

Analisis Sentimen Penggemar Treasure Di Karnaval Mandiri Menggunakan Naïve Bayes

Chulyatunni'mah¹, Rudi Kurniawan², Saeful Anwar³

^{1,2,3}Teknik Informatika, STMIK IKMI Cirebon

Email: ¹rainrizky1209@gmail.com, ²rudikurniawan@gmail.com, ^{3,*}saefulanwar419@gmail.com

Email Penulis Korespondensi: rainrizky1209@gmail.com

Abstrak

Fenomena K-pop yang terus berkembang telah menarik perhatian dunia, termasuk di Indonesia, di mana Treasure adalah salah satu boy group yang memiliki banyak penggemar aktif. Dengan menggunakan algoritma Naive Bayes, penelitian ini mengkaji sentimen penggemar terhadap penampilan Treasure di media sosial X pada Karnaval Mandiri 2024. Metode ini dipilih karena mudah digunakan untuk mengklasifikasikan teks. Metode crawling digunakan untuk mendapatkan dataset penelitian, yang menghasilkan 1.018 tweet yang relevan. Proses preprocessing termasuk pembersihan, normalisasi, penghapusan stopwords, tokenisasi, dan stemming. Selanjutnya, pustaka TextBlob digunakan untuk melabelkan tweet ke dalam kategori positif, netral, dan negatif. Hasil analisis menunjukkan bahwa sentimen netral mendominasi (38,1%), diikuti oleh sentimen positif (35,3%) dan negatif (26,6%). Dengan akurasi evaluasi sebesar 96%, model Naive Bayes menunjukkan bahwa mayoritas penggemar bersikap netral terhadap penampilan Treasure. Sementara sentimen positif menunjukkan antusiasme yang besar, sentimen negatif menunjukkan bahwa ekspektasi tidak terpenuhi. Studi ini menunjukkan bahwa algoritma Naive Bayes adalah metode yang berguna untuk menganalisis sentimen di media sosial. Karena itu, manajemen artis dan penyelenggara acara dapat memanfaatkan temuan ini untuk membuat strategi komunikasi yang lebih baik dan meningkatkan pengalaman penggemar. Penelitian ini juga menambah literatur tentang analisis sentimen industri hiburan, terutama yang berkaitan dengan fenomena K-pop di Indonesia.

Kata Kunci: Sentimen Penggemar, Naïve Bayes, Treasure, Karnaval Mandiri, Media Sosial X

Abstract

The growing K-pop phenomenon has attracted worldwide attention, including in Indonesia, where Treasure is one of the boy groups that has many active fans. Using the Naive Bayes algorithm, this research examines fan sentiment towards Treasure's performance on social media X at Carnival Mandiri 2024. This method was chosen because it is easy to use for classifying text. The crawling method was used to obtain the research dataset, which resulted in 1,018 relevant tweets. The preprocessing process included cleaning, normalization, stopwords removal, tokenization, and stemming. Next, the TextBlob library was used to label the tweets into positive, neutral, and negative categories. The analysis results show that neutral sentiments dominate (38.1%), followed by positive (35.3%) and negative (26.6%) sentiments. With an evaluation accuracy of 96%, the Naive Bayes model showed that the majority of fans were neutral towards Treasure's performance. While positive sentiment indicates great enthusiasm, negative sentiment indicates that expectations were not met. This study shows that the Naive Bayes algorithm is a useful method for analyzing sentiment on social media. As such, artist management and event organizers can leverage these findings to create better communication strategies and improve fan experience. This study also adds to the literature on sentiment analysis of the entertainment industry, especially with regard to the K-pop phenomenon in Indonesia.

Keywords: Fan Sentiment, Naïve Bayes, Treasure, Mandiri Carnival, Social Media X

1. PENDAHULUAN

Fenomena K-pop telah menjadi bagian yang tak terpisahkan dari budaya pop di seluruh dunia. K-pop telah berhasil membangun basis penggemar yang aktif di seluruh dunia, termasuk Indonesia, berkat visual yang menarik, musik yang energik, dan hubungan yang kuat antara artis dan penggemarnya [1]. Salah satu yang paling populer saat ini adalah grup Treasure, telah menarik perhatian besar karena penampilan mereka yang luar biasa dan interaksi intens mereka melalui berbagai platform, seperti Weverse dan media sosial X (sebelumnya Twitter) [2]. Penampilan mereka di Karnaval Mandiri 2024 adalah momen penting dalam perjalanan mereka, yang memicu antusiasme besar dan tanggapan luas dari penggemar di media sosial. Namun, belum ada upaya yang dilakukan untuk mempelajari secara menyeluruh reaksi emosional yang diberikan penggemar melalui komentar dan unggahan.

Salah satu masalah utama dalam penelitian ini adalah kurangnya pemahaman mendalam tentang reaksi penggemar Treasure terhadap penampilan mereka di Karnaval Mandiri 2024. Analisis sentimen dapat menjadi metode yang tepat untuk menggali pola emosi penggemar, yang mencakup sentimen positif, negatif, dan netral, yang terungkap dalam unggahan yang diposting di media sosial [3]. Pemahaman terhadap pola ini sangat penting bagi manajemen artis dan penyelenggara acara untuk mengevaluasi serta meningkatkan strategi komunikasi dan pengalaman penggemar [4]. Penelitian ini menawarkan solusi dengan menggunakan algoritma Naïve Bayes, yang dikenal handal dalam mengklasifikasikan data teks besar dan beragam, dengan dukungan pustaka TextBlob untuk pelabelan otomatis yang akurat.

Penelitian sebelumnya telah menunjukkan relevansi pendekatan ini. Nurmawati et al. [5] menemukan dominasi sentimen positif dalam analisis sentimen penggemar BTS di Twitter dengan Naïve Bayes. Rizkina dan Hasan [4] melihat dominasi sentimen negatif dalam reaksi netizen terhadap pembubaran konser NCT 127. Fudholi et al. [6] menemukan bahwa pengalaman konser memengaruhi sentimen penggemar Coldplay di Twitter. Selain itu, Astuti & Astuti [7] menemukan bahwa algoritma Naïve Bayes yang dioptimalkan dengan Particle Swarm Optimization (PSO) dapat

meningkatkan akurasi analisis sentimen. Di sisi lain, Ginabila & Fauzi [8] membandingkan kinerja Naïve Bayes dengan SVM, menunjukkan bahwa Naïve Bayes memiliki keunggulan dalam klasifikasi teks.

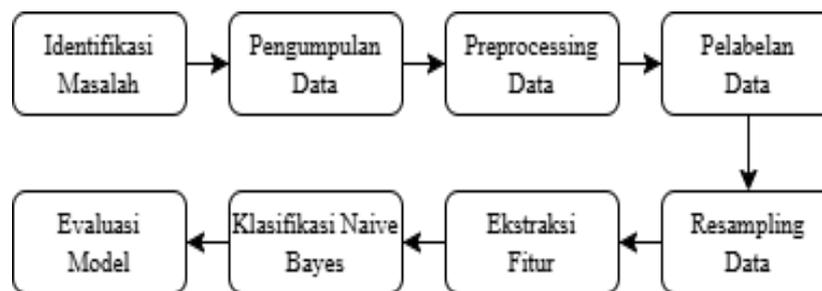
Meskipun penelitian tersebut telah mempelajari berbagai aspek analisis sentimen, masih ada ruang untuk memahami reaksi khusus penggemar terhadap penampilan grup K-pop tertentu di acara festival seperti Karnaval Mandiri 2024. Selain itu, sentimen netral sering kali dianggap tidak penting dalam penelitian sebelumnya, padahal dapat memberikan wawasan penting tentang emosi penggemar [9]. GAP ini menjadi alasan utama untuk mempelajari distribusi sentimen lebih lanjut dan menyelidiki faktor yang memengaruhi perasaan penggemar Treasure.

Penelitian ini menggunakan algoritma Naïve Bayes karena efisiensinya dalam menangani data teks besar [10]. Analisis mencakup proses preprocessing data, pelabelan sentimen menggunakan pustaka TextBlob, serta evaluasi akurasi model. Untuk menangani ketidakseimbangan data, teknik resampling SMOTE juga digunakan, memastikan model yang dihasilkan lebih optimal [11]. Dengan metode ini, penelitian diharapkan dapat mempelajari pola emosi penggemar secara menyeluruh dan menganalisis sentimen mereka dengan akurat.

Tujuan utama penelitian ini adalah untuk mengidentifikasi distribusi sentimen penggemar terhadap penampilan Treasure, mengevaluasi akurasi algoritma analisis sentimen Naïve Bayes, serta menganalisis faktor-faktor yang memengaruhi sentimen tersebut. Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan kepada literatur yang berkaitan dengan analisis sentimen berbasis media sosial, khususnya yang berkaitan dengan budaya K-pop di Indonesia [12]. Selain itu, hasil penelitian juga diharapkan dapat membantu manajemen artis dan penyelenggara acara untuk meningkatkan interaksi dengan penggemar serta merancang strategi komunikasi yang lebih efektif [13].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1 Tahapan Penelitian

- Identifikasi Masalah**
Menganalisis sentimen penggemar terhadap penampilan Treasure di Karnaval Mandiri2024, serta mengidentifikasi tantangan utama yang dihadapi selama proses analisis data.
- Pengumpulan Data**
Mengumpulkan data berupa postingan atau komentar dari media sosial X menggunakan teknik crawling, mencakup informasi relevan yang terkait dengan penampilan Treasure di Karnaval Mandiri 2024.
- Preprocessing Data**
Membersihkan data dari elemen tidak relevan seperti tanda baca, emoji, dan kata-kata kosong. Untuk memastikan kualitas dan konsistensi data yang akan digunakan dalam proses analisis selanjutnya.
- Pelabelan Data**
Memberikan label positif, negatif, atau netral secara otomatis pada data berdasarkan isi postingan atau komentar, guna mempermudah proses klasifikasi sentimen menggunakan algoritma yang telah direncanakan.
- Resampling Data**
Melakukan penyeimbangan jumlah data pada setiap kategori sentimen untuk mengatasi masalah data tidak seimbang yang dapat memengaruhi akurasi model kalsifikasi yang digunakan.
- Ekstraksi Fitur**
Mengubah data teks menjadi representasi numerik melalui teknik bagOf-words, sehingga dapat diproses oleh algoritma klasifikasi secara lebih efektif dan efisien.
- Klasifikasi Naive Bayes**
Menggunakan algoritma Naive Bayes untuk mengklasifikasikan sentimen data yang telah diproses, karena algoritma ini sederhana, cepat, dan memiliki performa baik pada data teks.
- Evaluasi model**
Mengukur kinerja model menggunakan metrik seperti akurasi, presisi, recall, dan f1 score untuk memastikan model bekerja optimal dalam mengklasifikasikan sentimen penggemar secara tepat.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Analisis

Penelitian ini bersifat kuantitatif dan data yang dihasilkan akan berbentuk angka. Berdasarkan data yang diperoleh, dilakukan analisis dengan menggunakan bahasa pemrograman Python. Penelitian ini bertujuan untuk mengukur sentimen penggemar Treasure pada Karnaval Mandiri 2024 di media sosial X. Hasil penelitian diperoleh melalui beberapa tahap, mulai dari pengumpulan data hingga evaluasi model. Berikut adalah tahapan-tahapan yang dilakukan selama penelitian:

3.1.1 Identifikasi Masalah

Langkah pertama yang dilakukan adalah mengidentifikasi tujuan utama dari penelitian, yaitu menganalisis persepsi penggemar terhadap penampilan Treasure di Karnaval Mandiri 2024. Dalam proses ini, latar belakang masalah digali untuk menentukan alasan pentingnya analisis tersebut, ini termasuk bagaimana respons penggemar dapat memengaruhi persepsi grup atau keefektifan acara. Peneliti juga menentukan ruang lingkup penelitian, seperti platform media sosial yang akan digunakan, rentang waktu pengumpulan data, dan kategori sentimen yang ingin diidentifikasi (positif, negatif, atau netral). Data tidak seimbang, keberagaman bahasa, dan volume data yang besar adalah masalah yang dapat muncul ketika proses analisis. Oleh karena itu, identifikasi masalah merupakan dasar untuk membuat strategi penelitian yang tepat dan menghasilkan hasil yang bermanfaat dan relevan.

3.1.2 Pengumpulan Data

langkah kedua adalah melakukan pengumpulan data. Data yang digunakan berasal dari media sosial X yang diambil menggunakan teknik crawling. Pustaka tweet-harvest pada google colab dimanfaatkan untuk melakukan pengambilan data secara otomatis sesuai dengan kata yang ditentukan. Langkah-langkah pada pengumpulan data bisa dilihat pada Gambar 2 dan Gambar 3 berikut:

```
[ ] twitter_auth_token = 'e8d5a59861512e9b322481275e6bf42bbd8848bd'

!pip install pandas
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" | sudo
!sudo apt-get update
!sudo apt-get install nodejs -y
!node -v
```

Gambar 2 Autentifikasi dan Instal Pustaka

```
filename = 'Data Penelitian Treasure.csv'
search_keyword = 'Treasure Mandiri Karnaval lang:Id'
limit = 1000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Gambar 3 Crawling Data

Dari proses crawling tersebut, diperoleh sebanyak 1.529 data yang akan digunakan sebagai bahan analisis. Contoh hasil crawling bisa dilihat pada Gambar 4 berikut:

id	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen
0	1851693591405656079	Wed Oct 30 18:32:09 +0000 2024	0	KELAKUAN 2024 (dulu drmm ini bje?) Jan - st.	1851693591405656079	NaN	
1	1851461589100110263	Wed Oct 30 05:06:53 +0000 2024	0	@Jaw17969 @10048ban @caratsalk ini waktu ban...	1851490940604958001	NaN	.Jaw17
2	1851439650499690992	Wed Oct 30 01:43:05 +0000 2024	2	doyung at Mandiri Kamaval 2024 #TREASURE #MA	1851439650499690992	https://pbs.twimg.com/ext_tw_video_thumb/18514...	
3	1850053020102357322	Sat Oct 26 15:44:02 +0000 2024	4	@mucclaws real waktu nnton treasure di kamava...	18500201733034697069	NaN	ttm
4	1850149654803226912	Sat Oct 26 12:17:06 +0000 2024	0	Treasure waktu mandiri kamaval juga perform 4...	1850149654803226912	NaN	

Gambar 4 Hasil Crawling Data

3.1.3 Preprocessing Data

Tahap ketiga adalah melakukan preprocessing data. Tahapan ini bertujuan membersihkan dan mempersiapkan data mentah agar siap digunakan dalam proses analisis. Preprocessing data melibatkan beberapa langkah penting, mulai pembersihan, normalisasi, stopwords, tokenisasi, hingga stemming. Setiap langkah dirancang untuk memastikan data bebas dari elemen-elemen yang tidak relevan sehingga lebih representatif untuk analisis. Berikut adalah langkah-langkah dalam preprocessing data:

1. Pembersihan Data

Membersihkan data dari elemen yang tidak relevan, seperti noise (mention, hashtag, URL, simbol, emoji, atau karakter khusus yang tidak diperlukan), data duplikat dan data kosong yang dapat memengaruhi hasil analisis, serta perubahan semua huruf menjadi huruf kecil untuk menjaga konsistensi format data. Tahapan bisa dilihat pada Gambar 5 dan Gambar 6 berikut:

```
data = data.drop_duplicates(subset=['full_text'])
data = data.dropna()
data['full_text'] = data['full_text'].str.lower()
data.head(10)
```

Gambar 5 Penghapusan nilai duplikat dan nilai kosong serta mengubah huruf menjadi kecil

```
def clean_twitter_text(text):
    text = re.sub(r'@[A-Za-z0-9_]+', '', text)
    text = re.sub(r'#\w+', '', text)
    text = re.sub(r'#[\s]+', '', text)
    text = re.sub(r'https?://\S+', '', text)
    text = re.sub(r'^A-Z0-9+', '', text)
    text = re.sub(r'\s+', '', text).strip()

    return text

data['full_text'] = data['full_text'].apply(clean_twitter_text)
```

Gambar 6 Pembersihan Data

Dari proses pembersihan data, tersisa sebanyak 1018 data yang relevan untuk analisis lebih lanjut

2. Normalisasi

Mengubah kata-kata atau frasa tidak baku menjadi bentuk baku sesuai dengan konteks analisis. Proses ini melibatkan pendefinisian kamus normalisasi (normalization dictionary) yang berisi kata tidak baku dan kata baku. Kamus ini kemudian diterapkan pada teks untuk menggantikan kata-kata yang tidak baku dengan bentuk yang lebih standar. Tahapan bisa dilihat pada Gambar 7 berikut:

```
norm = {
    'a thread' : 'kumpulan tweet', 'aaaaas' : 'aaaaakkk', 'aaaaakk' : 'aaaaakk', 'abs gini' : 'habis ini', 'account' : 'akun',
    'bjri' : 'bikin', 'bikin' : 'bikin', 'olio' : 'beliau', 'bls' : 'belum', 'bneran' : 'benaran', 'bngtiti' : 'banget',
    'fancam' : 'rekaman video', 'fancamku' : 'rekaman videoku', 'fancam/jisupportreasure' : 'rekam video suji dukung treasure',
    'high note' : 'nada tinggi', 'highnote' : 'nada tinggi', 'hiks' : 'hu kantangnya', 'hp lemahnya' : 'ht', 'has' : 'has',
    'launching' : 'peluncuran', 'ldr' : 'hubungan jarak jauh', 'leader' : 'pesipolin', 'legend' : 'legenda', 'lg' : 'lg',
    'no caption needed' : 'tidak perlu keterangan', 'nwh' : 'tuh', 'nontoo' : 'nonton', 'nontonn' : 'nontan', 'npy' : 'npy',
    'see u agallinnnn' : 'saepl jumpa lagi', 'see u' : 'saepl jumpa', 'see you again' : 'saepl jumpa', 'seekk' : 'a',
    'tour' : 'tur', 'tp' : 'taul', 'tstp' : 'sementar', 'treasureee' : 'treasure', 'treat' : 'perlakukan', 'trejo' : 'trejo'
}

def normalisasi(str_text):
    for i in norm:
        str_text = str_text.replace(i, norm[i])
    return str_text

data['full_text'] = data['full_text'].apply(lambda x: normalisasi(x))
```

Gambar 7 Normalisasi

3. Stopwords

Menghapus kata-kata umum yang tidak memberikan kontribusi makna signifikan pada analisis, seperti “yang”, “dan”, atau “di”, akan dihapus dari teks. Proses ini dilakukan untuk mengurangi gangguan yang disebabkan oleh kata-kata yang sering muncul tetapi tidak memberikan kontribusi signifikan terhadap konteks atau makna teks. Tahapan bisa dilihat pada Gambar 4.15 dan hasil bisa dilihat pada Gambar 8 berikut:

```
stop_words = set(stopwords.words('Indonesian'))

def remove_stopwords(text):
    tokens = word_tokenize(text)
    filtered_text = [word for word in tokens if word not in stop_words]
    return ' '.join(filtered_text)

data['full_text'] = data['full_text'].apply(remove_stopwords)
```

Gambar 8 Stopwords

4. Tokenize

Memecah teks menjadi unit-unit kecil berupa kata-kata atau token tersendiri. Proses ini penting untuk mempersiapkan data teks untuk analisis lebih lanjut, seperti perhitungan frekuensi kata, pengenalan pola, atau penerapan algoritma pembelajaran machine learning. Tahapan bisa dilihat pada Gambar 9 berikut:

```
data['tokenized_text'] = data['full_text'].apply(word_tokenize)
```

Gambar 9 Tokenize

5. Stemming

Mengubah setiap kata dalam teks ke bentuk dasar, sehingga variasi bentuk kata yang memiliki makna serupa dapat disederhanakan. Misalnya, kata-kata seperti “menampilkan”, “penampilan”, dan “tampilan” akan diubah ke bentuk dasar yaitu “tampil”. Tahapan bisa dilihat pada Gambar 10 berikut:

```
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stem_text(text):
    tokens = word_tokenize(text)
    stemmed_text = [stemmer.stem(word) for word in tokens]
    return ' '.join(stemmed_text)

data['stemmed_text'] = data['full_text'].apply(stem_text)
```

Gambar 10 Stemming

3.1.4 Pelabelan Data

Tahap keempat adalah melakukan pelabelan data. Tahap ini merupakan langkah penting dalam analisis sentimen karena menentukan cara mengklasifikasikan data teks berdasarkan polaritas sentimennya. Pada tahap ini, data mentah yang telah melalui preprocessing diberi label kategori sentimen positif, negatif, dan netral. Pelabelan ini dilakukan secara otomatis menggunakan algoritma TextBlob yang dirancang untuk mendeteksi pola sentimen dalam teks. Tahapan bisa dilihat pada Gambar 11 berikut:

```
def get_sentiment(text):
    blob = TextBlob(text)
    polarity = blob.sentiment.polarity
    if polarity > 0:
        return 'positif'
    elif polarity < 0:
        return 'negatif'
    else:
        return 'netral'

data['sentiment'] = data['full_text'].apply(get_sentiment)
```

Gambar 11 Pelabelan Data

Fungsi `get_sentiment(text)` digunakan untuk menentukan sentimen teks berdasarkan nilai polaritasnya. Fungsi ini menerima parameter berupa teks (String) yang sentimennya akan dianalisis. Dalam prosesnya, fungsi tersebut membuat objek `TextBlob` dari teks yang diberi dan kemudian mendapatkan nilai polaritas menggunakan atribut `sentiment.polarity`. Nilai polaritas ini berada di antara 1 (Positif) dan -1 (Negatif). Berdasarkan nilai polaritas yang diperoleh, sentimen teks ditentukan sebagai berikut:

- a. Polarity > 0 = Sentimen Positif.
- b. Polarity < 0 = Sentimen Negatif.
- c. Polarity == 0 = Sentimen Netral.

Hasil analisis sentimen berupa label (“Positif”, “Negatif”, atau “Netral”) kemudian dikembalikan oleh fungsi `data['sentiment'] = data['full_text'].apply(get_sentiment)`. Baris ini bertujuan untuk mengetahui sentimen teks yang terdapat pada kolom `full_text` dataset, kemudian hasil analisis sentimen nya disimpan pada kolom baru yang disebut `sentiment`. Tahapan bisa dilihat pada Tabel 1 berikut:

Tabel 1 Hasil Pelabelan Data

Full Text	Sentiment
sayang jeongwoo mandiri karnaval kredit saudara perempuan x	Netral
nyata nonton treasure karnaval mandiri diam beneran nyanyi bagus	Positif
ngomong ngomong bagi fancam treasure mandiri karnaval heuheu stupid	Negatif
surup manggil doyoung maafin ya dek bocor video doyoung mandiri karnaval	Netral
doyoung sayang terima kasih senyum cinta kau teume terus senyum terus bahagia bangga pada cinta pirotongtong treasure mandiri karnaval	Positif
minimal kasihan waras treasure mandiri karnaval	Negatif

3.1.5 Resampling Data

Tahap keempat adalah resampling data. Resampling merupakan teknik yang digunakan untuk mengatasi ketidakseimbangan data, dimana distribusi kategori sentimen tidak merata. Ketidakseimbangan ini dapat mempengaruhi performa model dalam memprediksi kelas minoritas karena cenderung lebih fokus pada kelas mayoritas. Tahapan bisa dilihat pada Gambar 12 hingga Gambar 14 berikut:

```
oversample = SMOTE(sampling_strategy='not majority')
undersample = RandomUnderSampler(sampling_strategy='not minority')
smote_enn = SMOTEENN()

X = data['stemmed_text']
y = data['sentiment']

vectorizer = CountVectorizer()
X_vect = vectorizer.fit_transform(X)
```

Gambar 12 Mendefinisikan Variabel Input dan Label

```
X_resampled, y_resampled = smote_enn.fit_resample(X_vect, y)
```

Gambar 13 Resampling Menggunakan SMOOTEENN

```
print("Before Resampling:")
print(y.value_counts())
print("\nAfter Resampling:")
print(pd.Series(y_resampled).value_counts())
```

```
Before Resampling:
sentiment
netral    931
positif    75
negatif    12
Name: count, dtype: int64

After Resampling:
netral    367
positif   359
negatif   292
Name: count, dtype: int64
```

Gambar 14 Hasil Resampling Data

3.1.6 Ekstraksi Fitur

Tahap keenam adalah ekstraksi fitur. Tahap ini bertujuan untuk mengubah data teks menjadi representasi numerik yang dapat diproses menggunakan algoritma machine learning. Metode yang digunakan adalah Bag of Words (BoW), dimana setiap kata dalam teks diubah menjadi fitur berdasarkan frekuensinya. Proses ini membantu model memahami informasi dalam teks secara terstruktur dan analitis. Langkah-langkah dalam ekstraksi fitur bisa dilihat pada Gambar 15 dan Gambar 16 berikut:

```
[ ] X_train = [" ".join(vectorizer.inverse_transform(doc)[0]) for doc in X_train_sparse]
X_test = [" ".join(vectorizer.inverse_transform(doc)[0]) for doc in X_test_sparse]
```

```
[ ] vectorizer_bow = CountVectorizer()
X_train_bow = vectorizer_bow.fit_transform(X_train)
X_test_bow = vectorizer_bow.transform(X_test)
```

```
[ ] feature_names = vectorizer_bow.get_feature_names_out()
feature_weights = X_train_bow.toarray()
```

Gambar 15 Ekstraksi Fitur Menggunakan Bag-of-Words

```
[ ] vocab_df = pd.DataFrame(list(vectorizer.vocabulary_.items()), columns=['word', 'index'])
print(vocab_df)
```

```
word  index
0     laku    756
1     2024     9
2     dujii   340
3     darimana 285
4     januari 587
...    ...    ...
1482  percaya 1046
1483  duhh    329
1484  mikanya 895
1485  wuhh   1410
1486  man     832
```

[1487 rows x 2 columns]

Gambar 16 Hasil Ekstraksi Fitur

3.1.7 Klasifikasi Naïve Bayes

Tahap ketujuh adalah klasifikasi sentimen menggunakan algoritma Naïve Bayes. Algoritma Naïve Bayes adalah salah satu algoritma yang populer dalam pemrosesan bahasa alami karena kesederhanaan dan efektivitasnya dalam pemrosesan data teks. Naïve bayes beroperasi berdasarkan prinsip teori probabilitas, yang mana setiap fitur dianggap independen satu sama lain, meskipun pada kenyataannya fitur-fitur tersebut mungkin terkait.

Proses klasifikasi terdiri dari dua tahap, yaitu pelatihan dan pengujian. Pada tahap pelatihan, model dilatih menggunakan data pelatihan untuk mempelajari pola distribusi kata dan hubungannya dengan kategori sentimen. Kemudian, pada tahap pengujian, model yang dilatih dievaluasi menggunakan data pengujian untuk menilai performanya dalam memprediksi sentimen secara akurat.

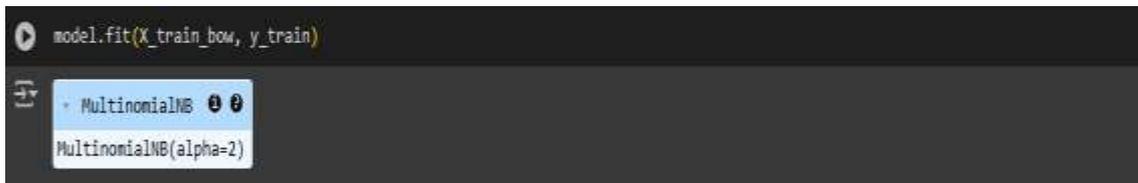
Berikut adalah langkah-langkah dalam klasifikasi sentimen menggunakan algoritma Naïve Bayes:

1. Membuat model Naïve Bayes menggunakan MultinomialNB. Salah satu yang dapat diatur adalah nilai alpha, yang digunakan untuk mengontrol proses smoothing atau penghalusan data. Tahapan bisa dilihat pada Gambar 17 berikut:

```
model = MultinomialNB(alpha=2)
```

Gambar 17 Membuat Model Naive Bayes

2. Melatih model Naïve Bayes menggunakan data latih yang telah di ekstraksi fiturnya. Tahapan bisa dilihat pada Gambar 18 berikut:



Gambar 18 Pelatihan Model Naive Bayes

3. Menguji model menggunakan data uji untuk menilai keakuratan model dalam memprediksi sentimen. Pada langkah ini, data pengujian yang terpisah dari proses pelatihan digunakan untuk menguji sejauh mana model dapat menerapkan pola yang dipelajari dari data pelatihan. Tahapan bisa dilihat pada Gambar 19 berikut:



Gambar 19 Pengujian Model Naive Bayes

3.1.8 Evaluasi Model

Tahap kedelapan adalah evaluasi model. Melihat performa model uji menggunakan data pengujian untuk memastikan seberapa baik model dapat memprediksi sentimen. Evaluasi ini dilakukan dengan berbagai metrik, seperti precision, recall, f1-score, dan akurasi. Tahapan bisa dilihat pada Gambar 20 berikut:



	precision	recall	f1-score	support
negatif	1.00	1.00	1.00	60
netral	0.97	0.99	0.98	76
positif	0.99	0.97	0.98	68
accuracy			0.99	204
macro avg	0.99	0.99	0.99	204
weighted avg	0.99	0.99	0.99	204

Akurasi Pengujian: 0.9852941176470589

Gambar 20 Evaluasi Model naive Bayes

3.2 Pembahasan

Pada bagian ini, dilakukan interpretasi terhadap hasil penelitian yang telah diperoleh untuk memberikan pemahaman lebih mendalam mengenai pengukuran sentimen penggemar Treasure pada Karnaval Mandiri 2024 di media sosial X.

3.2.1 Interpretasi Hasil Analisis

Model yang digunakan menunjukkan performa yang sangat baik dengan rata-rata nilai precision, recall, dan f1-score di atas 0,97 untuk ketiga kategori sentimen (positif, netral, dan negatif). Tingginya akurasi model ini menunjukkan bahwa algoritma Naïve Bayes mampu menangkap pola pada data secara efektif, meskipun pada kategori sentimen netral hanya terdapat sedikit kesalahan klasifikasi (confusion matrix menunjukkan 1 data netral diklasifikasikan sebagai positif dan 2 data positif diklasifikasikan sebagai netral).

Dari total 1018 kalimat dalam dataset, distribusi sentimen menunjukkan bahwa kalimat dengan sentimen netral memiliki proporsi terbesar (36,1%), diikuti oleh sentimen positif (35,3%). Hal ini mencerminkan fakta bahwa percakapan penggemar Treasure di media sosial lebih cenderung mengungkapkan opini netral atau positif dibandingkan opini negatif.

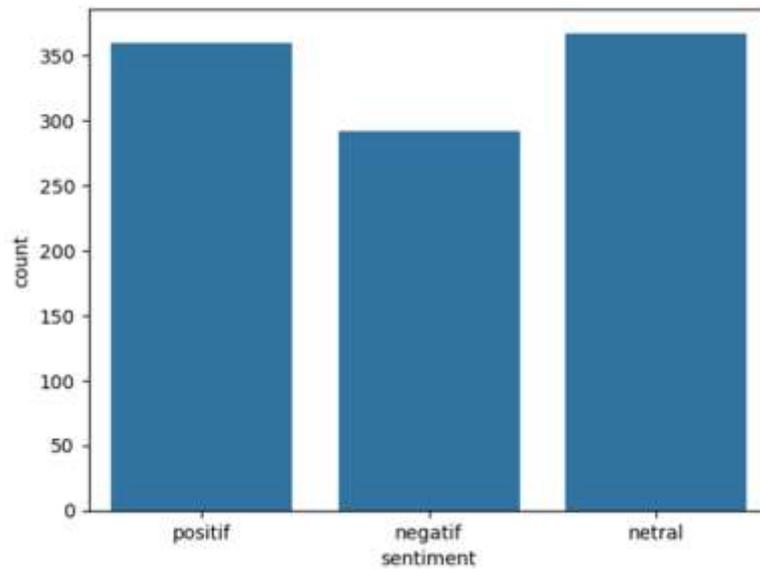
Untuk meningkatkan pemahaman terhadap hasil analisis, visualisasi wordcloud digunakan untuk mengidentifikasi kata-kata yang paling sering muncul dalam dataset. Hasil visualisasi wordcloud menunjukkan bahwa kata-kata yang paling sering muncul mencakup istilah yang terkait dengan topik utama yang dibahas, seperti nama-nama anggota, istilah tentang konser, dan kata-kata yang berkaitan dengan pengalaman penggemar.

3.2.2 Visualisasi Data

Berikut adalah visualisasi hasil penelitian untuk memperkuat penafsiran:

1. Distribusi Sentimen

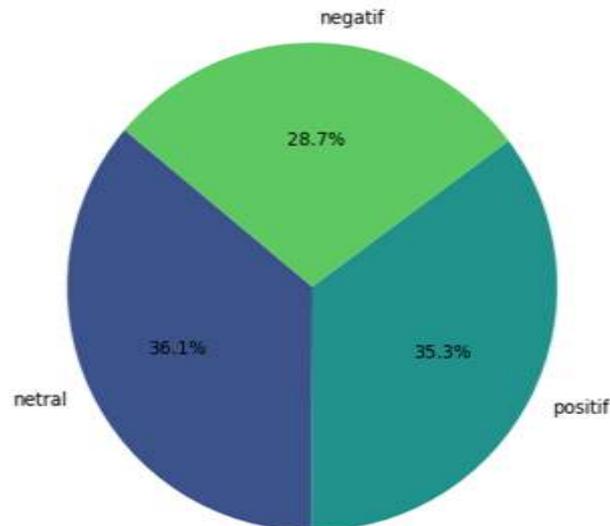
Distribusi sentimen menggambarkan bagaimana data terbagi ke dalam kategori sentimen, misalnya negatif, netral, atau positif. Visualisasi ini membantu memvalidasi bahwa proses resampling berhasil menjaga keseimbangan data. Visualisasi bisa dilihat pada Gambar 21 berikut:



Gambar 21 Bagan Distribusi Sentimen

2. Persentase Distribusi

Persentase distribusi sentimen memberikan gambaran proporsi masing-masing kategori sentimen secara keseluruhan. Hal ini penting untuk melihat representasi data yang lebih terperinci. Visualisasi bisa dilihat pada Gambar 22 berikut:



Gambar 22 Bagan Persentase Distribusi

3. Confusion Matrix

Menunjukkan performa model klasifikasi dengan membandingkan hasil prediksi dan data aktual untuk tiap kelas sentimen. Visualisasi bisa dilihat pada Gambar 23 berikut:

keandalan algoritma dalam mengklasifikasikan sentimen positif, negatif, dan netral dalam data penelitian. Dengan akurasi 96,08% dan nilai precision, recall, dan f1-score yang tinggi, algoritma Naive Bayes menunjukkan kinerja yang sangat baik dalam analisis sentimen. Nilai f1-nya rata-rata 0,99, menunjukkan bahwa algoritma dapat mengklasifikasikan sentimen positif, negatif, dan netral dalam data penelitian.

UCAPAN TERIMAKASIH

Dengan penuh rasa syukur, penulis mengucapkan terima kasih kepada Allah SWT atas rahmat, hidayah, dan kekuatan yang telah diberikan sehingga artikel ini dapat terselesaikan dengan baik. Ucapan terima kasih juga penulis sampaikan kepada orang tua dan keluarga tercinta yang selalu memberikan doa, dukungan, dan cinta tanpa batas. Penghargaan yang mendalam juga penulis sampaikan kepada dosen pembimbing yang telah memberikan bimbingan, ilmu, serta arahan yang berharga selama proses penyusunan artikel ini. Tidak lupa, terima kasih kepada teman-teman yang selalu memberikan semangat, masukan, dan motivasi. Terakhir, penulis juga ingin mengucapkan terima kasih kepada para idola yang menjadi inspirasi dalam perjalanan ini, karena karya dan dedikasi mereka telah memberikan motivasi untuk terus berusaha mencapai yang terbaik.

DAFTAR PUSTAKA

- [1] M. A. Z. Larasati, N. A. S. Winarsih, M. S. Rohman, and G. W. Saraswati, "Penerapan Metode K-Means Clustering Dalam Menganalisis Sentimen Masyarakat Terhadap K-Popers Pada Twitter," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, pp. 201–210, 2022, doi: 10.35889/progresif.v18i2.877.
- [2] R. Aditriyani and K. Sinduwiatmo, "Weverse Menghubungkan Boygroup Treasure dan Para Penggemar Mereka," *Interact. Commun. Stud. J.*, vol. 1, no. 2, pp. 82–96, 2024, doi: 10.47134/interaction.v1i2.2881.
- [3] N. Raisa and N. Riza, "Sentimen Analisis Terhadap Opini Masyarakat Mengenai Drama Korea Pada Twitter Menggunakan Metode Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1312–1320, 2023, doi: 10.36040/jati.v7i2.6765.
- [4] 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.
- [5] P. Nurmawati, E. Supriyati, and T. Listiyorini, "Analisis Sentimen Terhadap Penggemar K-Pop Di Media Sosial Twitter Menggunakan Naive Bayes (Studi Kasus Penggemar Grup Bts)," *J. Inf. Eng. Educ. Technol.*, vol. 4, no. 2, pp. 86–89, 2020, doi: 10.26740/jieet.v4n2.p86-89.
- [6] L. A. Fudholi, N. Rahaningsih, and R. D. Dana, "SENTIMEN ANALISIS PERILAKU PENGGEMAR COLDPLAY DI MEDIA SOSIAL," *J. Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 4150–4159, 2024.
- [7] T. Astuti and Y. Astuti, "Analisis Sentimen Review Produk Skincare Dengan Naïve Bayes Classifier Berbasis Particle Swarm Optimization (PSO)," *J. Media Inform. Budidarma*, vol. 6, no. 4, pp. 1806–1815, 2022, doi: 10.30865/mib.v6i4.4119.
- [8] Ginabila and A. Fauzi, "Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine," *J. Ilm. Ilk. - Ilmu Komput. dan Inform.*, vol. 6, no. 2, pp. 111–122, 2023.
- [9] R. Slamet, W. Gata, A. Novtariany, K. Hilyati, and F. A. Jariyah, "Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 145–153, 2022, doi: 10.31539/intecom.v5i1.3933.
- [10] A. R. Isnain, N. S. Marga, and D. Alita, "Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 15, no. 1, pp. 55–64, 2021, doi: 10.22146/ijccs.60718.
- [11] Suryani, M. F. Fayyad, D. T. Savra, V. Kurniawan, and B. H. Estanto, "Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm," *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 1, no. 1, pp. 1–9, 2023, doi: 10.57152/predatecs.v1i1.814.
- [12] P. R. A. Savitri, I. M. A. D. Suarjaya, and W. O. Vihikan, "Sentiment Analysis of X (Twitter) Comments on The Influence of South Korean Culture in Indonesia," *J. Inf. Syst. Informatics*, vol. 6, no. 2, pp. 979–991, 2024, doi: 10.51519/journalisi.v6i2.749.
- [13] G. Maria, Garno, and T. N. Padilah, "ANALISIS SENTIMEN FENOMENA FOMO PADA KONSER K-POP DI INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 4, pp. 2548–2554, 2024, doi: 10.36040/jati.v7i4.7104.