

Penggunaan Algoritma Naïve Bayes dengan Polarity Textblob untuk Analisis Sentimen pada Acara ASEAN CUP 2024 U-16 di Media Sosial Twitter

Arya Erlangga¹, Yani Parti Astuti², Etika Kartikadarma³, Sindhu Rakasiwi⁴, Egia Rosi Subhiyakt⁵

^{1,2}Teknik Informatika, Universitas Dian Nuswantoro, Indonesia

E-mail: 111201911940@mhs.dinus.ac.id¹, yanipartiastuti@dsn.dinus.ac.id², etika.kartikadarma@dsn.dinus.ac.id³, sindhu.rakasiwi@dsn.dinus.ac.id⁴, egia@dsn.dinus.ac.id⁵

Abstract: Football is a popular sport in the world and is enjoyed by people of all ages. The Indonesia U-16 national team played in the ASEAN CUP 2024 event in this field. Twitter users gave their support through #timnasday during the event. This provided many forms of support for the Indonesian national team which made it difficult to identify positive, neutral, and negative sentiments. This requires the use of lexicon-based textblob to perform automatic labeling. In the labeling results using textblob from a total of 1138 user tweet data resulted in positive sentiment values of 50.9% or 579 positive data, neutral 33.7% or 384 neutral data, and negative 15.4% or 175 negative data. In the test results using one of the machine learning from the naïve bayes classifier, namely gaussian naïve bayes with the division of test data and training data of 0.3 and 0.7, the accuracy value is 98.53%

Keywords: #timnasday, Textblob, naïve bayes, gaussian naïve bayes, twitter

Abstrak: Sepak bola merupakan olahraga yang populer di dunia dan di nikmati oleh semua kalangan usia. Tim nasional Indonesia U-16 bermain pada acara ASEAN CUP 2024 dalam bidang ini. Pengguna twitter memberikan dukungan mereka melalui #timnasday selama acara berlangsung. Hal ini memberikan banyak bentuk dukungan untuk timnas Indonesia yang sulit diketahui sentimen positif, netral, dan negatif. Hal ini memerlukan penggunaan textblob yang berbasis lexicon-based untuk melakukan labelling otomatis. Dalam hasil pelabelan menggunakan textblob dari total 1138 data tweet pengguna menghasilkan nilai sentiment positif 50.9% atau 579 data positif, netral 33.7% atau 384 data netral, dan negatif 15.4% atau 175 data negatif. Dalam hasil pengujian menggunakan salah satu machine learning dari naïve bayes classifier yaitu gaussian naïve bayes dengan pembagian data tes dan data training sebesar 0.3 dan 0.7 menghasilkan nilai akurasi sebesar 98.53%

Kata kunci: #timnasday, Textblob, naïve bayes, gaussian naïve bayes, twitter

1. PENDAHULUAN

Sepak bola Salah satu olahraga paling populer di dunia dan memiliki penggemar dari berbagai usia dari anak-anak hingga orang dewasa bahkan orang – orang yang sudah tua. Banyak kejuaraan yang diadakan oleh berbagai negara, salah satunya adalah Piala ASEAN U-16. Peserta dari kejuaraan ASEAN U-16 diikuti oleh tim sepak bola dari negara – negara ASEAN termasuk Indonesia. Dukungan masyarakat Indonesia terhadap tim nasional U-16 dalam mengikuti kejuaraan tersebut sangat luar biasa. Dukungan tersebut berupa dukungan dengan cara menonton langsung maupun melalui komentar – komentar di media sosial. Salaha satu media sosial yang andil dalam komentar tersebut adalah Twitter.

Twitter sebagai platform media sosial yang berfokus pada penyebaran informasi cepat dan singkat, telah berkembang menjadi ruang komunikasi digital yang digunakan

oleh berbagai kalangan untuk berbagai tujuan, termasuk berbagi informasi, berdiskusi, dan bahkan berjualan [1]. Dalam konteks sepak bola, khususnya Piala ASEAN 2024 U-16, twitter menjadi sarana bagi masyarakat untuk menyuarakan pendapat dan dukungan mereka terhadap Timnas Indonesia U-16. Melalui analisis sentimen, tanggapan yang tersebar di media sosial ini dapat diklasifikasikan untuk memahami bagaimana dukungan publik dapat mempengaruhi perkembangan dan performa tim.

Dalam penelitian ini, pendekatan data mining dengan Natural Language Processing (NLP) digunakan untuk menganalisis sentimen masyarakat yang diungkapkan melalui tweet dengan tagar #timnasday. Metode Naïve Bayes, yang dikenal dengan keakuratannya dalam analisis sentimen, diterapkan untuk mengklasifikasikan teks menjadi sentimen positif, netral, dan negatif. Hasil dari analisis ini diharapkan dapat memberikan wawasan bagi pihak-pihak terkait, termasuk manajemen tim dan pendukung, untuk memperbaiki strategi komunikasi dan mendukung perkembangan tim nasional secara lebih efektif.

2. LANDASAN TEORI

Penelitian terkait yang dilakukan oleh (Azhar et al., 2022) yang menggunakan data tweet dengan tagar #crypto sebanyak 1032 tweet. Setelah pelabelan otomatis menggunakan textblob, jumlah sentimen positif sebanyak 61.24%, sentimen netral sebanyak 26.68%, dan sentimen negatif sebanyak 10.07%. Nilai akurasi dengan naive bayes classifier sebesar 71.98%, precision sebesar 83.04%, recall sebesar 60.88%, dan nilai f1_score 71.98%. Nilai akurasi, precision, recall, dan f1_score masing-masing data test dan data training adalah 0,2 dan 0,8 [2]. Penelitian terkait lainnya yang dilakukan oleh (Kurniawan et al., 2024) dengan data tweet mengenai AFC U-23 Asian Cup 2024 sejumlah 925 tweet. Nilai sentimen yang didapatkan berupa 287 sentimen positif, 37 sentimen negatif, dan 318 sentimen netral dengan menggunakan algoritma naive bayes mendapatkan nilai akurasi sebesar 78% [3].

2.1 Text Mining

Text mining adalah bagian dari data mining yang melibatkan proses ekstraksi pengetahuan dan informasi dari pola-pola yang ditemukan dalam sekumpulan dokumen teks dengan menggunakan alat analisis tertentu. Text mining juga dapat digunakan untuk beberapa kebutuhan, seperti summarization, pencarian dokumen teks, dan analisis sentiment [4]. Dalam text mining memiliki fungsi untuk menganalisis informasi yang ada pada teks dalam mendapatkan pola, tren, dan informasi yang bermanfaat. Sentimen analisis memiliki pengaruh dalam pengambilan

Keputusan berdasarkan opini publik terhadap produk, layanan, atau isu – isu tertentu.

2.2 Textblob

Textblob adalah library dalam python yang dirancang untuk memproses data tekstual dan memiliki fitur dasar pemrosesan bahasa natural (NLP) [5]. dengan API yang dapat digunakan dalam bidang pemrosesan bahasa alami (NLP), seperti ekstraksi frasa kata benda, analisis sentimen, klasifikasi, penerjemahan, dan sebagainya . Hasil objek TextBlob digunakan untuk memproses pembelajaran bahasa alami, dan library TextBlob hanya dapat memahami Bahasa Inggris

2.3 Gaussian Naive bayes .

Naive Bayes adalah teknik pengklasifikasian yang paling umum dan memiliki tingkat keakuratan yang tinggi. Banyak penelitian tentang pengklasifikasian telah menggunakan algoritma naive bayes. Berbeda dengan metode pengklasifikasian yang menggunakan logistic regression ordinal dan nominal, algoritma ini tidak membutuhkan pemodelan atau uji statistik. Dalam proses klasifikasi penelitian, metode Naive Bayes digunakan karena bekerja dengan cepat dan efisien sebagai metode pengklasifikasian teks[3].

gaussian naïve bayes yang digunakan dalam pengklasifikasian data dengan menghitung probabilitas menggunakan Teorema Bayes. Berikut adalah persamaan yang digunakan dalam gaussian naïve bayes:

$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp \left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2} \right) \quad (1)$$

Keterangan :

$p(x_i | y)$: Probabilitas x_i terhadap kelas y

x_i : Nilai Kelas- i

σ : Nilai Varian

μ : Rata – rata Kelas y

π : Phi (3.14)

2.4 Confusion Matrix

Confusion matrix adalah alat evaluasi yang digunakan untuk menilai kinerja model klasifikasi. Ini adalah tabel yang menunjukkan jumlah prediksi yang benar dan

salah yang dibuat oleh model, terpisah berdasarkan kelas atau kategori. Matrik confusion membantu kita memahami seberapa baik model mengklasifikasikan data ke dalam kategori yang benar. Data positif yang diprediksi positif akan dianggap true positif, sedangkan data negatif yang diprediksi negatif akan dianggap true negatif, dan data positif yang diprediksi negatif akan dianggap false positif begitu juga dengan data negatif yang di prediksi positif akan dianggap false negatif [2].

Dengan menggunakan confusion matriks bisa dilakukan perhitungan menggunakan rumus:

a. Akurasi

Akurasi menghitung nilai persentase prediksi yang dibuat oleh model.

Perhitungan dilakukan dengan rumus:

$$akurasi = \frac{\text{Prediksi benar}}{\text{total prediksi}} \quad (2)$$

b. Presisi

Presisi menghitung nilai sentiment yang sesuai dengan sentimennya.

Contohnya seperti data yang memiliki nilai negatif di prediksi dengan hasil nilai negatif. Perhitungan dilakukan dengan rumus:

$$Presisi = \frac{\text{True Negatif (TNeg)}}{\text{True Negatif (TNeg)} + \text{False Negatif (FNeg)}} \quad (3)$$

c. Recall

recall menghitung nilai sentimen yang berhasil di prediksi dengan benar.

Perhitungan dilakukan dengan rumus:

$$Recall = \frac{\text{True positif (Tpos)}}{\text{True positif (Tpos)} + \text{False Netral (FNet)} + \text{False Negatif (FNeg)}} \quad (4)$$

d. F1-score

F1-score merupakan nilai rata – rata dari presisi dan recall. Perhitungan dilakukan dengan rumus:

$$f1 - score = 2 \left(\frac{\text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} \right) \quad (5)$$

3. METODE PENELITIAN

Tinjauan Data

Data yang digunakan merupakan data tweet dengan tagar #timnasday pada tanggal

1 juli 2024 hingga 3 juli 2024 dengan jumlah 1145 data tweet. Data yang sudah didapatkan disimpan dalam format CSV (Comma Separated Value).

conversation_id	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name
1,80865E+23	Wed Jul 03 23:44:04	0	Hasil Pertandingan	1,80865E+23	https://pbs.t	
1,80864E+23	Wed Jul 03 23:20:14	0	Games online dap	1,80864E+23	https://pbs.t	
1,80864E+23	Wed Jul 03 22:59:39	0	Meski tidak juara	1,80864E+23		
1,80863E+23	Wed Jul 03 22:40:35	0	Timnas U16 Menar	1,80863E+23	https://pbs.t	
1,80863E+23	Wed Jul 03 22:25:19	7	Yesss ..Yess...kita r	1,80863E+23	https://pbs.t	
1,80858E+23	Wed Jul 03 18:56:24	1	Anak bekasi nih bc	1,80858E+23	https://pbs.t	
1,80857E+23	Wed Jul 03 18:53:26	0	Garuda Muda • #	1,80857E+23	https://pbs.t	
1,80857E+23	Wed Jul 03 18:19:40	1	Info game gacor ha	1,80857E+23	https://pbs.t	
1,80815E+23	Wed Jul 03 17:36:11	0	@BlogBayuWin @	1,80856E+23		BlogBayuWin
1,80815E+23	Wed Jul 03 17:35:52	0	@BlogBayuWin @	1,80856E+23		BlogBayuWin
1,80855E+23	Wed Jul 03 17:01:22	2	Menyala Dek GHOL	1,80855E+23	https://pbs.t	
1,80852E+23	Wed Jul 03 15:29:23	0	Gholy best player	1,80852E+23		
1,80852E+23	Wed Jul 03 15:28:51	0	Timnas indonesia	1,80852E+23		
1,80852E+23	Wed Jul 03 15:28:06	2	selamattt garuda n	1,80852E+23	https://pbs.t	
1,80808E+23	Wed Jul 03 15:25:43	0	@luqman_am @ar	1,80852E+23		luqman_am
1,80852E+23	Wed Jul 03 15:24:34	2	Congratss buat tim	1,80852E+23	https://pbs.t	
1,8083E+22	Wed Jul 03 15:23:27	0	@ino_wilmac @co	1,80852E+23		ino_wilmac
1,80777E+23	Wed Jul 03 15:23:15	0	Menyala Gholy BE	1,80852E+23		BJ_NJIR

Gambar 1 Raw Data

Selanjutnya dilakukan preprocessing data dengan mereduksi data yang diperlukan yaitu atribut `full_text`. Atribut `full_text` berisi tentang unggahan para pengguna sosial media twitter yang digunakan untuk penelitian ini.

full_text
Hasil Pertandingan Sepak Bola hari ini . . . Fo
Games online dapat uang rupiah nyata! Mud
Meski tidak juara AFF-U16 sudah sukses mer
Timnas U16 Menang Atas Vietnam dan Berha
Yesss ..Yess...kita memang harus menunjukk
Anak bekasi nih bos senggol dong ! ĩ, õõ¥õõõ
õ†õõõõ«õ°õõõõ õõõõ°õõõõ â€£ #KitaGaruda â€
Info game gacor hari -PRAGMATIC PLAY -PG :
@BlogBayuWin @ardisatriawan Cocoklogi le
@BlogBayuWin @ardisatriawan Kalau ngelia
Menyala Dek GHOLY #TimnasDay https://t.α
Gholy best player seneng banget liatnya huh
Timnas indonesia 3rd place !!! Selamat !!! #1

Gambar 2 reduksi data

Metode pengumpulan data

1. Metode Literatur

pada tahap ini perlu melakukan pemahaman dan pengkajian terhadap Scrapping

data, dan Textblob. Hal ini di perlukan untuk peneliti menentukan metode yang sesuai dengan data yang akan digunakan.

2. Metode Observasi

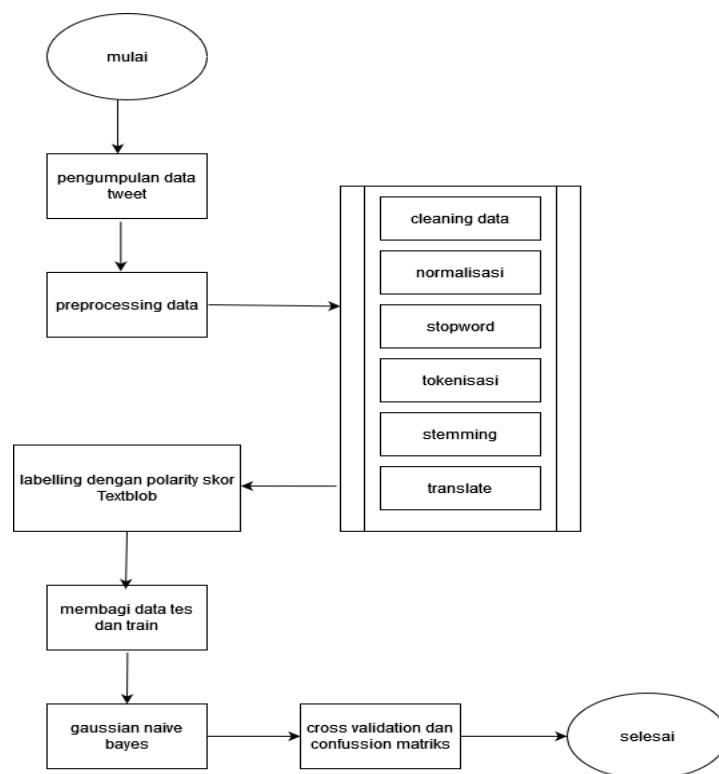
Proses ini dilakukan untuk memverifikasi keakuratan dan kepercayaan data yang diperoleh. Selain itu, observasi dilakukan untuk mengidentifikasi variabel tambahan yang mungkin berdampak pada data.

3. Scrapping Data Tweet

Pada tahap ini dilakukan pengambilan data tweet menggunakan tweet harvest. Tweet harvest merupakan library yang digunakan untuk scrapping data dengan token twitter menggunakan kata kunci untuk data yang ingin digunakan sebagai data penelitian.

Metode penelitian

Metode penelitian yang akan digunakan dalam penelitian ini digambarkan pada gambar 3 sebagai berikut :



Gambar 3 metode yang diusulkan

4. HASIL DAN PEMBAHASAN

Pegolahan pada data menggunakan IDE (Integrated Developement Enviroment) jupyter notebook sebagai tools untuk pengolahan dan penujian pada data menggunakan

bahasa pemrograman python.

4.1 preprocessing

Sebelum melakukan penelitian dilakukan persiapan *library* yang akan digunakan terlebih dahulu.

```
#Library yang digunakan
import pandas as pd
import re
import seaborn as sns
import matplotlib.pyplot as plt
import Sastrawi
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemover
from Sastrawi.StopWordRemover.StopWordRemoverFactory import ArrayDictionary
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from googletrans import Translator
from textblob import TextBlob
from wordcloud import WordCloud, STOPWORDS
from collections import Counter
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, classification_report
import numpy as np
from sklearn.model_selection import cross_val_score, KFold
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import learning_curve
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
```

Gambar 4 Library yang digunakan

Pada tahap preprocessing dilakukan beberapa tahapan, antara lain:

1. cleaning data dan case folding

Pada cleaning data dilakukan penghapusan data duplikat, data null, menghapus username, menghapus tagar '#', menghapus retweet, menghapus URL, menghapus angka dan simbol, menghapus format tebal, dan menghapus bendera.

Tabel 1 proses Cleaning Data dan case folding

Raw Data	Cleaning Data
Meski tidak juara AFF-U16 sudah sukses membantai Vietnam dengan skor 5-0 itu sudah luar biasa. Selamat Timnas Indonesia U-16 meraih tiga terbaik! #TimnasDay	meski tidak juara affu sudah sukses membantai vietnam dengan skor itu sudah luar biasa selamat timnas indonesia u meraih tiga terbaik timnasday
Anak bekasi nih bos senggol dong ! Player of the Tournament [5] main [5] gholy ▸ #KitaGaruda #timnasU16 #TimnasDay https://t.co/xbUUGNSYal	anak bekasi nih bos senggol dong player of the tournament main gholy kitagaruda timnasu timnasday
Garuda Muda ▸ #KitaGaruda ▸ #TimnasDay #TUMI @1KinLabs #1KIN #MoonBase #GR1D \$GRID https://t.co/FxFHT0i2iS	garuda muda kitagaruda timnasday tumi kin moonbase grd grid

2. Normalisasi Data

Pada proses normalisasi data dilakukan perubahan kata yang *typo* menjadi kata umum.

3. Stopword

Pada tahap ini stopword dilakukan untuk menghilangkan kata – kata yang tidak memiliki makna yang jelas pada data teks twitter. Proses ini dilakukan dengan *library* sastrawi.

4. Tokenisasi

Proses tokenisasi merupakan proses mengubah kalimat menjadi beberapa kata. Proses ini dilakukan untuk memudahkan proses stemming.

Tabel 2 proses Tokenisasi

Cleaning Data	Tokenisasi
meski tidak juara affu sudah sukses membantai vietnam dengan skor itu sudah luar biasa selamat timnas indonesia u meraih tiga terbaik timnasday	['meski', 'juara', 'affu', 'sukses', 'membanicetryai', 'vietnam', 'skor', 'sudah', 'luar', 'biasa', 'selamat', 'tim', 'nasional', 'indonesia', 'kamu', 'meraih', 'tiga', 'terbaik']
anak bekasi nih bos senggol dong player of the tournament main gholy kitagaruda timnasu timnasday	['anak', 'bekasi', 'bos', 'senggol', 'player', 'the', 'tournamenicetry', 'main', 'gholy', 'kitagaruda', 'timnasu']
garuda muda kitagaruda timnasday tumi kin moonbase grd grid	['garuda', 'muda', 'kitagaruda', 'tumi', 'kin', 'moonbase', 'grd', 'grid']

5. Stemming

Pada tahap ini melakukan perubahan pada kata yang memiliki imbuhan menjadi kata dasar. Proses ini menggunakan *library* sastrawi,

Tabel 3 proses Stemming

Tokenisasi	Stemming
['meski', 'juara', 'affu', 'sukses', 'membanicetryai', 'vietnam', 'skor', 'sudah', 'luar', 'biasa', 'selamat', 'tim', 'nasional', 'indonesia', 'kamu', 'meraih', 'tiga', 'terbaik']	meski juara affu sukses membanicetryai vietnam skor sudah luar biasa selamat tim nasional indonesia kamu raih tiga baik
['anak', 'bekasi', 'bos', 'senggol', 'player', 'the', 'tournamenicetry', 'main', 'gholy', 'kitagaruda', 'timnasu']	anak bekas bos senggol player the tournamenicetry main gholy kitagaruda timnasu
['garuda', 'muda', 'kitagaruda', 'tumi', 'kin', 'moonbase', 'grd', 'grid']	garuda muda kitagaruda tum kin moonbase grd grid

6. Translate

Pada tahap ini menggunakan library googletrans untuk mengubah kata dari bahasa indonesia menjadi bahasa inggris. Proses ini dilakukan karena pada pelabelan menggunakan Textblob hanya mengenali bahasa inggris dapat dilihat pada gambar 5.

```
from googletrans import Translator

def convert_to_english(text):
    translator = Translator()
    try:
        translated = translator.translate(text, src='id', dest='en')
        return translated.text # Return only the translated text
    except Exception as e:
        print(f"Error translating text: {text} - {e}")
        return text
```

Gambar 5 proses terjemahan

Setelah melakukan proses preprocessing dilakukan pengecekan nilai NULL atau data kosong.

```
data = data.dropna()
data.shape

(1138, 1)
```

Gambar 6 menghapus data NULL

4.2 Labelling Textblob

Labelling dilakukan untuk mendapat nilai polarity skor yang ada pada textblob. Polarity skor textblob digunakan untuk menentukan sentimen yang ada pada teks.

```
data_tweet = list(data['en_garuda'])
polaritas = 0

status = []
tot_pos = tot_neg = tot_net = total = 0

for i, tweet in enumerate(data_tweet):
    analisis = TextBlob(tweet)
    polaritas += analisis.polarity

    if analisis.sentiment.polarity > 0.0:
        tot_pos += 1
        status.append('positif')
    elif analisis.sentiment.polarity == 0.0:
        tot_net += 1
        status.append('netral')
    else:
        tot_neg += 1
        status.append('negatif')

    total += 1
```

Gambar 7 proses sentimen textblob

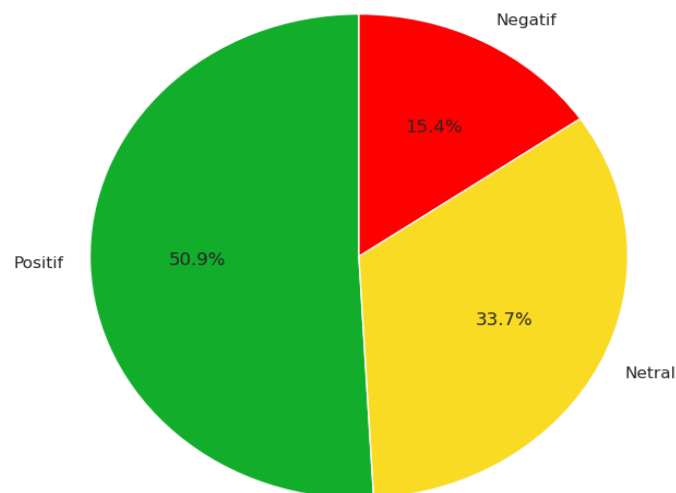
Pada gambar 7 proses sentimen dengan textblob untuk menentukan positif, netral, dan negatif berdasarkan pada polarity skor yang ada pada textblob. Proses selanjutnya mendapatkan nilai polarity skor textblob sesuai pada gambar 8.

```
def get_polarity(text):  
    return TextBlob(text).sentiment.polarity  
  
data['Polarity'] = data['en_garuda'].apply(get_polarity)
```

Gambar 8 mendapatkan nilai polarity skor

Hasil sentimen analisis menggunakan textblob dengan sentimen positif, netral, dan negatif mendapatkan 579 data positif, 385 netral, dan 174 negatif. Hasil dari persentase banyaknya label dapat dilihat pada gambar 9.

Persentase Sentimen Timnas Indonesia U-16 pada piala AFF 2024



Gambar 9 persentase sentimen

4.3 Algoritma Gaussian Naive Bayes

Pada proses implementasi algoritma gaussian naive bayes, langkah awal yang perlu dilakukan adalah pembagian data tes dan data train. Pembagian pada penelitian ini menggunakan 30% data tes dan 70% data training. Proses pembagian dapat dilihat pada gambar 9.

```
# Split data untuk validasi  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

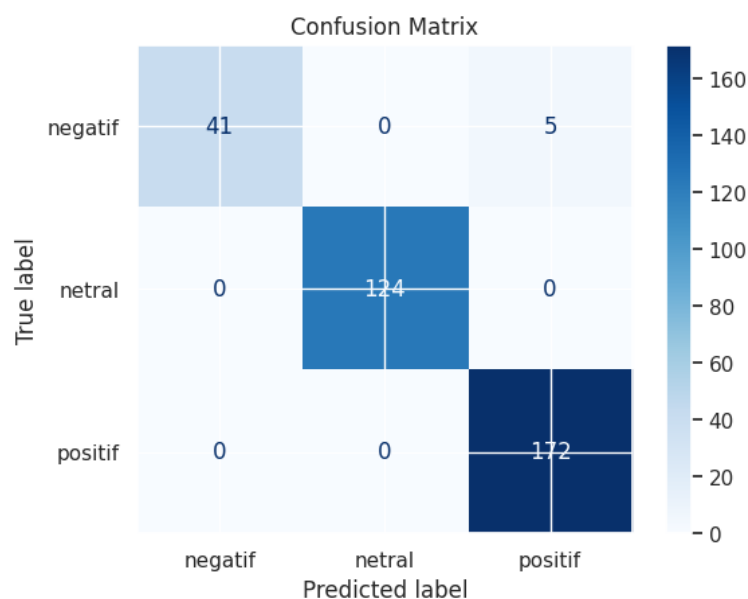
Gambar 10 pembagian data

Setelah melakukan pembagian data dengan perbandingan 30:70, dilakukan implementasi dengan menggunakan algoritma gaussian naive bayes. Proses implementasi algoritma ditampilkan pada gambar 11.

```
# Train Gaussian Naive Bayes
nb_model = GaussianNB()
nb_model.fit(X_train, y_train)
```

Gambar 11 implementasi gaussian naive bayes

Agar proses dapat dilihat dengan lebih baik, dapat dilihat pada visualisasi confusion matrix pada gambar 12.



Gambar 12 confusion matrix

Pada gambar 11 confusion matrix, ditampilkan dari jumlah data tes sejumlah 342 terbagi dalam True negatif sebanyak 41, True netral 124, dan True positif. Pada False positif dan False negatif terdapat kesalahan prediksi sejumlah 5 data.

4.4 Evaluasi

Setelah proses implementasi algoritma dilakukan, selanjutnya dilakukan proses evaluasi. Hasil evaluasi menampilkan akurasi algoritma gaussian naive bayes sebesar 98.53%. hasil akurasi pada sentimen negatif sebesar 100%, netral 100%, dan positif sebesar 97%. Untuk recall negatif sebesar 89%, netral 100%, dan positif 100%. Untuk F1_score negatif sebesar 94%, netral 100%, positif 99%. Hasil evaluasi dapat dilihat pada gambar 13.

Accuracy: 0.9853801169590644				
	precision	recall	f1-score	support
negatif	1.00	0.89	0.94	44
netral	1.00	1.00	1.00	123
positif	0.97	1.00	0.99	175
accuracy			0.99	342
macro avg	0.99	0.96	0.98	342
weighted avg	0.99	0.99	0.99	342

Gambar 13 akurasi algoritma

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, mendapatkan hasil dari penelitian berupa:

1. Dengan jumlah data sebanyak 1145 data tweet, akurasi menggunakan algoritma gaussian naive bayes senilai 98.53% pembagian data tes dan training 30% dan 70%.
2. hasil akurasi pada sentimen negatif sebesar 100%, netral 100%, dan positif sebesar 97%. Untuk recall negatif sebesar 89%, netral 100%, dan positif 100%. Untuk F1_score negatif sebesar 94%, netral 100%, positif 99%.

Adapun saran yang dapat diberikan pada penelitian selanjutnya, sebagai berikut:

1. Memperluas dataset yang digunakan untuk menampilkan keakuratan yang lebih baik dan memberikan hasil yang lebih baik terhadap persebaran kata pada sentiment analisis.
2. Pelabelan pada data bisa menggunakan metode lain dengan penyesuaian pada algoritma yang digunakan agar dapat menghasilkan model yang baik.

DAFTAR PUSTAKA

- Azhar, R., Surahman, A., & Juliane, C. (2022). Analisis sentimen terhadap cryptocurrency berbasis Python TextBlob menggunakan algoritma Naïve Bayes. Jurnal Teknologi dan Informasi.
- Kurniawan, M. A., Yuniarti, S., & Makin, S. (2024). Analisis sentimen AFC U23 Asian Cup 2024 menggunakan algoritma Naïve Bayes. Jurnal IPSIKOM, 12(1).
- Lestari, N. A., et al. (2020). Metode Naive Bayes classifier dengan TextBlob untuk analisis sentimen terhadap pelayanan Indihome dan First Media. Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK), 4(1). Retrieved from <https://t.co/Ws2wOyU5kz>

- Muzaki, A., & Witanti, A. (2021). Sentiment analysis of the community in the Twitter to the 2020 election in pandemic COVID-19 by method Naive Bayes classifier. *Jurnal Teknik Informatika (Jutif)*, 2(2), 101–107. <https://doi.org/10.20884/1.jutif.2021.2.2.51>
- Nugroho, H., & Sari, D. P. (2023). Penerapan Naïve Bayes untuk analisis sentimen dalam pemilihan umum. *Jurnal Komunikasi dan Media*, 10(3), 112–120.
- Pratama, R., & Lestari, M. (2023). Penggunaan algoritma Naïve Bayes dalam analisis sentimen terhadap opini masyarakat terhadap produk baru. *Jurnal Sistem Informasi dan Teknologi*, 8(1), 65–72.
- Santoso, B., & Hidayat, A. (2021). Penerapan analisis sentimen untuk monitoring tren topik di Twitter. *Jurnal Ilmu Komputer*, 9(2), 54–61.
- Saputra, M. (2022). Analisis sentimen produk digital di Twitter dengan Naïve Bayes. *Jurnal Analisis Data dan Komputer*, 5(4), 233–245.
- Simbolon, C. A. D., & Siahaan, C. (2021). Penggunaan komunikasi media sosial Twitter di kalangan remaja di Kecamatan Cibinong, Kabupaten Bogor. *JISIP*.
- Wibowo, A., & Yuliana, R. (2020). Implementasi TextBlob pada analisis sentimen di media sosial. *Jurnal Teknologi Komputer*, 15(2), 78–85.