk mengukur performa *accuracy*,

precision, recall dan f1-score. Evaluasi model ini mencakup tiga kategori kelas yaitu positif, negatif, dan netral sehingga confusion matrix yang dihasilkan memiliki ordo 3x3. Dalam struktur tabel confusion matrix terdapat data aktual dan prediksi yang akan menghasilkan nilai rata-rata accuracy, precision, recall dan f1-score. Nilai ini memberikan gambaran komprehensif tentang kinerja model kedua kelas tersebut.

#### IV. HASIL DAN PEMBAHASAN

##### A. Pengumpulan Data

Pada penelitian ini, dataset didapatkan melalui proses crawling data Twitter dengan bantuan tools Google Colab menggunakan bahasa pemrograman Python. Proses pengumpulan data dilakukan pada tanggal 01 Januari 2023 – 31 Desember 2023 dengan kata kunci yang sangat spesifik. Kata kunci yang digunakan, di antaranya artis live tiktok shop (76 dataset), artis jualan sembako (175 dataset), artis jualan tiktok (463 dataset), artis tiktok shop (260 dataset), artis umkm tiktok (157 dataset), dan produk artis tiktok (52 dataset). Hasil keseluruhan dari kata kunci yang digunakan yaitu 1189 data.

##### B. Pre-Processing Text

Tahap pre-processing text adalah tahap pembersihan data yang sebelumnya tidak terstruktur menjadi data yang berkualitas. Proses ini akan menampilkan informasi singkat terkait dataset sebelum melewati proses pembersihan atau lainnya. Data awal pada proses ini berjumlah 1189 data. Implementasi pre-processing dapat dilihat pada Gambar 2.

```
[1] import pandas as pd

data = pd.read_csv("DATASET_ERNI_20103277.csv", sep=";")
data.info()

<<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1189 entries, 0 to 1188
Data columns (total 12 columns):
#   Column              Non-Null Count  Dtype
---  ---             
0   created_at          1189 non-null  object
1   id_str              1189 non-null  object
2   full_text           1189 non-null  object
3   quote_count        1189 non-null  object
4   reply_count        1189 non-null  object
5   retweet_count      1189 non-null  object
6   favorite_count     1189 non-null  object
7   lang                1189 non-null  object
8   user_id_str        1189 non-null  object
9   conversation_id_str 1189 non-null  object
10  username            1189 non-null  object
11  tweet_url           1189 non-null  object
dtypes: object(12)
memory usage: 111.6+ KB
```

Gambar 2 Tahap awal Pre-Processing Text

##### 1) Cleaning

Cleaning adalah proses membersihkan dataset dengan menghilangkan karakter yang tidak diperlukan. Seluruh dataset akan mengalami banyak perubahan seperti URL, HTML dan karakter lainnya. Script yang digunakan untuk membersihkan URL dan HTML dapat dilihat pada Gambar 3.

```
import re
import string
import nltk

# Fungsi untuk menghapus URL
def remove_URL(tweet):
    if tweet is not None and isinstance(tweet, str):
        url = re.compile(r'https?://\S+|www\.\S+')
        return url.sub('', tweet)
    else:
        return tweet

# Fungsi untuk menghapus HTML
def remove_html(tweet):
    if tweet is not None and isinstance(tweet, str):
        html = re.compile(r'<.*>')
        return html.sub('', tweet)
    else:
        return tweet
```

Gambar 3 Pembersihan URL dan HTML

##### 2) Case Folding

Case Folding adalah proses perubahan huruf kapital menjadi huruf kecil pada seluruh dataset. Proses ini menyamaratakan seluruh dataset dengan susunan huruf yang sama dan menghapus karakter lain karena dianggap sebagai pembatas. Implementasi case folding dapat dilihat pada Gambar 4.

```
[11] def case_folding(text):
    lowercase_text = text.lower()
    return lowercase_text

def case_folding(tweet):
    if tweet is not None and isinstance(tweet, str):
        return case_folding(tweet)
    else:
        return tweet
```

Gambar 4 Proses Case Folding

##### 3) Tokenizing

Tokenizing adalah proses pemecahan kalimat menjadi kata tunggal atau biasa disebut dengan token. Pemisahannya menggunakan tanda koma. Tokenizing dapat dilihat pada Gambar 5.

```
[1] def tokenize(text):
    tokens = text.split()
    return tokens

def tokenize(tweet):
    if tweet is not None and isinstance(tweet, str):
        return tokenize(tweet)
    else:
        return tweet
```

Gambar 5 Proses Tokenizing

##### 4) Normalisasi Kata

Normalisasi kata adalah perubahan kata-kata tidak baku menjadi kata baku. Pada proses ini menggunakan bantuan kamus katabaku agar data atau kata mudah dipahami. Kamus kata baku akan mengubah kata-kata yang salah menjadi kata baku yang sesuai dengan ejaan Bahasa Indonesia. Script normalisasi kata dapat dilihat pada Gambar 6.

```

import pandas as pd

# Fungsi penggantian kata tidak baku
def replace_taboo_words(text, kamus_tidak_baku):
    if isinstance(text, str):
        words = text.split()
        replaced_words = []
        kalimat_baku = []
        kata_diganti = []
        kata_tidak_baku_hash = []

        for word in words:
            if word in kamus_tidak_baku:
                baku_word = kamus_tidak_baku[word]
                if isinstance(baku_word, str) and all(char.isalpha() for char in baku_word):
                    replaced_words.append(baku_word)
                    kalimat_baku.append(baku_word)
                    kata_diganti.append(word)
                    kata_tidak_baku_hash.append(hash(word))
                # else:
                #     replaced_words.append('karakter eksplisit')
                #     kalimat_baku.append(word)
            else:
                replaced_words.append(word)
        replaced_text = ' '.join(replaced_words)
    else:
        replaced_text = ''
        kalimat_baku = []
        kata_diganti = []
        kata_tidak_baku_hash = []

    return replaced_text, kalimat_baku, kata_diganti, kata_tidak_baku_hash
    
```

Gambar 6 Penggantian Kata Tidak Baku

5) Stopword Removal/ Filtering

Stopword removal/filtering adalah menghapus kata-kata yang tidak memiliki makna penting dalam dataset. Kata-kata tersebut seperti “dan”, “atau”, “yang”. *DataFrame* *stopword removal/filtering* dapat dilihat pada Gambar 7.

```

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer

# Data Data (menyaring kata-kata yang tidak penting)
data_ahli_bahasa = data_ahli_bahasa.dropna()

# Memasukkan data ke dalam DataFrame
df = pd.DataFrame(data_ahli_bahasa)

# Menampilkan DataFrame
df
    
```

Gambar 7 Proses Stopword Removal/Filtering

6) Stemming

Stemming adalah penyederhanaan kata-kata ke dalam bentuk dasarnya. Penyederhanaan kata tersebut seperti kekayaan menjadi kata kaya. *Stemming* dapat dilihat pada Gambar 8.

```

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer

# Data Data (menyaring kata-kata yang tidak penting)
data_ahli_bahasa = data_ahli_bahasa.dropna()

# Memasukkan data ke dalam DataFrame
df = pd.DataFrame(data_ahli_bahasa)

# Menampilkan DataFrame
df
    
```

Gambar 8 Proses Stemming

7) Pelabelan Data

Pelabelan data dalam penelitian ini menggunakan dua jenis pelabelan yaitu pelabelan *TextBlob* (mesin) dan ahli bahasa (manusia). Tujuan dua pelabelan ini untuk

membandingkan akurasi kedua pelabelan tersebut dalam menentukan keakuratan yang lebih tinggi.

8) Splitting Data

Tujuan dua pelabelan ini untuk membandingkan akurasi kedua pelabelan tersebut dalam menentukan keakuratan yang lebih tinggi.

a) Splitting Data Ahli Bahasa

Langkah awal memulai proses *splitting data* ahli bahasa dengan memasukkan informasi hasil labelling data ahli bahasa. Seluruh informasi seperti tanggal, waktu, username, dan jumlah data tersaji. Informasi data ahli bahasa dapat dilihat pada Gambar 9.

```

import pandas as pd

data_ahli_bahasa = pd.read_csv("Hasil_Labelling_Data_Ahli_Bahasa_Erni1.csv")
data_ahli_bahasa.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 778 entries, 0 to 769
Data columns (total 11 columns):
 # Column Non-Null Count Dtype
---
 0 tanggal 778 non-null object
 1 waktu 778 non-null object
 2 username 778 non-null object
 3 quote_count 778 non-null int64
 4 reply_count 778 non-null int64
 5 retweet_count 778 non-null int64
 6 favorite_count 778 non-null int64
 7 stemming_data 778 non-null object
 8 sentiment_polarity 778 non-null object
 9 sentiment_textblob 778 non-null object
 10 sentiment_ahli_bahasa 778 non-null object
dtypes: int64(4), object(7)
memory usage: 66.3+ KB

[45] data_ahli_bahasa_baru = data_ahli_bahasa.dropna()

data_ahli_bahasa_baru.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 778 entries, 0 to 769
Data columns (total 11 columns):
 # Column Non-Null Count Dtype
---
 0 tanggal 778 non-null object
 1 waktu 778 non-null object
 2 username 778 non-null object
 3 quote_count 778 non-null int64
 4 reply_count 778 non-null int64
 5 retweet_count 778 non-null int64
 6 favorite_count 778 non-null int64
 7 stemming_data 778 non-null object
 8 sentiment_polarity 778 non-null object
 9 sentiment_textblob 778 non-null object
 10 sentiment_ahli_bahasa 778 non-null object
dtypes: int64(4), object(7)
memory usage: 66.3+ KB
    
```

Gambar 9 Informasi Data Ahli Bahasa

b) Splitting Data TextBlob

Langkah awal memulai proses *splitting data TextBlob* dengan memasukkan informasi hasil labelling data *Textblob*. Seluruh informasi seperti tanggal, waktu, username, dan jumlah data tersaji. Informasi data *TextBlob* dapat dilihat pada Gambar 10.

```

import pandas as pd

data_textblob = pd.read_csv("Hasil_Labelling_Data_Textblob_Erni1.csv")
data_textblob.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 778 entries, 0 to 769
Data columns (total 11 columns):
 # Column Non-Null Count Dtype
---
 0 tanggal 778 non-null object
 1 waktu 778 non-null object
 2 username 778 non-null object
 3 quote_count 778 non-null int64
 4 reply_count 778 non-null int64
 5 retweet_count 778 non-null int64
 6 favorite_count 778 non-null int64
 7 stemming_data 778 non-null object
 8 sentiment_polarity 778 non-null object
 9 sentiment_textblob 778 non-null object
 10 sentiment_ahli_bahasa 778 non-null object
dtypes: int64(4), object(7)
memory usage: 66.3+ KB

[79] data_textblob_baru = data_textblob.dropna()

data_textblob_baru.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 778 entries, 0 to 769
Data columns (total 11 columns):
 # Column Non-Null Count Dtype
---
 0 tanggal 778 non-null object
 1 waktu 778 non-null object
 2 username 778 non-null object
 3 quote_count 778 non-null int64
 4 reply_count 778 non-null int64
 5 retweet_count 778 non-null int64
 6 favorite_count 778 non-null int64
 7 stemming_data 778 non-null object
 8 sentiment_polarity 778 non-null object
 9 sentiment_textblob 778 non-null object
 10 sentiment_ahli_bahasa 778 non-null object
dtypes: int64(4), object(7)
memory usage: 66.3+ KB
    
```

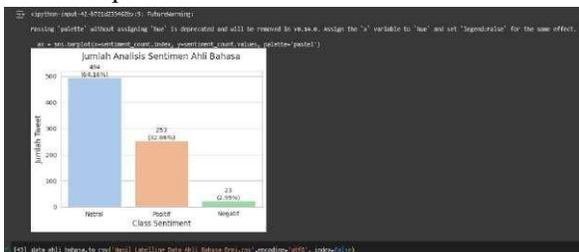
Gambar 10 Informasi Data TextBlob

9) *Perhitungan Klasifikasi Naïve Bayes Classifier*

Hasil tampilan bar plot sentimen ahli bahasa menunjukkan bahwa sentimen netral memiliki 494 data (64.16 %), positif memiliki 253 data (32.86%), dan negatif memiliki 23 data (2.99%). Tampilan bar plot ini memberikan visualisasi sentimen dalam dataset. Tampilan lengkap dari bar plot *labelling* ahli bahasa dapat dilihat pada Gambar

10) *Evaluasi Model*

Evaluasi model data ahli bahasa melibatkan pengujian akurasi prediksi sentimen menggunakan klasifikasi *Naïve Bayes (Gaussian, Multinomial, Bernoulli)*. Langkah ini membandingkan kemampuan tiga model dalam mengklasifikasikan sentimen berdasarkan dataset ahli bahasa. Proses awal evaluasi model dataset ahli bahasa dapat dilihat pada Gambar 11.



Gambar 11 Tampilan Bar Plot Analisis Sentimen

V. KESIMPULAN

Berdasarkan analisis sentimen terhadap opini masyarakat di media sosial Twitter mengenai fenomena artis yang berjualan di TikTok Shop menggunakan model-model *Naïve Bayes Classifier*, model *Multinomial Naive Bayes* yang dilakukan oleh ahli bahasa menunjukkan performa paling unggul dan akurat dengan akurasi 88.31% daripada pelabelan mesin (*TextBlob*) dengan akurasi tertinggi 86.36%.

DAFTAR PUSTAKA

[1] D. Pebrimireni and D. S. Fauziya, “Analisis Penggunaan Aplikasi Tiktok Sebagai Media Pembelajaran Bahasa Indonesia Dalam Upaya Peningkatan Motivasi Belajar Siswa,” *J. Bima Pus. Publ. Ilmu Pendidik. Bhs. dan Sastra*, vol. 2, no. 3, pp. 169–178, 2024, [Online]. Available: <https://journal.aripi.or.id/index.php/Bima/article/view/1040>

[2] P. S. Rosiana, A. R. Nurhidayat, A. A. Mohsa, and A. A. Ridha, “Komputer Menggunakan Evaluasi Heuristic,” *Anal. Apl. Tiktok Berdasarkan Prinsip Dan Paradig. Interak. Mns. Dan Komput. Menggunakan Eval. Heuristic*, vol. 11, no. 3, 2023.

[3] R. Ferira, “Analisis Aplikasi Tiktok Sebagai Platform,” *J. Dakwah dan Komun.*, no. June, 2022.

[4] Y. N. Bulele and T. Wibowo, “Analisis Fenomena Sosial Media Dan Kaum Milenial: Studi Kasus Tiktok,” *Conf. Business, Soc. Sci. Innov. Technol.*, vol. 1, pp. 565–572,

2020, [Online]. Available: <http://journal.uib.ac.id/index.php/cbsisit>

[5] S. M. Salsabila, A. Alim Murtopo, and N. Fadhilah, “Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier,” *J. Minfo Polgan*, vol. 11, no. 2, pp. 30–35, 2022, doi: 10.33395/jmp.v11i2.11640.

[6] Y. U. Budiman, M. F. Santoso, T. Pattiasina, and A. J. Wahidin, “Aplikasi Tiktok Sebagai Media Promosi Kelompok Masyarakat Peduli Iklim Bumiku Satu,” *J-ABDI J. Pengabd. Kpd. Masy.*, vol. 2, no. 5, pp. 5085–5092, 2022, doi: 10.53625/jabdi.v2i5.3677.

[7] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.

[8] M. A. Djamaludin, A. Triayudi, and E. Mardiani, “Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan Metode Naïve Bayes Classifier,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 2, pp. 161–166, 2022, doi: 10.35870/jtik.v6i2.398.

[9] F. R. Rakhman, R. W. Ramadhani, and Y. A. Kuncoroyakti, “Analisis Sentimen dan Opini Digital Kampanye 3M di Masa Covid-19 Melalui Media Sosial Twitter,” *Komunikologi J. Ilm. Ilmu Komun.*, vol. 18, no. 1, pp. 8–20, 2021.

[10] F. D. Ananda and Y. Pristyanto, “Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.

[11] A. Wibowo, Firman Noor Hasan, Rika Nurhayati, and Arief Wibowo, “Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier,” *J. Asimetri J. Ilm. Rekayasa Inov.*, vol. 4, pp. 239–248, 2022, doi: 10.35814/asiimetrik.v4i1.3577.

[12] N. Q. Rizkina and F. N. Hasan, “Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes,” *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1136–1144, 2023, doi: 10.47065/josh.v4i4.3803.

[13] T. Walasary, “Survey Paper tentang Analisis Sentimen,” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 201–206, 2022, doi: 10.24002/konstelasi.v2i1.5378.

[14] H. Syah and A. Witanti, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm),” *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 59–67, 2022, doi: 10.47080/simika.v5i1.1411.

[15] A. Septi Rizqiah dan Kadafi, Abdul Rahman, “Analisis Sentimen Pengguna Twitter Terhadap Vaksin Booster Menggunakan Metode Naive Bayes Dan K-Nearest Neighbor,” *J. Sains dan Teknol.*, vol. 4, no. 1, pp. 1–5, 2022.