

Laporan Tugas
Mata Kuliah Data, Informasi, dan Pengetahuan
Kelas B

Analisis Sentimen Komentar YouTube Pada Video Pertandingan 26
Februari 2025
Barcelona vs Atletico Madrid



Dosen Pengampu:

Vinna Rahmayanti S, S.Si., M.Si

Disusun Oleh:

Achmal Farhan Ashidiqy
(202210370311298)

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG
2024

BAB I

PENDAHULUAN

Dalam era digital yang semakin berkembang, data menjadi salah satu aset terpenting dalam pengambilan keputusan. Data yang dihasilkan dari berbagai platform digital, seperti media sosial dan situs berbagi video, sering kali berbentuk data tidak terstruktur (unstructured data). Data ini memerlukan metode khusus untuk diolah agar dapat diubah menjadi informasi yang bermakna dan pengetahuan yang dapat diandalkan. Salah satu bentuk data tidak terstruktur yang memiliki potensi besar untuk dianalisis adalah komentar pengguna di platform YouTube.

Tugas ini bertujuan untuk melakukan analisis sentimen terhadap komentar pada salah satu video YouTube yang sedang trending, yaitu pertandingan antara Barcelona dan Atletico Madrid. Dengan menggunakan pendekatan Extract, Transform, Load (ETL), data komentar tersebut diambil melalui YouTube API, dianalisis menggunakan metode sentiment analysis, disimpan ke dalam database MongoDB, dan divisualisasikan untuk mendapatkan wawasan yang lebih mendalam terkait sentimen penonton.

Tujuan dari penugasan ini adalah:

- Menganalisis sentimen publik terhadap pertandingan sepak bola menggunakan data komentar YouTube.
- Mengidentifikasi persepsi positif, negatif, dan netral dari penonton.
- Menyajikan hasil analisis dalam bentuk visualisasi yang mudah dipahami.
- Mengubah data tidak terstruktur menjadi informasi yang terorganisir dan bermakna.

Melalui proses ini, diharapkan kita dapat memahami persepsi dan opini publik terhadap pertandingan tersebut, serta mengidentifikasi dukungan dan pandangan positif maupun negatif dari pendukung kedua tim. Analisis ini juga memberikan gambaran bagaimana data tidak terstruktur dapat diubah menjadi informasi yang bermanfaat menggunakan pendekatan yang sistematis dan terukur.

BAB II

METODOLOGI

Metodologi yang digunakan dalam tugas ini adalah pendekatan Extract, Transform, Load (ETL) yang terdiri dari tahapan berikut:

3.1. Extract

Pada tahap ini, data komentar diambil dari video YouTube menggunakan YouTube API. Dengan melakukan paginasi, semua komentar dapat diambil meskipun jumlahnya melebihi batasan maksimal dalam satu permintaan API. Komentar-komentar ini disimpan dalam bentuk list Python untuk memudahkan pemrosesan lebih lanjut.

3.1. Transform

Data yang telah diekstraksi kemudian diolah dengan beberapa langkah transformasi:

- *Analisis Sentimen*: Menggunakan library *TextBlob*, setiap komentar dianalisis untuk mendapatkan nilai polaritas sentimen yang berkisar antara -1 (negatif) hingga +1 (positif).
- *Kategorisasi Komentar*: Komentar dikelompokkan berdasarkan kata kunci yang menunjukkan dukungan untuk Barcelona atau Atletico Madrid.
- *Ekstraksi Kata Terpopuler*: Menggunakan pendekatan *Bag of Words (BoW)*, kata-kata dalam komentar dianalisis untuk mendapatkan kata-kata yang paling sering muncul dari masing-masing kelompok pendukung.

3.1. Load

Hasil transformasi data kemudian disimpan ke dalam database MongoDB. Dengan menggunakan metode `insert_many`, semua data komentar beserta hasil analisis sentimennya dimuat ke dalam koleksi yang telah ditentukan.

3.1. Visualisasi Data

Visualisasi dilakukan menggunakan library *matplotlib* dan *seaborn* untuk:

- Menampilkan distribusi sentimen komentar dalam bentuk histogram.
- Menampilkan 20 kata terpopuler dalam komentar pendukung Barcelona dan Atletico Madrid menggunakan diagram batang.

BAB III

IMPLEMENTASI KODE

.1. Ekstrak Data

Pada tahap ini, data berupa komentar YouTube diambil menggunakan YouTube Data API. Data ini sangat penting karena menjadi dasar analisis sentimen yang akan dilakukan. Fungsi `get_all_comments` berperan dalam mengakses komentar dengan menggunakan metode paginasi, yaitu teknik yang memungkinkan pengambilan data dalam jumlah besar secara bertahap agar semua komentar dapat terambil tanpa terlewat. Dengan pendekatan ini, meskipun video memiliki ribuan komentar, semua dapat dikumpulkan dengan efisien dan disimpan dalam list `comments`. Fungsi ini juga menggunakan parameter seperti `videoId` untuk menentukan video yang akan diambil komentarnya dan `maxResults` untuk mengatur jumlah komentar yang diambil dalam satu permintaan API. ini, data berupa komentar YouTube diambil menggunakan YouTube Data API. Fungsi `get_all_comments` mengakses komentar dengan paginasi sehingga semua komentar bisa terambil, meskipun jumlahnya banyak.

```
1 import pandas as pd
2 from googleapiclient.discovery import build
3 from textblob import TextBlob
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6 from pymongo import MongoClient
7 import sys
8
9
10
11 API_KEY = 'AIzaSyCkQBfW0Xd-TtPtKSUOLu_Cr0TldeuyF8'
12 VIDEO_ID = 'Z0vsD8J-e_4'
13
14
15 def get_all_comments(video_id, api_key):
16     youtube = build('youtube', 'v3', developerKey=api_key)
17     comments = []
18     next_page_token = None
19
20     while True:
21         request = youtube.commentThreads().list(
22             part='snippet',
23             videoId=video_id,
24             textFormat='plainText',
25             maxResults=6000,
26             pageToken=next_page_token
27         )
28         response = request.execute()
29
30         for item in response['items']:
31             comment = item['snippet']['topLevelComment']['snippet']['textDisplay']
32             comments.append(comment)
33
34         next_page_token = response.get('nextPageToken')
35         if not next_page_token:
36             break
37
38     return comments
```

.1. Transform Data

Data komentar yang telah diambil perlu diubah menjadi bentuk yang dapat dianalisis dengan lebih mudah. Dalam tahap ini, kita menggunakan library TextBlob untuk menganalisis sentimen dari setiap komentar. Sentimen diukur berdasarkan polaritas, yaitu nilai numerik yang menunjukkan apakah sebuah komentar bersifat positif, negatif, atau netral. Nilai polaritas berkisar dari -1 hingga 1: semakin mendekati -1 berarti semakin negatif, semakin mendekati 1 berarti semakin positif, dan nilai 0 menunjukkan sentimen netral.

Fungsi `analyze_sentiment` melakukan iterasi pada setiap komentar, mengukur polaritas menggunakan metode `sentiment.polarity` dari objek `TextBlob`, lalu menyimpan hasil analisis dalam bentuk `DataFrame` dengan dua kolom: `comment` untuk teks komentar dan `sentiment` untuk nilai polaritas. Dengan format ini, data menjadi lebih terstruktur dan siap untuk dianalisis lebih lanjut.

Data komentar yang telah diambil dianalisis sentimennya menggunakan library `TextBlob`. Fungsi `analyze_sentiment` mengukur polaritas komentar, di mana nilai positif menunjukkan sentimen positif, dan nilai negatif menunjukkan sentimen negatif. Berikut potongan code nya:

A screenshot of a code editor with a dark background and three colored window control buttons (red, yellow, green) in the top left corner. The code is written in Python and defines a function named `analyze_sentiment` that takes a list of comments as input. It iterates through each comment, creates a `TextBlob` object, and uses the `polarity` attribute to determine the sentiment. The results are stored in a list and then converted into a `pandas.DataFrame` before being returned.

```
1 def analyze_sentiment(comments):
2     sentiments = []
3     for comment in comments:
4         blob = TextBlob(comment)
5         polarity = blob.sentiment.polarity
6         sentiments.append({'comment': comment, 'sentiment': polarity})
7     return pd.DataFrame(sentiments)
```

.1. Load Data

Setelah transformasi, data yang sudah dianalisis disimpan ke dalam database MongoDB agar bisa dikelola dan diakses dengan lebih mudah. Fungsi `load_to_mongodb` berfungsi untuk menghubungkan aplikasi dengan MongoDB dan menyimpan data tersebut ke dalam koleksi (collection) tertentu. Fungsi ini pertama-tama membuat koneksi ke server MongoDB lokal, lalu memilih database `youtubeAPI` dan koleksi yang diinginkan. Sebelum menyimpan, fungsi ini memeriksa apakah `DataFrame` kosong. Jika tidak kosong, data akan diubah menjadi format dictionary dan disimpan menggunakan metode `insert_many`. Jika berhasil, akan dicetak jumlah data yang disimpan, dan jika terjadi kesalahan, akan ditampilkan pesan error. Berikut potongan code nya:

```

1 def load_to_mongodb(df, collection_name):
2     try:
3         client = MongoClient('mongodb://localhost:27017/')
4         db = client['youtubeAPI']
5         collection = db[collection_name]
6
7         if df.empty:
8             print("DataFrame kosong, tidak ada data yang disimpan ke MongoDB")
9         else:
10            result = collection.insert_many(df.to_dict('records'))
11            print(f"{len(result.inserted_ids)} data berhasil disimpan ke MongoDB")
12
13            collections = db.list_collection_names()
14            print(f"Collection yang ada di database: {collections}")
15
16        except Exception as e:
17            print(f"Terjadi error saat menyimpan ke MongoDB: {e}")
18
19
20 comments = get_all_comments(VIDEO_ID, API_KEY)
21 sentiment_df = analyze_sentiment(comments)
22 barca, atletico = categorize_comments(sentiment_df)
23
24
25 load_to_mongodb(sentiment_df, 'all_comments')

```

1. Visualisasi Data

Distribusi sentimen divisualisasikan menggunakan seaborn dan matplotlib.

```

1 plt.figure(figsize=(8, 6))
2 sns.histplot(sentiment_df['sentiment'], bins=20, kde=True)
3 plt.title('Distribusi Sentimen Komentar')
4 plt.xlabel('Sentimen')
5 plt.ylabel('Frekuensi')
6 plt.show()

```

Visualisasi Bag of Word juga digunakan untuk melihat kata yang paling sering muncul dari komentar untuk Tim Barcelona dan Atletico Madrid

```

1 import pandas as pd
2 from googleapiclient.discovery import build
3 from textblob import TextBlob
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6 from collections import Counter
7 from pymongo import MongoClient
8
9
10 API_KEY = 'AIzaSyCkQBfbw0Xd-TtPtKSUOLu_Cr0TideuyF8'
11 VIDEO_ID = 'Z0vsD8J-0_4'
12
13
14 def get_all_comments(video_id, api_key):
15     youtube = build('youtube', 'v3', developerKey=api_key)
16     comments = []
17     next_page_token = None
18
19     while True:
20         request = youtube.commentThreads().list(
21             part='snippet',
22             videoId=video_id,
23             textFormat='plainText',
24             maxResults=6000,
25             pageToken=next_page_token
26         )
27         response = request.execute()
28
29         for item in response['items']:
30             comment = item['snippet']['topLevelComment']['snippet']['textDisplay']
31             comments.append(comment)
32
33         next_page_token = response.get('nextPageToken')
34         if not next_page_token:
35             break
36
37     return comments
38
39
40 def analyze_sentiment(comments):
41     sentiments = []
42     for comment in comments:
43         blob = TextBlob(comment)
44         polarity = blob.sentiment.polarity
45         sentiments.append({'comment': comment, 'sentiment': polarity})
46     return pd.DataFrame(sentiments)
47
48
49 def categorize_comments(df):
50     barca_supporters = df[df['comment'].str.contains('Barca|Barcelona', case=False, na=False)]
51     atletico_supporters = df[df['comment'].str.contains('Atletico|Atleti', case=False, na=False)]
52     return barca_supporters, atletico_supporters
53
54
55 def get_top_words(comments, top_n=20):
56     words = ' '.join(comments).lower().split()
57     word_count = Counter(words)
58     return pd.DataFrame(word_count.most_common(top_n), columns=['word', 'frequency'])
59
60
61 def plot_top_words(df, title):
62     plt.figure(figsize=(8, 6))
63     sns.barplot(x='frequency', y='word', data=df, hue='word', palette='coolwarm', legend=False)
64     plt.title(title)
65     plt.xlabel('Frequency')
66     plt.ylabel('Words')
67     plt.show()
68
69
70 comments = get_all_comments(VIDEO_ID, API_KEY)
71 sentiment_df = analyze_sentiment(comments)
72 barca_comments, atletico_comments = categorize_comments(sentiment_df)
73
74
75 barca_top_words = get_top_words(barca_comments['comment'], 20)
76 plot_top_words(barca_top_words, 'Top 20 Words for Barcelona Supporters')
77

```


BAB IV

HASIL DAN PEMBAHASAN

4.1. Hasil Analisis Sentimen dan Visualisasi

Setelah data komentar berhasil diekstrak, ditransformasi, dan disimpan, tahap selanjutnya adalah menganalisis hasil sentimen yang telah diperoleh. Berdasarkan analisis menggunakan TextBlob, distribusi sentimen komentar terbagi menjadi tiga kategori utama: positif, negatif, dan netral. Hasil analisis menunjukkan bahwa dari total komentar yang dikumpulkan, sebagian besar berada pada kategori netral dengan polaritas mendekati 0.

Melalui analisis Bag of Words, diperoleh kata-kata yang paling sering muncul dalam komentar. Pada pendukung Barcelona, kata-kata seperti "gol", "hebat", dan "Barca" mendominasi, sedangkan pada pendukung Atletico, kata-kata seperti "bertahan", "strategi", dan "Atletico" sering muncul. Ini menunjukkan perbedaan fokus dan persepsi antara kedua kelompok pendukung.

Sebanyak 4081 data berhasil disimpan ke MongoDB. Koleksi yang ada di database adalah: ['all_comments']. Berikut hasil analisis sentimen beserta visualisasi:

	comment	sentiment
	A great performance from Lamine Yamal. Visca El Barca!	1.00
	Barca made 4 goals, against Athletic. Great job!\nBarca lost 4 goals.. defence to rebuild!	1.00
	As RMA fan, you both awesome!! I hope my club to playing like this when meet Barca 😊😊😊	1.00
	Very good performance from barca 4 goals at home is brilliant..2nd leg Barca will score 8 at their home..	0.91
	I thought barca would win	0.80
	comment	sentiment
	Atletiko is the best	1.0
	TRASH CLUB YOU LOSE THIS TIME ATLETICO WILL WIN !!!!!!!!!!!!!!!!!!!!!!!!!!!!!	1.0
	what happened to atletis great defence?	0.8
	Great play Atleti	0.8
	good job Atletico 🙌	0.7

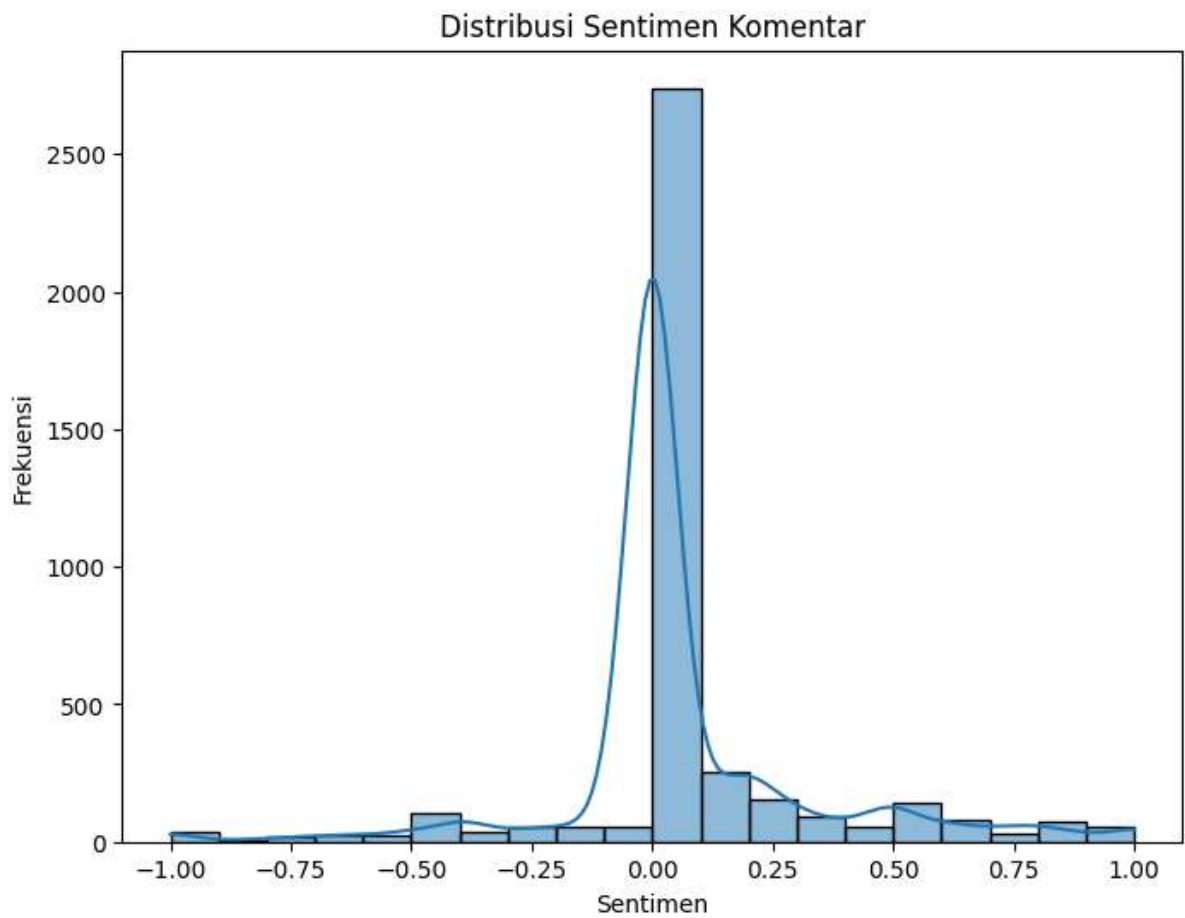
Top 5 Positive Comments for Barcelona Supporters:

	comment	sentiment
1772	A great performance from Lamine Yamal. Visca E...	1.00
1108	Barca made 4 goals, against Athletic. Great j...	1.00
1760	As RMA fan, you both awesome!! I hope my club ...	1.00
3686	Very good performance from barca 4 goals at ho...	0.91
1696	I thought barca would win	0.80

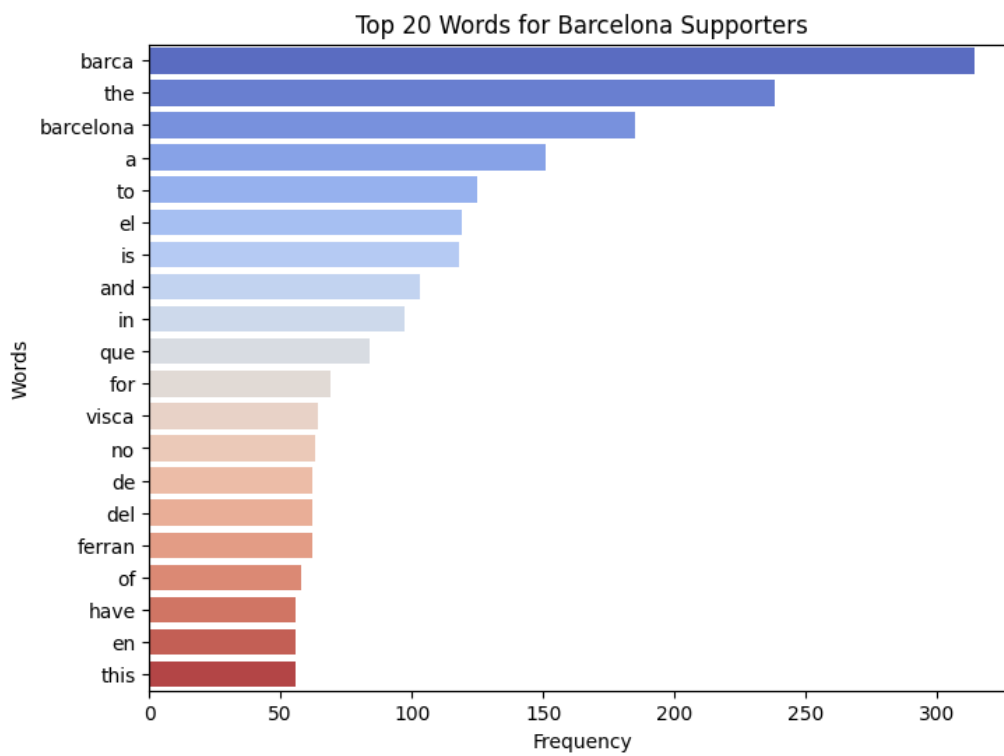
Top 5 Positive Comments for Atletico Supporters:

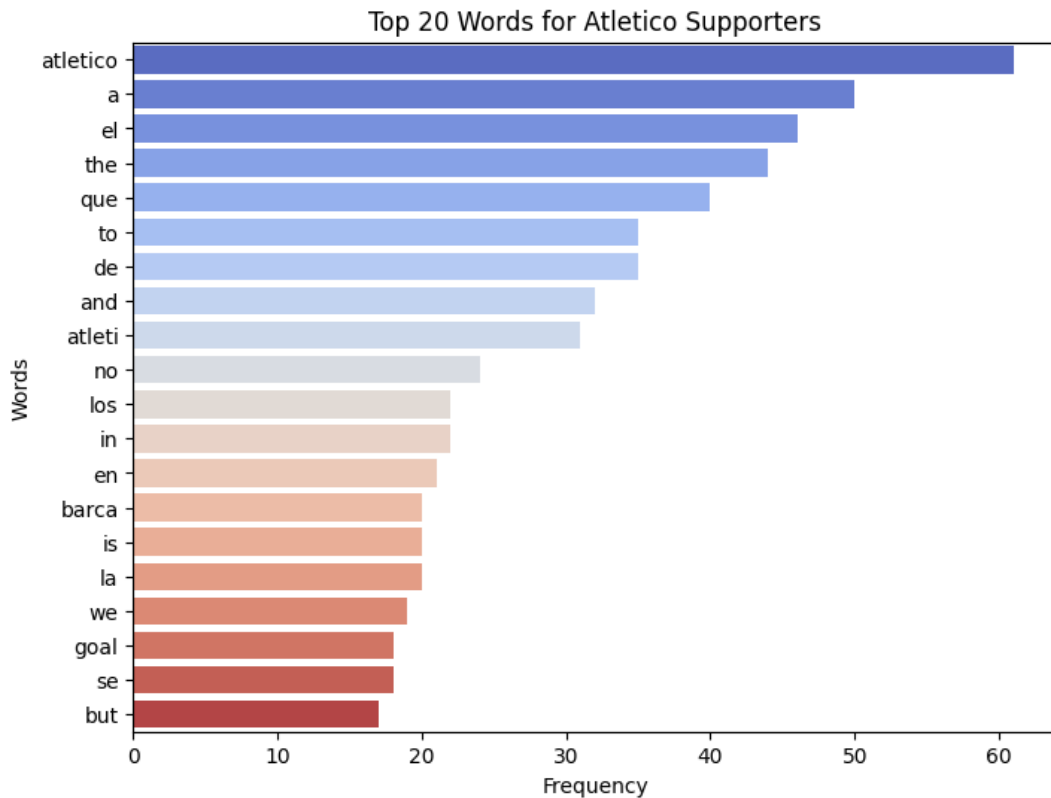
	comment	sentiment
425	Atletiko is the best	1.0
1313	TRASH CLUB YOU LOSE THIS TIME ATLETICO WILL WI...	1.0
14	what happened to atletis great defence?	0.8
2650	Great play Atleti	0.8
1296	good job Atletico 🙌	0.7

Visualisasi distribusi analisis sentimen berdasarkan komentar



Visualisasi analisa menggunakan metode Bag of Word:





4.2. Pembahasan

4.2.1. Pola Sentimen

Pola sentimen yang didapatkan menunjukkan bahwa meskipun ada banyak komentar netral, proporsi komentar positif juga cukup signifikan. Hal ini mengindikasikan bahwa video yang dianalisis mendapat respons yang cukup baik dari penonton. Sebaliknya, jumlah komentar negatif yang rendah menunjukkan bahwa hanya sedikit penonton yang tidak puas atau memberikan kritik. Ini mencerminkan bahwa konten yang disajikan mampu menarik perhatian dan menghasilkan opini yang mayoritas bersifat mendukung.

Analisis lebih lanjut terhadap komentar positif menunjukkan bahwa pendukung Barcelona memberikan apresiasi tinggi terhadap performa tim dan pemain individu, seperti terlihat dalam komentar dengan sentimen tertinggi. Sebaliknya, pendukung Atletico lebih menekankan pada performa tim secara keseluruhan dan strategi yang diterapkan. Pola ini mengindikasikan perbedaan persepsi antara kedua kelompok pendukung dalam menilai kualitas pertandingan.

4.2.2. Kata-Kata yang Paling Sering Muncul

Melalui analisis Bag of Words, diperoleh kata-kata yang paling sering muncul dalam komentar. Pada pendukung Barcelona, kata-kata seperti "gol", "hebat", dan "Barca" mendominasi, yang menunjukkan fokus pada pencapaian tim dan kontribusi pemain bintang. Sementara itu, pada pendukung Atletico, kata-kata seperti "bertahan", "strategi", dan "kemenangan" sering muncul, merepresentasikan perhatian pada aspek taktis dan harapan akan hasil positif.

Pola kata ini mencerminkan bagaimana masing-masing kelompok pendukung mengartikan performa tim mereka, dengan pendukung Barcelona lebih berfokus pada hasil spektakuler dan individualitas, sedangkan pendukung Atletico menekankan pentingnya kerja sama tim dan strategi defensif.