

ANALISIS SENTIMEN TERHADAP KONTROVERSI PERGANTIAN PELATIH TIMNAS INDONESIA MENGGUNAKAN METODE KNN (K - NEAREST NEIGHBOUR)

Aulia Marcha Artamevia, Jati Sasongko Wibowo

Teknik Informatika, Universitas Stikubank

Jalan Tri Lomba Juang, Mugassari, Kota Semarang, Jawa Tengah 50241

marchartamevia@gmail.com

ABSTRAK

Sepak bola merupakan salah satu cabang olahraga yang memiliki banyak penggemar di berbagai belahan dunia, termasuk Indonesia. Antusiasme masyarakat terhadap Tim Nasional Indonesia sangat tinggi, khususnya ketika terjadi perubahan penting seperti pergantian pelatih. Penelitian ini bertujuan untuk mengevaluasi kecenderungan sentimen publik terhadap pergantian pelatih Timnas Indonesia melalui analisis komentar – komentar di platform YouTube. Fokus permasalahan mencakup tahapan pengumpulan serta pengolahan data komentar, dan bagaimana kecenderungan opini publik terbentuk berdasarkan hasil analisis. Sebanyak 1.000 komentar dikumpulkan dari beberapa video yang membahas topik tersebut. Analisis sentimen dilakukan menggunakan algoritma K-Nearest Neighbour (KNN), melalui tahapan preprocessing seperti pembersihan teks, tokenisasi, penghapusan kata umum (stopword), dan stemming. Komentar kemudian diklasifikasikan ke dalam tiga kategori: sentimen positif, negatif, dan netral. Hasil pengujian menunjukkan bahwa model KNN mencapai akurasi sebesar 68%, dengan precision tertinggi 0,79 dan recall tertinggi 0,90. Nilai rata – rata F1-score makro sebesar 0,60 menunjukkan bahwa performa model tergolong baik, terutama dalam mengidentifikasi sentimen positif dan negatif. Penelitian ini memberikan gambaran kuantitatif mengenai persepsi masyarakat terhadap perubahan pelatih Tim Nasional Indonesia.

Kata kunci : Analisis Sentimen, Sepak Bola, KNN, Timnas Indonesia, Pelatih, YouTube

1. PENDAHULUAN

Sepak bola merupakan salah satu cabang olahraga yang memiliki daya tarik besar secara global, termasuk di Indonesia. Di Tanah Air, sepak bola telah menjadi bagian dari budaya populer yang mempersatukan berbagai lapisan masyarakat. Salah satu bentuk nyata dari antusiasme tersebut terlihat dalam dukungan yang diberikan masyarakat terhadap Tim Nasional Indonesia. Setiap dinamika yang terjadi dalam tubuh tim nasional, terutama peristiwa besar seperti pergantian pelatih, hampir selalu memicu respons publik yang luas, mulai dari dukungan penuh hingga kritik tajam [1]. Kondisi ini semakin menarik perhatian ketika performa tim tidak konsisten, sehingga setiap kebijakan federasi menjadi sorotan masyarakat.

Dalam era digital saat ini, media sosial berperan penting sebagai sarana masyarakat untuk mengekspresikan opini secara terbuka dan masif. YouTube, sebagai salah satu platform yang paling sering digunakan, menyediakan ruang diskusi melalui kolom komentar di berbagai konten seperti konferensi pers, highlight pertandingan, atau pernyataan resmi dari federasi [2]. Ribuan komentar sering kali muncul dalam waktu singkat, yang mencerminkan beragam pandangan publik terhadap keputusan federasi. Namun, tingginya volume data ini menjadi tantangan tersendiri, karena proses analisis secara manual akan memakan waktu dan rentan terhadap bias subjektif.

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan otomatis dan sistematis yang mampu menganalisis opini publik dalam jumlah besar secara efisien dan akurat. Salah satu metode yang

sesuai adalah *sentiment analysis*, yaitu teknik dalam bidang pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang bertujuan mengidentifikasi dan mengklasifikasikan sikap atau emosi dalam teks menjadi kategori tertentu seperti positif, negatif, atau netral [3]. Melalui pendekatan ini, opini publik yang terkandung dalam komentar-komentar dapat dimetakan secara kuantitatif, sehingga memberikan wawasan terhadap persepsi kolektif masyarakat.

Dalam penelitian ini, algoritma *K-Nearest Neighbour* (KNN) digunakan sebagai metode utama untuk klasifikasi sentimen. KNN merupakan salah satu algoritma machine learning berbasis *supervised learning* yang bekerja dengan cara menghitung jarak kemiripan antar data untuk menentukan kelas dari data yang belum diketahui labelnya [4]. Algoritma ini banyak digunakan dalam analisis teks karena kesederhanaannya, efektivitasnya dalam menangani data *non-linear*, serta fleksibilitasnya dalam berbagai skenario klasifikasi. Penelitian ini menerapkan proses *crawling* untuk mengumpulkan komentar YouTube, diikuti dengan tahap *preprocessing* teks, lalu mengubah data ke bentuk numerik dengan teknik TF-IDF sebelum akhirnya diklasifikasikan menggunakan KNN. Melalui pendekatan ini, diharapkan dapat diperoleh pemetaan persepsi publik yang lebih objektif dan dapat dijadikan referensi dalam evaluasi kebijakan oleh pihak federasi.

2. TINJAUAN PUSTAKA

Penelitian ini mengacu pada sejumlah studi sebelumnya yang membahas tentang analisis sentimen yang dilakukan pada data yang diambil dari media

sosial. Dalam hal ini, metode yang diterapkan adalah pengklasifikasian teks, dengan menggunakan algoritma *K-Nearest Neighbour* (KNN).

2.1. Penelitian Terdahulu

Beberapa penelitian sebelumnya telah menunjukkan bahwa algoritma *K-Nearest Neighbour* (KNN) memiliki kinerja yang cukup baik dalam analisis sentimen berbasis data media sosial. KNN termasuk dalam metode pembelajaran terawasi yang bekerja berdasarkan kedekatan antar data dan mampu menangani data non-linear tanpa memerlukan asumsi distribusi awal. Dalam studi yang dilakukan oleh Pamungkas dan Kharisudin, KNN diterapkan untuk mengklasifikasikan sentimen masyarakat Indonesia terhadap pandemi Covid-19 melalui Twitter. Dengan nilai *K* sebesar 20 dan teknik validasi silang 10-Fold *Cross Validation*, algoritma KNN berhasil mencapai akurasi sebesar 62,10% [5].

Puspita dan Widodo juga melakukan penelitian yang membandingkan performa tiga algoritma klasifikasi, yakni *KNN*, *Decision Tree*, dan *Naïve Bayes*, pada data komentar terkait layanan BPJS. Dari 903 data yang dianalisis, metode KNN menunjukkan performa terbaik dengan akurasi mencapai 95,58%, terutama dalam mengklasifikasikan sentimen netral dan negatif yang umum dijumpai dalam isu layanan publik [6].

Sementara itu, studi oleh Dwiki et al. meneliti ulasan pengguna pada dua aplikasi investasi, yaitu Bibit dan Bareksa, menggunakan algoritma KNN dengan pendekatan CRISP-DM. Dengan rasio data pelatihan dan pengujian 60:40, hasil klasifikasi pada aplikasi Bibit menunjukkan akurasi sebesar 85,14%, precision 91,91%, dan recall 76,44%. Sedangkan pada aplikasi Bareksa, diperoleh akurasi 81,70%, precision 87,15%, dan recall 75,73%. Temuan ini menunjukkan bahwa KNN dapat diandalkan dalam klasifikasi sentimen ulasan aplikasi finansial [7].

Hasil dari berbagai studi tersebut memperkuat dasar pemilihan algoritma KNN dalam penelitian ini untuk mengklasifikasikan komentar publik di YouTube terkait pergantian pelatih Tim Nasional Indonesia

2.2. Analisis Sentimen

Analisis sentimen merupakan salah satu sektor dalam data mining yang bertujuan untuk meneliti dan mengumpulkan data teks berkaitan dengan pandangan, penilaian, sikap, perasaan, serta emosi seseorang terhadap produk, individu, organisasi, atau isu tertentu [5]. Dalam penelitian ini, analisis akan dilakukan terhadap tanggapan masyarakat Indonesia di platform YouTube mengenai pergantian pelatih Timnas Indonesia, yang bertujuan untuk mengidentifikasi tiga klasifikasi sentimen: positif, negatif, dan netral

2.3. YouTube

YouTube adalah situs asal Amerika yang digunakan untuk berbagi video. Didirikan pada bulan

Februari tahun 2005 oleh tiga mantan pegawai *PayPal*. Di platform ini, orang-orang dapat mengunggah, menonton, dan membagikan berbagai macam video. Selain itu, platform ini juga menawarkan konten yang dihasilkan oleh pengguna biasa, seperti vlog, video pendek orisinal, dan video untuk tujuan Pendidikan [8].

2.4. Sepak Bola

Secara umum pengertian Sepak bola adalah permainan tim yang dimainkan oleh dua tim berisi sebelas pemain, dengan tujuan mencetak gol ke gawang lawan. Permainan ini menuntut kemampuan fisik, mental, motorik, dan kerjasama tim. Pemain menggunakan kaki untuk mengontrol bola, kecuali penjaga gawang yang boleh menggunakan tangan. Menurut *FIFA (Federation Internationale de Football Association)* mengungkapkan bahwa sepak bola adalah suatu permainan yang menggunakan bola sepak dan dimainkan oleh sebelas pemain dalam satu tim serta dimainkan di atas lapangan rumput atau turf dengan ukuran panjang lapangan 90 – 120 meter dan lebar 45 – 90 meter.

2.5. Text Mining

Beberapa orang menyatakan bahwa text mining adalah kegiatan mengolah teks untuk menemukan informasi yang bermanfaat dalam kumpulan dokumen. Dengan cara ini, kita bisa mendapatkan pola, tren, atau hubungan antar teks. Hasil dari text mining biasanya adalah teks yang mengandung banyak gangguan, sehingga diperlukan proses preprocessing teks [5].

2.6. K-Nearest Neighbour

Algoritma *K – Nearest Neighbor (KNN)* merupakan algoritma yang mengklasifikasikan objek berdasarkan data pelatihan yang paling dekat dengan objek yang diinginkan. *Algoritma K-Nearest Neighbour (KNN)* merupakan metode yang mengklasifikasikan objek berdasarkan data latih yang paling dekat dengan objek tersebut. Perhitungan dilakukan menggunakan Nilai jarak (*euclidean distance*) terhadap masing – masing objek data yang diberikan. Rumus untuk menghitung euclidean distance sebagai berikut [5], [6], [9].

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

2.7. Confusion Matrix

Confusion Matrix menunjukkan hasil prediksi dengan membandingkan nilai sebenarnya dan nilai yang diprediksi oleh model untuk menghasilkan evaluasi seperti di bawah [7].

- Accuracy*: *Accuracy* adalah seberapa dekat nilai yang diprediksi dengan nilai yang sebenarnya.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (2)$$

- b. *Precision*: *Precision* adalah ukuran seberapa baik model dapat memberikan jawaban yang benar kepada pengguna.

$$Precision = \frac{(TP)}{(TP + FP)} \quad (3)$$

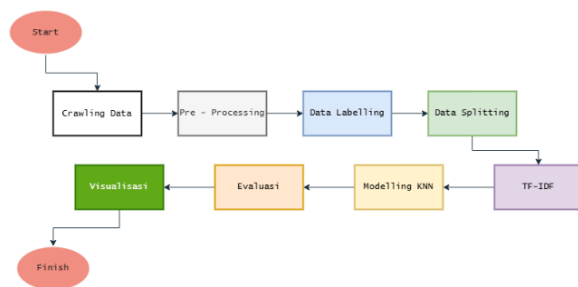
- c. *Recall*: *Recall* adalah tingkat keberhasilan sebuah model.

$$Recall = \frac{(TP)}{(TP + FN)} \quad (4)$$

- d. *F1-Score*: *F1-Score* merupakan hasil dari evaluasi yang membandingkan nilai *precision* dengan nilai *recall*.

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (5)$$

3. METODE PENELITIAN



Gambar 1. Alur penelitian

3.1. Crawling Data

Pada tahap *crawling*, peneliti memanfaatkan *google collab* dengan *python* untuk mengumpulkan data sebanyak 1000 komentar. *Crawling data* merupakan sebuah cara untuk mengumpulkan informasi dari situs web secara otomatis dengan memanfaatkan perangkat lunak. Pendekatan ini memungkinkan pengambilan informasi dari berbagai sumber dengan cara yang efisien dan terorganisir, sehingga dapat dimanfaatkan untuk berbagai tujuan seperti riset, analisis, dan pengembangan aplikasi [10].

3.2. Pre-Processing

Pre-processing data memiliki tujuan untuk mengubah data teks yang tidak terorganisir menjadi bentuk yang teratur. Pada fase ini, dilakukan serangkaian tindakan untuk menghasilkan data yang bersih dan lebih terorganisir guna memperoleh informasi yang lebih tepat. Berikut tahapan pada proses *pre-processing*:

- a. *Data Cleaning*: Tahap pembersihan data dilakukan dengan menghapus informasi tidak bermakna seperti emoji, URL, tanda baca, dan simbol yang

tidak relevan agar data lebih akurat dan sesuai konteks.

- b. *Case Folding*: Tahap *case folding* mengubah seluruh teks menjadi huruf kecil untuk menyeragamkan data.
- c. *Tokenizing*: Tahap *tokenizing* membagi teks menjadi kata-kata untuk memisahkan informasi dalam kalimat.
- d. *Normalization*: Tahap normalisasi memperbaiki kata tidak resmi, singkatan, dan imbuhan agar sesuai dengan kaidah bahasa yang benar
- e. *Stemming*: Tahap *stemming* mengubah kata berimbuhan menjadi bentuk dasar menggunakan pustaka Sastrawi.
- f. *Stopword Removal* atau *Filtering*: Tahap *stopword removal* menghapus kata tidak signifikan seperti “dan” atau “yang” menggunakan kamus NLTK agar teks lebih relevan.

3.3. Data Labelling

Pada fase pelabelan data, atau yang dikenal sebagai data labelling, metode ini diterapkan untuk mengelompokkan suatu sentimen ke dalam kategori negatif, positif, dan netral dengan memanfaatkan teknik klasterisasi K-NN. Selanjutnya, proses pelabelan data akan dilakukan secara manual, di mana kita akan mengidentifikasi tiga kluster yang telah berhasil dikumpulkan ke dalam kategori sentimen negatif, positif, atau netral.

3.4. Data Splitting

Data Splitting merupakan langkah memisahkan dataset menjadi beberapa bagian terpisah untuk keperluan pelatihan serta pengujian model machine learning. Dalam tahap pemisahan data ini, dataset akan dibagi menjadi dua bagian, yaitu data training dan data testing dengan perbandingan 80:20 (80% untuk data training dan 20% untuk data testing). Data training adalah sekumpulan data yang dimanfaatkan untuk melatih model pengelompokan, sedangkan data testing adalah sekumpulan data yang akan diuji menggunakan bantuan dari data training.

3.5. TF-IDF

TF-IDF merupakan suatu metode yang digunakan untuk memberikan nilai pada setiap kata tergantung pada seberapa sering dan seberapa khas kata tersebut muncul di dalam dokumen. Dalam langkah ini, kita akan menerapkan pendekatan TF-IDF (*Term Frequency Inverse Document Frequency*) untuk melaksanakan penilaian terhadap kata. Hasil perhitungan ini akan menghasilkan angka yang semakin tinggi ketika jumlah kemunculan suatu kata dalam sebuah dokumen juga meningkat.

3.6. Modelling K-NN

Pada langkah ini, kumpulan data komentar YouTube yang telah diproses sebelumnya akan dilakukan pemodelan. Untuk proses pemodelan ini, akan diterapkan algoritma *K - Nearest Neighbour*.

3.7. Evaluasi

Evaluasi yang diterapkan pada model KNN dalam menganalisis sentimen terhadap komentar YouTube berkaitan dengan Analisis Sentimen Mengenai Pergantian Pelatih Tim Nasional Sepak Bola Indonesia memanfaatkan *Confusion Matrix*. *Confusion Matrix* akan diterapkan pada langkah ini untuk melakukan penilaian guna memahami performa model yang telah diterapkan dalam menilai nilai akurasi, presisi, *recall*, dan *f1-score*. Berikut ini adalah model *Confusion Matrix*.

Table 1. Model Confusion Matrix

Actual Class Data	Prediction Class Data	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

3.8. Visualization

Visualization merupakan metode untuk menampilkan outcome penilaian model pembelajaran mesin dalam format grafik atau gambar sehingga lebih mudah untuk dimengerti dan dianalisis. Dalam langkah memperlihatkan visualisasi mengenai hasil analisis sentimen yang diperoleh penulis adalah dengan memanfaatkan:

- WordCloud*, membantu memahami arah pandangan masyarakat tanpa harus melihat setiap komentar secara terpisah. untuk memberikan wawasan umum tentang kata-kata yang sering digunakan atau muncul paling banyak.
- Bar Plot*, pada tampilan ini dipergunakan untuk menyampaikan data terkait jumlah perasaan di setiap kategori.
- Pie Chart* (Diagram Lingkaran), menampilkan sentimen (berapa banyak komentar yang bersifat positif, negatif, dan netral). Memahami rasio sentimen dengan cepat dan dengan jelas.

4. HASIL DAN PEMBAHASAN

Dalam studi ini, algoritma *K-Nearest Neighbour* (KNN) diterapkan untuk mengelompokkan pendapat publik terkait Pergantian Pelatih Sepak Bola Timnas Indonesia yang diperoleh dari platform media sosial YouTube. Untuk mencapai klasifikasi yang diinginkan, diperlukan serangkaian tahapan pemrosesan data. Seluruh proses ini dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan *Google Collab*.

4.1. Pengumpulan Data

Pada tahap pengumpulan data, dilakukan proses *crawling web* dengan memanfaatkan *Application Programming Interface (API)* dari YouTube guna memperoleh data komentar yang berkaitan dengan kata kunci "Pergantian Pelatih Sepak bola Timnas Indonesia". Pengolahan data ini dilaksanakan menggunakan bahasa pemrograman *Python* dengan bantuan platform *Google Colaboratory*.

4.2. Dataset

Hasil dari kegiatan pemrograman data *crawling* menunjukkan output seperti yang terlihat di Gambar 2. Tampilan Data Frame untuk dataset keseluruhan mencakup 1000 data dan terdiri dari 5 kolom.

	author	published_at	text	video_id	sentiment
0	@danudandand9218	2025-04-10T00:06:43Z	Saya harapkan pelatih timnas U17 Nova Ariyanto...	MmVcbE3G_4i	NaN
1	@faruqvi270	2025-03-23T03:42:34Z	PSSI HARUS KEMBALIKAN JOBNYA SIN TAE YONG KE T...	MmVcbE3G_4i	NaN
2	@alimudarmad8741	2025-02-09T04:12:23Z	Mat bola bisa piala dunia siapa Shintayong 4th...	MmVcbE3G_4i	NaN
3	@gusrahman7066	2025-01-11T16:11:43Z	TETAP SEMANGAT BUAT PAK ERICK TOHR YG SUDAH M...	MmVcbE3G_4i	NaN
4	@SujOmo-mb2eb	2025-01-10T01:56:31Z	Semoga dengan gantinya pelatih Baru dari Eropa...	MmVcbE3G_4i	NaN
...	...	...	...	...	...
995	@zuffgr9648	2025-01-06T19:15:09Z	Mantab pak Erik Tohir mengambil resiko dari pa...	MmVcbE3G_4i	NaN
996	@Blacksafely	2025-01-06T19:15:08Z	Bung yuke luar biasa	MmVcbE3G_4i	NaN
997	@Indo17845	2025-01-06T19:12:34Z	Klo penggantian Patrick Kluvert sama aja di...	MmVcbE3G_4i	NaN
998	@hiz_kia798	2025-01-06T19:12:14Z	Interpretasi bung yuke paling mendekati kondis...	MmVcbE3G_4i	NaN
999	@wahyuprasetyo7933	2025-01-06T19:12:03Z	Tolong pak ET cari pengganti pelatih shin deng...	MmVcbE3G_4i	NaN

1000 rows x 5 columns

Jumlah baris dan kolom dalam dataset: (1000, 5)

Gambar 2. Tampilan 1000 Dataset

4.3. Pre-processing Data

Tahap ini adalah proses yang dilakukan untuk memastikan dataset lebih terorganisir dan mudah dipahami oleh sistem. Berikut adalah langkah - langkah yang dilakukan dalam Pre-processing data:

- Data Cleaning*: Ini adalah langkah untuk memastikan bahwa informasi yang telah disimpan adalah benar dan akurat.

```
# Import pustaka
import pandas as pd
import re

# Unggah file CSV (jalankan hanya di Google Colab)
from google.colab import files

# Baca file
df = pd.read_csv('crawlingyutubmarcha1.csv')

# Cek jumlah data awal
print("Jumlah data sebelum cleaning:", len(df))

# Drop duplikat dan baris kosong pada kolom 'text'
df.drop_duplicates(subset='text', inplace=True)
df.dropna(subset=['text'], inplace=True)

# Fungsi pembersihan teks
def clean_text(text):
    text = re.sub(r"http\S+", "", text)
    text = re.sub(r"<.*?>", "", text)
    text = re.sub(r'[\W\s]', '', text)
    text = re.sub(r'\d+', '', text)
    text = re.sub(r'[\x00-\x7F]+', '', text)
    text = re.sub(r'\s+', ' ', text).strip()
    return text
```

Gambar 3. Kode Python Data Cleaning

- Case Folding*: Tujuan dari penerapan case folding adalah untuk mengonversi seluruh bentuk data teks menjadi huruf kecil.

```
# Ubah menjadi huruf kecil semua (case folding)
df['casefolding'] = df['cleaned_text'].str.lower()

# Atur agar semua baris dan isi kolom ditampilkan tanpa terpotong
import pandas as pd
pd.set_option('display.max_rows', None) # Menampilkan semua baris
pd.set_option('display.max_columns', None) # Menampilkan semua kolom
pd.set_option('display.max_colwidth', None) # Menampilkan seluruh isi teks dalam sel

# Tampilkan seluruh data pada kolom 'cleaned_text' dan 'casefolding'
df[['cleaned_text', 'casefolding']]
```

Gambar 4. Kode Python Case Folding

- Tokenizing*: Dalam tahapan Tokenizing ini, teks atau kalimat dibagi menjadi kata-kata.

Gambar 5. Kode *Python* Tokenizing

Gambar 6. Kode *Python Normalization*

Gambar 7. Kode *Python Stemming*

Gambar 8. Kode *Python Filtering*

Gambar 9. *Output Pre-Processing Data*

Gambar 10. *Output Data Labelling*

Jumlah data uji : 197

8949

```

=== 10 Kata Paling Umum (IDF rendah) ===
latih      : 2.484
timnas     : 2.696
main       : 3.012
psi        : 3.144
indonesia  : 3.235
nya        : 3.375
yuke       : 3.461
ganti      : 3.476
piala      : 3.539
dunia      : 3.555

=== 10 Kata Paling Langka (IDF tinggi) ===
bisnis     : 6.279
gol         : 6.279
kualifikasi piala : 6.279
otak        : 6.279
rusak       : 6.279
cetak       : 6.567
cetak gol   : 6.567
gaal van    : 6.567
mikir       : 6.567
piala asia  : 6.567

```

Gambar 11. Output Python TF-IDF

4.7. Penerapan K – Nearest Neighbour

Pada fase ini, penerapan algoritma *K-Nearest Neighbor* (KNN) dilaksanakan untuk mengklasifikasikan sentimen komentar yang terdapat dalam dataset yang telah diproses sebelumnya. Algoritma KNN dimanfaatkan untuk klasifikasi dan prediksi baik pada data latih maupun data uji. Di dalam kedua pembagian data tersebut, terdapat tiga kategori sentimen yaitu positif, negatif, dan netral; sehingga penerapan algoritma ini bertujuan untuk memahami karakteristik dari setiap kata yang ada dalam masing-masing kategori sentimen.

Table 2. Hasil Akurasi K- NN

Data Latih : Data Uji	Akurasi K-NN
80 : 20	68%

```

=== Classification Report ===
              precision    recall  f1-score   support

negatif      0.73         0.90         0.80         106
netral        0.51         0.45         0.48          53
positif      0.79         0.39         0.53          38

accuracy          0.68         0.58         0.60         197
macro avg         0.68         0.58         0.60         197
weighted avg      0.68         0.68         0.66         197

```

Akurasi: 0.68

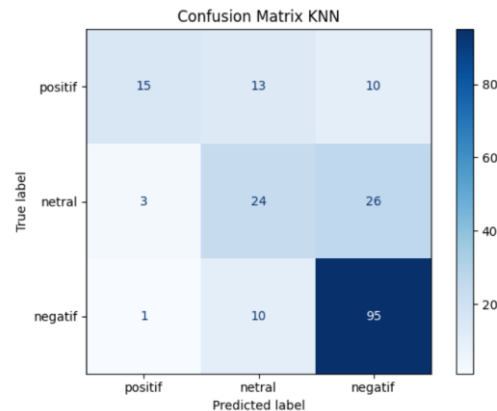
Gambar 12. Output Kode Python Penerapan K-NN

Model yang digunakan saat ini telah mengalami peningkatan akurasi hingga 68%, dengan performa paling baik pada kategori negatif (F1-score mencapai 0.80). Namun, model masih kurang efektif dalam mengenali komentar positif dan netral karena nilai recall yang rendah, yaitu 0.39 untuk positif dan 0.45 untuk netral. Kondisi ini kemungkinan besar disebabkan oleh distribusi data yang tidak seimbang, di mana jumlah komentar negatif jauh lebih banyak dibanding kelas lainnya. Walaupun presisi pada kelas positif cukup tinggi (0.79), banyak komentar positif yang gagal teridentifikasi. Secara umum, model ini dapat dijadikan baseline yang cukup baik, tetapi masih memerlukan sejumlah perbaikan, seperti penyeimbangan data, pengujian model alternatif

seperti SVM atau Logistic Regression, serta pemanfaatan pendekatan berbasis konteks seperti BERT agar analisis sentimen menjadi lebih akurat dan seimbang.

4.8. Confusion Matrix

Confusion Matrix akan diterapkan pada langkah ini untuk melakukan penilaian guna memahami performa model yang telah diterapkan dalam menilai nilai akurasi, presisi, recall, dan f1-score. Berikut ini adalah model *Confusion Matrix*.



Gambar 13. Output Kode Python Confusion Matrix

Dari gambar diatas bisa dilakukan perhitungan manual untuk mengetahui hasil perhitungan presisi, recall, dan f1-score sebagai berikut ini:

$$\text{Accuracy} = \frac{(\text{Jumlah Prediksi Benar})}{(\text{Jumlah Data})} = \frac{(15+24+95)}{(197)} = \frac{(134)}{(197)} \approx 0,6802$$

$$\text{Precision Positive} = \frac{(TP)}{(TP+FP)} = \frac{(15)}{(15+4)} = 0,789$$

$$\text{Precision Negative} = \frac{(TP)}{(TP+FP)} = \frac{(95)}{(95+36)} = 0,725$$

$$\text{Precision Neutral} = \frac{(TP)}{(TP+FP)} = \frac{(24)}{(24+23)} = 0,511$$

$$\text{Recall Positive} = \frac{(TP)}{(TP+FN)} = \frac{(15)}{(15+23)} = 0,395$$

$$\text{Recall Negative} = \frac{(TP)}{(TP+FN)} = \frac{(95)}{(95+11)} = 0,896$$

$$\text{Recall Neutral} = \frac{(TP)}{(TP+FN)} = \frac{(24)}{(24+29)} = 0,453$$

$$\text{F1-Score Positive} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} = 2 \times \frac{(0,789 \times 0,395)}{(0,789 + 0,395)} \approx 0,525$$

$$\begin{aligned}
 F1\text{-Score} &= 2x \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \\
 \text{Negative} &= 2x \frac{(0,725 \times 0,896)}{(0,725 + 0,896)} \\
 &\approx 0,802
 \end{aligned}$$

$$\begin{aligned}
 F1\text{-Score} &= 2x \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \\
 \text{Neutral} &= 2x \frac{(0,511 \times 0,453)}{(0,511 + 0,453)} \\
 &\approx 0,480
 \end{aligned}$$

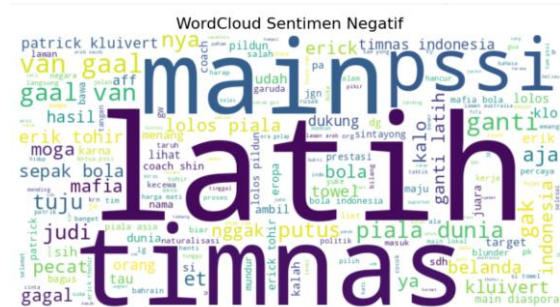
4.9. Visualization

Visualization merupakan metode untuk menampilkan outcome penilaian model pembelajaran mesin dalam format grafik atau gambar sehingga lebih mudah untuk dimengerti dan dianalisis.

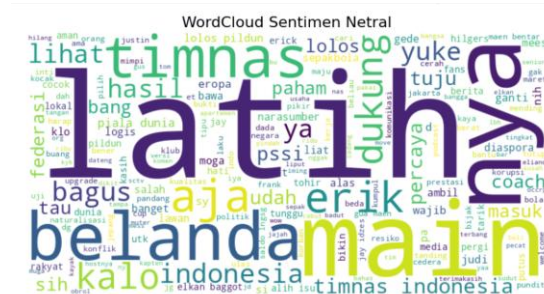
- WordCloud**, membantu memahami arah pandangan masyarakat tanpa harus melihat setiap komentar secara terpisah. untuk memberikan wawasan umum tentang kata – kata yang sering digunakan atau muncul paling banyak.



Gambar 14. WordCloud Sentiment Positive

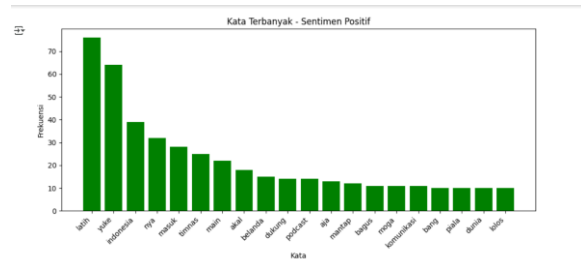


Gambar 15. WordCloud Sentiment Negative



Gambar 16. WordCloud Sentiment Neutral

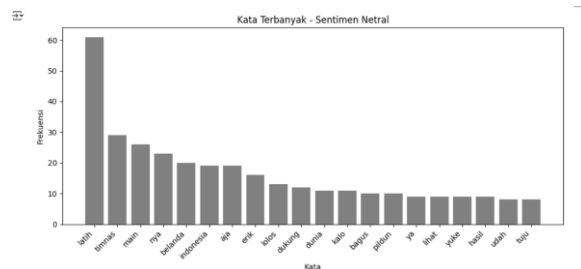
- Bar Plot**, pada tampilan ini dipergunakan untuk menyampaikan data terkait jumlah perasaan di setiap kategori.



Gambar 17. Bar Plot Sentiment Positive

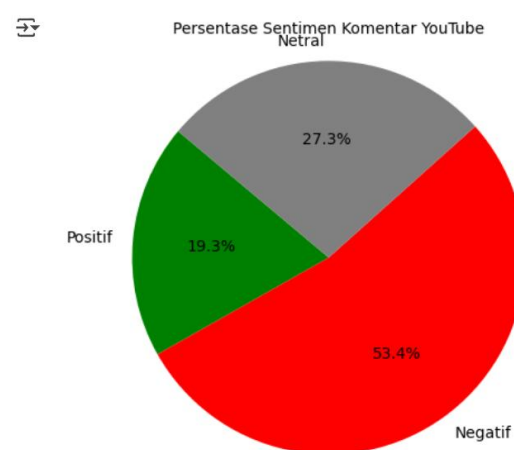


Gambar 18. Bar Plot Sentiment Negative



Gambar 19. Bar Plot Sentiment Neutral

- Pie Chart** (Diagram Lingkaran), menampilkan sentimen (berapa banyak komentar yang bersifat positif, negatif, dan netral). Memahami rasio sentimen dengan cepat dan dengan jelas.



Gambar 20. Output Kode Python Pie Chart

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa penerapan algoritma K-NN dalam sistem analisis sentimen terhadap komentar YouTube mengenai pergantian pelatih Timnas Indonesia telah berhasil dilaksanakan.

Proses pengumpulan data dilakukan melalui YouTube API dan diolah menggunakan *Python* di *Google Colab*, mencakup tahapan seperti *cleaning* teks, normalisasi, tokenisasi, penghapusan kata-kata tidak penting (*stopword*), serta stemming dengan bantuan library Sastrawi. Hasil analisis menunjukkan dominasi komentar negatif, ditandai dengan nilai *f1-score* tertinggi pada kategori negatif (0,80), sementara kategori netral dan positif memiliki tingkat akurasi yang lebih rendah. Temuan ini mengindikasikan kecenderungan publik untuk memberikan tanggapan yang kritis terhadap keputusan tersebut. Ke depan, disarankan agar kamus sentimen diperluas dengan kosakata informal, penggunaan algoritma lain seperti SVM atau LSTM dieksplorasi, serta aspek waktu dan metainformasi turut dipertimbangkan dalam analisis. Selain itu, penyetelan parameter TF-IDF dan nilai K dalam KNN, serta penerapan teknik lanjutan seperti *Named Entity Recognition (NER)* dan lemmatization, diharapkan dapat meningkatkan ketepatan dan pemahaman terhadap konteks sentimen yang dianalisis.

DAFTAR PUSTAKA

- [1] T. Z. Arifin, "EVALUASI KETERAMPILAN TEKNIK MENGOPER BOLA PADA PERMAINAN SEPAK BOLA KLUB MUTIARA RAYA FC BEUREUNUEN KABUPATEN PIDIE," *Jurnal Ilmiah Mahasiswa*, vol. 4, no. 1, 2023.
- [2] D. Ramadhan, D. Yudha Pranata, D. T. Sarwita, S. Bina, and B. Getsempena, "SURVEI KETERAMPILAN DASAR SEPAK BOLA PADA ATLET U 9-11 TAHUN DI LAMBHUK FOOTBALL ACADEMY TAHUN 2020," 2021.
- [3] D. Pranata, "Pengaruh Olahraga Dan Model Latihan Fisik Terhadap Kebugaran Jasmani Remaja: Literature Review," *Jurnal Kesehatan Olahraga*, vol. Vol. 10. No. 02, pp. 107–116, 2022.
- [4] F. A. Mangun and A. Subarkah, "PELATIHAN PENINGKATAN KONDISI FISIK SEBAGAI UPAYA MENUMBUHKAN KESADARAN BEROLAHRAGA PADA MASYARAKAT GANDARIA SELATAN DKI JAKARTA," *Prosiding Seminar Nasional Pengabdian kepada Masyarakat*, no. ISSN 2985-3648, pp. 247–254, 2023, [Online]. Available: <http://journal.unj.ac.id/unj/index.php/snppm>
- [5] F. S. Pamungkas and I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 628–634, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [6] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 646, Dec. 2021, doi: 10.32493/informatika.v5i4.7622.
- [7] A. Dwiki, A. Putra, and S. Juanita, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 2, 2021, [Online]. Available: <http://jurnal.mdp.ac.id>
- [8] R. S. Ainul Wildan, R. Adam Rajagede, and R. Rahmadi, "Analisis Sentimen Politik Berdasarkan Big Data dari Media Sosial Youtube: Sebuah Tinjauan Literatur," 2021.
- [9] B. Laurensz, A. Sentimen, and E. Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19 (Analysis of Public Sentiment on Vaccination in Efforts to Overcome the Covid-19 Pandemic)," 2021.
- [10] S. Algifari Rismawan, Y. Syahidin, P. Piksi Ganesha Bandung, J. Gatot Subroto No, and K. Batununggal, "Implementasi Website Berita Online Menggunakan Metode Crawling Data Dengan Bahasa Pemrograman Python," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 3, pp. 167–178, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>