

Analisis Sentimen Berita Piala Dunia di Twitter: Evaluasi Dampak Media melalui NLP

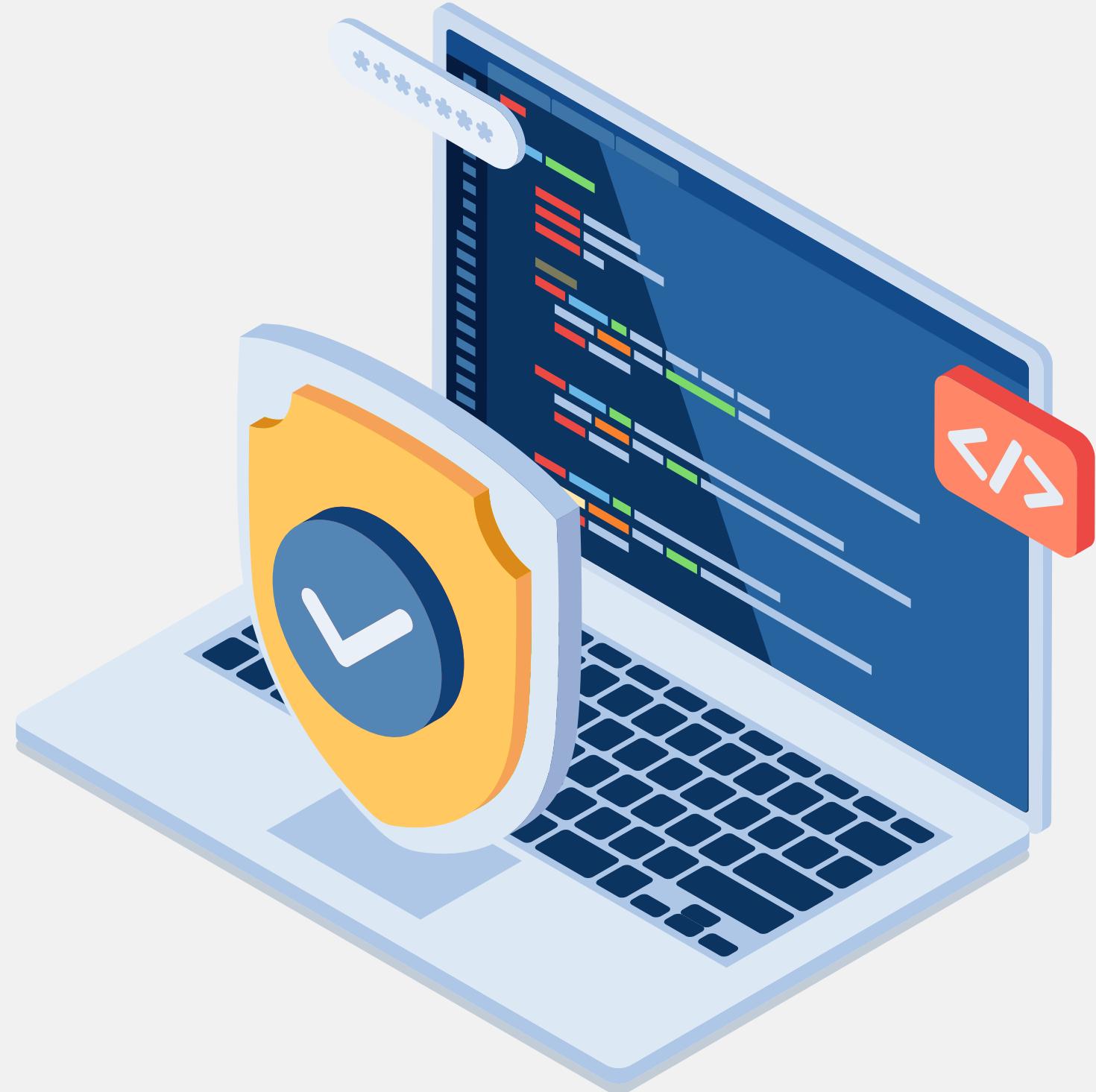


Table of Contents



- 01 *Business Understanding*
- 02 Teknik Pengumpulan Data
- 03 *Data Overview*
- 04 *Data Cleaning & Data Preprocessing*
- 05 *Feature Extraction*
- 06 *Modelling*

Table of Contents



- 07 *Visualization*
- 08 *Insights*
- 09 *Business Recommendation*

Business Understanding



Goal

Background

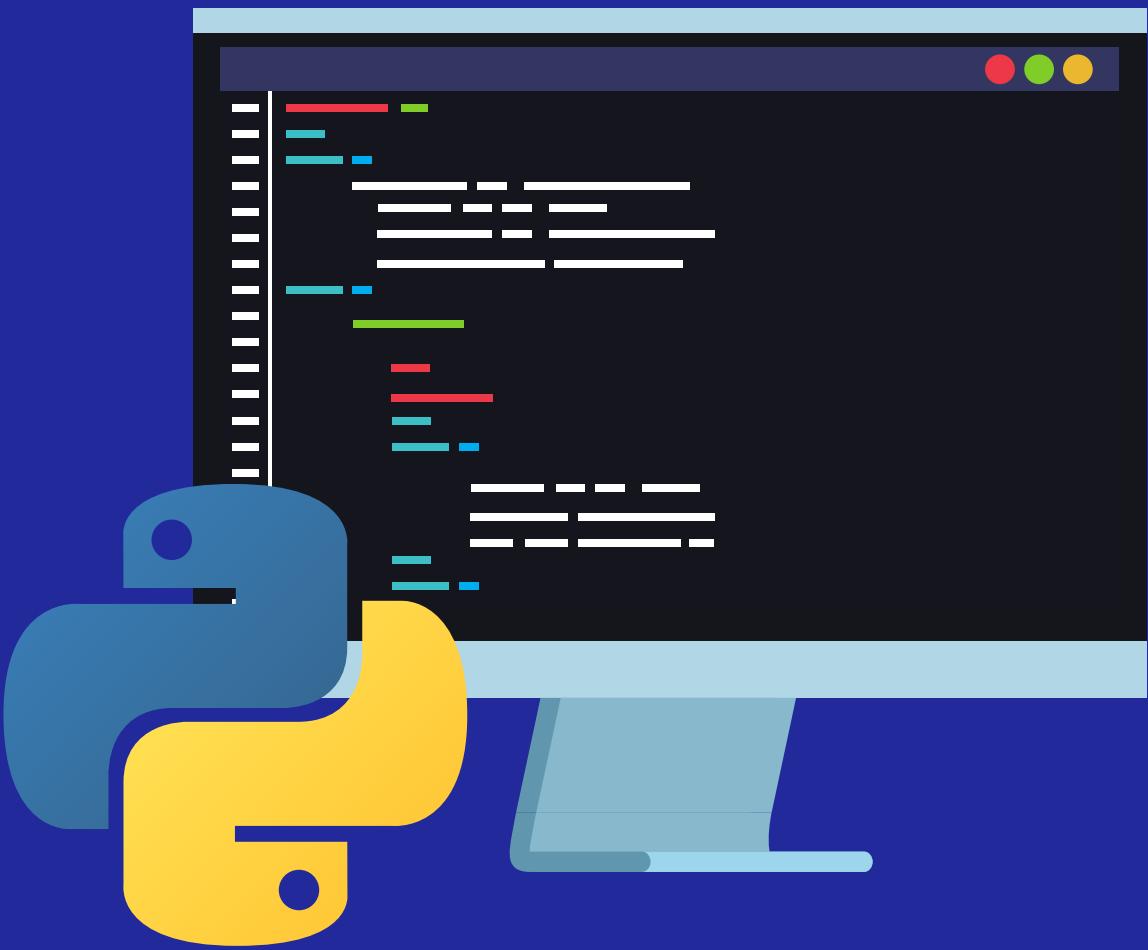
Piala Dunia merupakan acara olahraga terbesar di dunia yang sedang *trend* saat ini, karena Indonesia menjadi tuan rumah untuk Piala Dunia U17. Hal ini menarik perhatian penggemar sepak bola seluruh dunia. Selama acara ini berlangsung, sosial media, seperti platform Twitter menjadi pusat interaksi, dimana pengguna berbagi komentar atau pendapat mengenai peristiwa yang terjadi saat Piala Dunia U17.

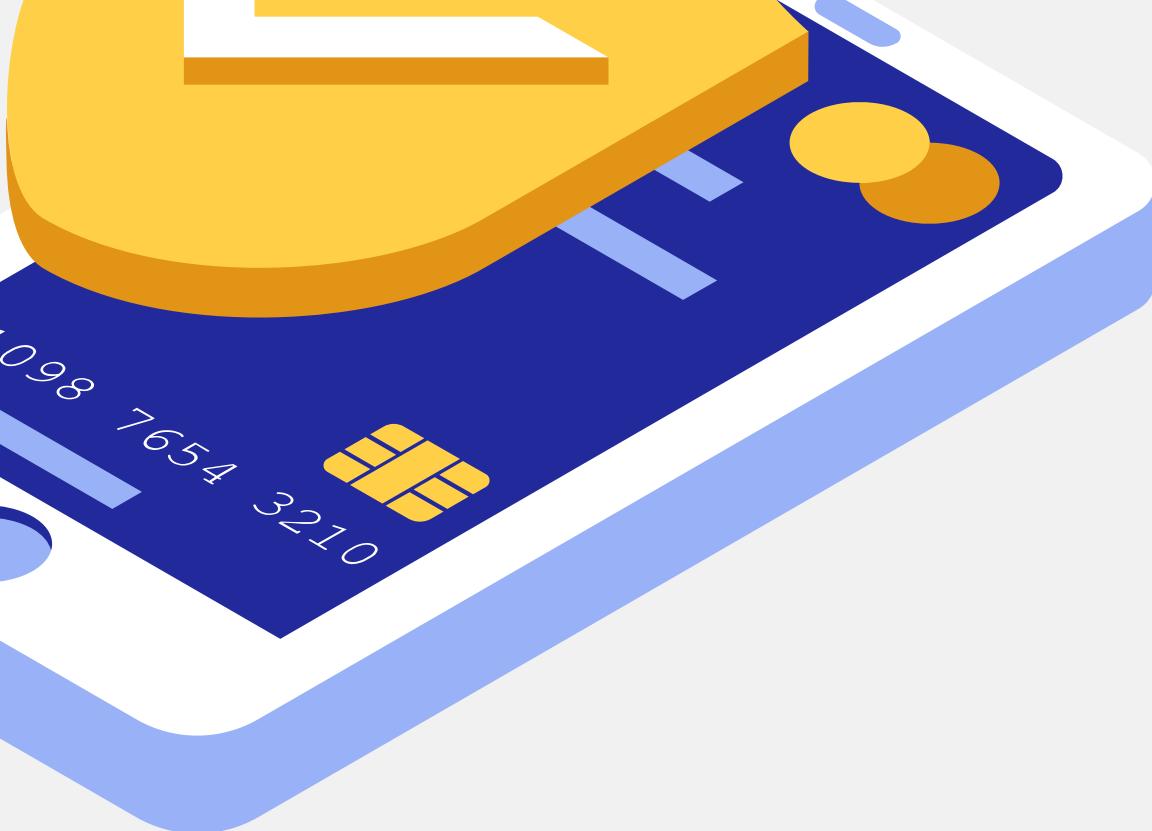
- Memberikan pemahaman tentang bagaimana Piala Dunia U17 mempengaruhi opini dan persepsi masyarakat.
- Mengevaluasi dampak berita terhadap popularitas, dukungan, dan opini yang tercipta.

Objective

- Membuat model *machine learning* yang bisa memprediksi komentar pengguna Twitter terhadap berita Piala Dunia.

Teknik Pengumpulan Data





Data Train

Data di samping diambil dari situs kagle,
<https://www.kaggle.com/datasets/tirendazacademy/fifa-world-cup-2022-tweets>

| Date Created | Number of Likes | Source of Tweet | Tweet | Sentiment |
|------------------------------|-----------------|---------------------|--|-----------|
| 0 2022-11-20 23:59:21+00:00 | 4 | Twitter Web App | What are we drinking today @TucanTribe | neutral |
| 1 2022-11-20 23:59:01+00:00 | 3 | Twitter for iPhone | Amazing @CanadaSoccerEN #WorldCup2022 launch video. Shows | positive |
| 2 2022-11-20 23:58:41+00:00 | 1 | Twitter for iPhone | Worth reading while watching #WorldCup2022 https://t.co/1SQrNa2dYI | positive |
| 3 2022-11-20 23:58:33+00:00 | 1 | Twitter Web App | Golden Maknae shinning bright | positive |
| 4 2022-11-20 23:58:28+00:00 | 0 | Twitter for Android | If the BBC cares so much about human rights, homosexual rights, and w | negative |
| 5 2022-11-20 23:57:32+00:00 | 0 | Twitter for Android | And like, will the mexican fans be able to scream "PUTO" now? Or is tha | negative |
| 6 2022-11-20 23:57:06+00:00 | 0 | Twitter for Android | Look like a only me and the Jamaican football team naw follow up worl | neutral |
| 7 2022-11-20 23:57:05+00:00 | 0 | Twitter for Android | Really? Football on a Monday morning at 9 and 12 and then 3? I need | negative |
| 8 2022-11-20 23:56:10+00:00 | 1 | Twitter for iPhone | As the World Cup starts in Qatar, itâ€™s Black Awareness Day in Brazilâ | positive |
| 9 2022-11-20 23:56:08+00:00 | 0 | Twitter for iPhone | #WorldCup2022 @ITVSport & @LFSYSTEMMUSIC go together so w | positive |
| 10 2022-11-20 23:56:05+00:00 | 1 | Twitter for iPhone | #Qatar tried to help its useless soccer team 5 minutes into the | negative |
| 11 2022-11-20 23:54:44+00:00 | 0 | Twitter for iPhone | What a happy night @socios @Socios_Turkiye | positive |
| 12 2022-11-20 23:54:31+00:00 | 0 | Twitter for iPhone | Letâ€™s go #TeamUSA ðŸ‡ðŸ‡,ðŸ‡ðŸ‡,ðŸ‡ðŸ‡,ðŸ‡ðŸ‡,#WorldCup2022 | positive |
| 13 2022-11-20 23:54:12+00:00 | 5 | Twitter for iPhone | Tyler Adams will have a CAPTAIN moment this #WorldCup2022. A mom | positive |
| 14 2022-11-20 23:53:52+00:00 | 0 | Twitter Web App | This is a good and promising project.should all look forward to the | positive |
| 15 2022-11-20 23:53:44+00:00 | 1 | Twitter for iPhone | What is your World Cup prediction? We have ðŸ‡@ðŸ‡ Å½ber ðŸ‡ðŸ‡ | neutral |
| 16 2022-11-20 23:53:27+00:00 | 0 | IFTTT | RT @Quicktake Qatari officials are set to ban the sale of alcohol within | neutral |
| 17 2022-11-20 23:53:13+00:00 | 2 | Twitter Web App | Nicki Minaj, Maluma & Myriam unite for #WorldCup2022 anthem # | neutral |

Data Predict

Data predict ini digunakan untuk memprediksi akurasi dari model yang sudah dibuat menggunakan *data train*. Proses pengumpulan data ini menggunakan metode *tweet harvest*.



| created_at | id_str | full_text | quote_count | reply_count | retweet_count | favorite_count |
|--------------------------------|-------------|--|-------------|-------------|---------------|----------------|
| Fri Aug 11 09:30:27 +0000 2023 | 1,68993E+18 | Sekali-sekali ngerasain jadi minoritas 🤦 | 5 | 30 | 1 | 9 |
| Fri Aug 11 07:46:02 +0000 2023 | 1,68991E+18 | lucu aja ya, event yang katanya anak pr... | 0 | 1 | 0 | 2 |
| Fri Aug 11 08:12:45 +0000 2023 | 1,68991E+18 | banyak masalah ini jambore... 1. knp ja... | 0 | 1 | 0 | 50 |
| Sat Aug 12 16:19:43 +0000 2023 | 1,6904E+18 | Ntr dibales aje kalau di cibubur | 0 | 0 | 0 | 0 |
| Sat Aug 12 11:24:26 +0000 2023 | 1,69032E+18 | Dah ga heran. ҪҪҪҪҪҪҪҪҪҪ | 0 | 0 | 0 | 0 |
| Fri Aug 11 19:24:00 +0000 2023 | 1,69008E+18 | Udh paling bener di jambore cibubur aj... | 15 | 6 | 51 | 378 |
| Fri Aug 11 15:37:24 +0000 2023 | 1,69002E+18 | Kenapa ga masak sendiri? Kan ada sup... | 15 | 46 | 0 | 1 |
| Sat Aug 12 02:29:41 +0000 2023 | 1,69019E+18 | babinya udah sahadat kali, jadi dianggap | 0 | 1 | 0 | 0 |
| Fri Aug 11 10:22:47 +0000 2023 | 1,68995E+18 | Dimana bumi berpijak jgn kau cba lang... | 2 | 9 | 0 | 0 |
| Fri Aug 11 07:54:47 +0000 2023 | 1,68991E+18 | Biasanya suruh menacari dan masak se... | 0 | 0 | 0 | 2 |
| Fri Aug 11 08:08:13 +0000 2023 | 1,68991E+18 | Besok2 pas jambore dunia ada diindonesia | 2 | 4 | 23 | 202 |
| Fri Aug 11 08:08:09 +0000 2023 | 1,68991E+18 | Makan aja kan darurat | 0 | 8 | 0 | 4 |
| Sat Aug 12 07:37:27 +0000 2023 | 1,69027E+18 | Korea memang racist kepada sesama orang... | 0 | 0 | 0 | 0 |
| Sat Aug 12 13:43:16 +0000 2023 | 1,69036E+18 | Dr komenan Trlihat yg intoletan justru... | 0 | 0 | 0 | 0 |
| Fri Aug 11 07:46:00 +0000 2023 | 1,68991E+18 | jambore itu emg kacau balau acaranya | 0 | 0 | 0 | 6 |
| Sat Aug 12 07:01:23 +0000 2023 | 1,69026E+18 | Yah gimana ya, denger katanya orang k... | 0 | 0 | 0 | 30 |
| Sat Aug 12 02:43:25 +0000 2023 | 1,69019E+18 | Jambore kan di ajarkan survive tong, kc... | 1 | 15 | 0 | 0 |
| Fri Aug 11 07:33:21 +0000 2023 | 1,6899E+18 | Nyammm | 0 | 0 | 0 | 0 |
| Fri Aug 11 09:28:52 +0000 2023 | 1,68993E+18 | Budaya KPop emang makan gtuan n m... | 0 | 0 | 0 | 0 |
| Sat Aug 12 02:18:29 +0000 2023 | 1,69019E+18 | Apa babik? Yaelah nunggu abis napa ng... | 0 | 0 | 0 | 0 |
| Sat Aug 12 11:20:48 +0000 2023 | 1,69032E+18 | entah kenapa beberapa kaum itu susah... | 0 | 0 | 1 | 28 |
| Fri Aug 11 16:52:56 +0000 2023 | 1,69004E+18 | seperti itulah rasanya ketika ibadah ag... | 0 | 6 | 0 | 1 |
| Sat Aug 12 08:00:57 +0000 2023 | 1,69027E+18 | Ini kyknya kmrn di akun mana ya,ada ya... | 0 | 0 | 1 | 11 |
| Sat Aug 12 15:00:47 +0000 2023 | 1,69038E+18 | Mampus korsel kualat dirujak sedunia, | 0 | 0 | 0 | 1 |
| Sat Aug 12 06:32:21 +0000 2023 | 1,69025E+18 | Kalau event international harusnya pre... | 0 | 0 | 5 | 179 |
| Fri Aug 11 07:48:17 +0000 2023 | 1,68991E+18 | Menjadi pelajaran, bahwa negara pela... | 0 | 1 | 1 | 36 |
| Fri Aug 11 09:15:17 +0000 2023 | 1,68993E+18 | Wah tdk masuk syurga dong, tubuh dat... | 0 | 3 | 0 | 0 |
| Sat Aug 12 02:13:31 +0000 2023 | 1,69018E+18 | Gimana ya rasax jd MINORITAS ?? | 0 | 3 | 0 | 0 |
| Sat Aug 12 12:07:32 +0000 2023 | 1,69033E+18 | Orang korea nganggep orang-asia te... | 0 | 0 | 0 | 1 |
| Fri Aug 11 07:51:30 +0000 2023 | 1,68991E+18 | Pramuka kok disediain makanan jadi, n... | 3 | 8 | 2 | 30 |
| Sun Aug 13 06:17:49 +0000 2023 | 1,69061E+18 | Jambore paling gagal wkwwkw... | 0 | 0 | 0 | 0 |
| Sat Aug 12 02:12:58 +0000 2023 | 1,69018E+18 | Tidak salah kalau saya menilai Japan... | 0 | 1 | 0 | 0 |

Data Overview



Data Train

```
df.info()  
[ ] df["Sentiment"].value_counts()
```

<class 'pandas.core.frame.DataFrame'>
Index: 7000 entries, 0 to 6999
Data columns (total 7 columns):
 # Column Non-Null Count Dtype

 0 Unnamed: 0 7000 non-null int64
 1 Date Created 7000 non-null object
 2 Number of Likes 7000 non-null int64
 3 Source of Tweet 7000 non-null object
 4 Tweet 7000 non-null object
 5 Sentiment 7000 non-null object
 6 Indonesian_Text 7000 non-null object
dtypes: int64(2), object(5)
memory usage: 437.5+ KB

```
[ ] df["Sentiment"].value_counts()  
  
Sentiment  
neutral     2710  
positive    2240  
negative    2050  
Name: count, dtype: int64
```

Data train berjumlah 7000 entries yang terdiri dari 2 data bertipe integer dan 5 data bertipe string.

Pada kolom sentiment, menunjukkan bahwa sentimen neutral merupakan sentimen yang paling sering muncul, tetapi dapat disimpulkan bahwa kolom sentimen tidak mengalami imbalance karena jumlah antar setiap sentimen yang berdekatan.



Data Cleaning & Data Preprocessing

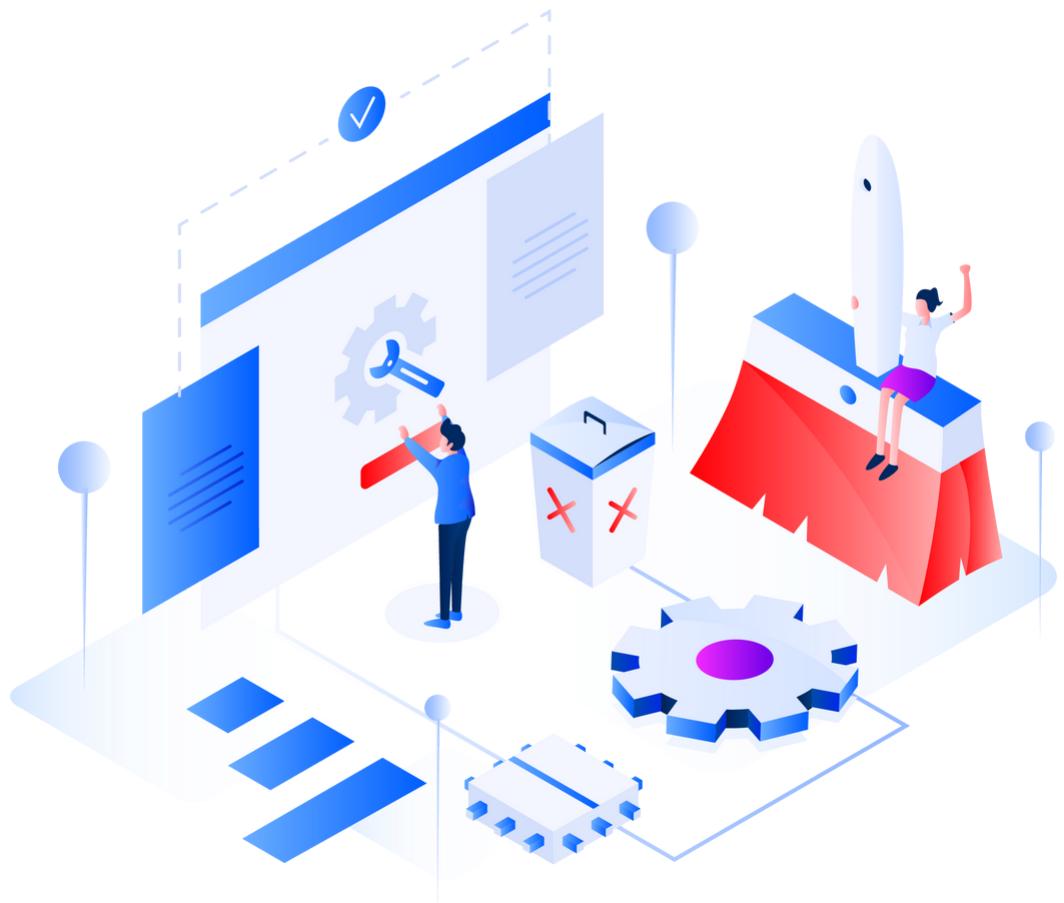


Data Cleaning

```
[ ] df = df.dropna()
```

```
[ ] df = df.drop_duplicates()
```

Kode ini bertujuan untuk membersihkan data, arti membersihkan data yang dimaksud adalah menghapus data-data yang *duplicate* dan data-data yang *missing values*.



Normalization Text

Data Preprocessing

```
▶ import re

def normalize_tweet(tweet):
    # Hapus tanda baca menggunakan regular expression
    tweet_without_punct = re.sub(r'[^w\s]', '', tweet)

    # Hapus emotikon menggunakan regular expression
    tweet_without_emoticons = re.sub(r'[\U0001F600-\U0001F650]', '', tweet_without_punct)

    # Hilangkan nama pengguna Twitter (@username)
    tweet_without_usernames = re.sub(r'@\S+', '', tweet_without_emoticons)

    # Hilangkan hashtag (#tag)
    tweet_without_hashtags = re.sub(r'#\S+', '', tweet_without_usernames)

    # Hilangkan alamat web atau link
    tweet_without_links = re.sub(r'http\S+', '', tweet_without_hashtags)

    # Konversi semua huruf menjadi huruf kecil
    tweet_lowercase = tweet_without_links.lower()

    return tweet_lowercase

df["Indonesian_Text"] = df["Indonesian_Text"].astype(str)
df["Indonesian_Text"] = [normalize_tweet(text) for text in df["Indonesian_Text"]]
```

Definisi

Langkah yang digunakan untuk menyederhanakan atau membersihkan teks supaya lebih mudah dipahami oleh mesin.

Tujuan

Mengatasi variasi dalam bentuk kata.

Teknik yang digunakan

- Pembersihan tanda baca dan karakter khusus
- Penghapusan emoji
- Menghilangkan nama pengguna Twitter
- Menghilangkan hastag (#)
- Penghapusan URL dan referensi
- Lowercasing

Encoding Label

Data Preprocessing

```
[ ] from sklearn.preprocessing import LabelEncoder  
  
# Inisialisasi LabelEncoder  
label_encoder = LabelEncoder()  
  
# Mengatur urutan yang diinginkan  
urutan = ["negative", 'neutral', "positive"]  
  
# Melakukan transformasi data berdasarkan urutan yang ditentukan  
label_encoder.fit(urutan)  
  
# Memetakan teks ke angka sesuai urutan yang telah ditentukan  
df["Sentiment"] = label_encoder.transform(df["Sentiment"])
```

Definisi dan Tujuan

Encoding label adalah proses mengonversi label atau kategori menjadi representasi numerik yang dapat dimengerti oleh *machine learning*.

Metode yang digunakan

Metode yang digunakan untuk encoding label adalah label encoding. Pada metode ini, setiap kategori diberikan label numerik unik bersadarkan urutan. Kategori negative = 0, neutral = 1, dan positive = 2.

Feature Extraction



Stop-words

```
[ ] from nltk.corpus import stopwords  
from string import punctuation  
  
sw = stopwords.words("indonesian") + list(punctuation)
```

Stop-words

Kode di atas membuat sebuah list yang berisi *stopword*.
Stopword adalah kata-kata umum yang sering diabaikan atau tidak memiliki arti penting untuk sentimen analisis.

Tokenization

```
▶ kalimat = [  
    word for word in word_tokenize(tweet.lower()) if word not in sw]  
    for tweet in tqdm(df['Indonesian_Text'])  
]  
kalimat[:5]
```

Tokenization

Kode di atas bertujuan untuk memisahkan suatu kalimat menjadi kata. Jika terdapat kata yang berisi *stopword* maka akan dihilangkan, supaya mempermudah dalam mengambil kata-kata yang memiliki arti penting.

FastText Model

```
[ ] # melatih Fast Text  
  
model_ft = FastText(kalimat, vector_size=300, window=3, min_count=5, epochs= 500, sg=0, hs=0)  
  
[ ] # Menyimpan Model Fast Text  
  
model_ft.save('/Users/nikitaandrea/Documents/LOMBA/GELAR RASA/model_ft.bin')  
  
[ ] ft = model_ft.wv
```

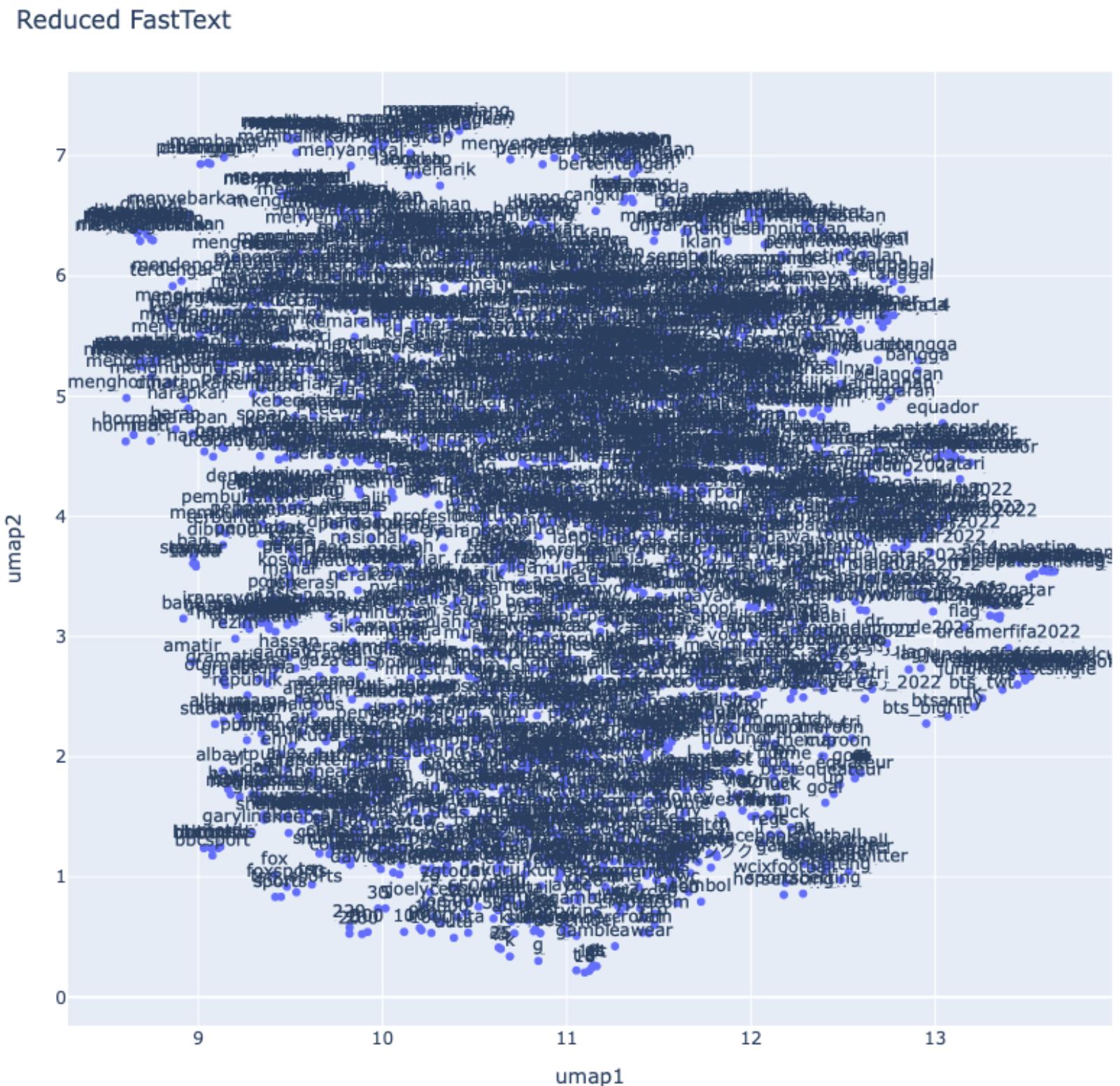
FastText pada NLP bekerja dengan cara membagi kata menjadi subkata atau "n-grams" dan memberikan vektor representasi untuk setiap subkata. Ini memungkinkan model memahami konteks dan makna kata yang lebih baik. Dengan melatih vektor kata untuk subkata, FastText dapat menangkap informasi morfologis dan semantik lebih baik daripada metode tradisional. Hasilnya adalah representasi kata yang kaya dan dapat digunakan untuk tugas NLP seperti klasifikasi teks atau analisis sentimen.



FastText Visualization

Pada visualisasi di samping, setiap titik mewakili satu kata. Kata-kata yang lebih dekat satu sama lain pada grafik menunjukkan bahwa kata-kata tersebut sering muncul bersama dalam tweet

Visualisasi ini dapat digunakan untuk memahami tren dan minat masyarakat Indonesia terhadap Piala Dunia.



Convert Word Vector to Sentence Vector

```
[ ] from collections import Counter
import numpy as np

def get_sentence_vector(sentence, model, weighted=False):
    # Split kalimat menjadi kata-kata
    words = sentence.split()

    # Hitung frekuensi kemunculan kata dalam kalimat
    word_freq = Counter(words)

    # Inisialisasi vektor kalimat
    sentence_vector = [model.wv[word] * word_freq[word] for word in word_freq]

    if weighted:
        # Hitung weighted average vektor kata dalam kalimat
        total_weight = sum(word_freq.values())
        sentence_vector = sum(sentence_vector) / total_weight

    return sentence_vector

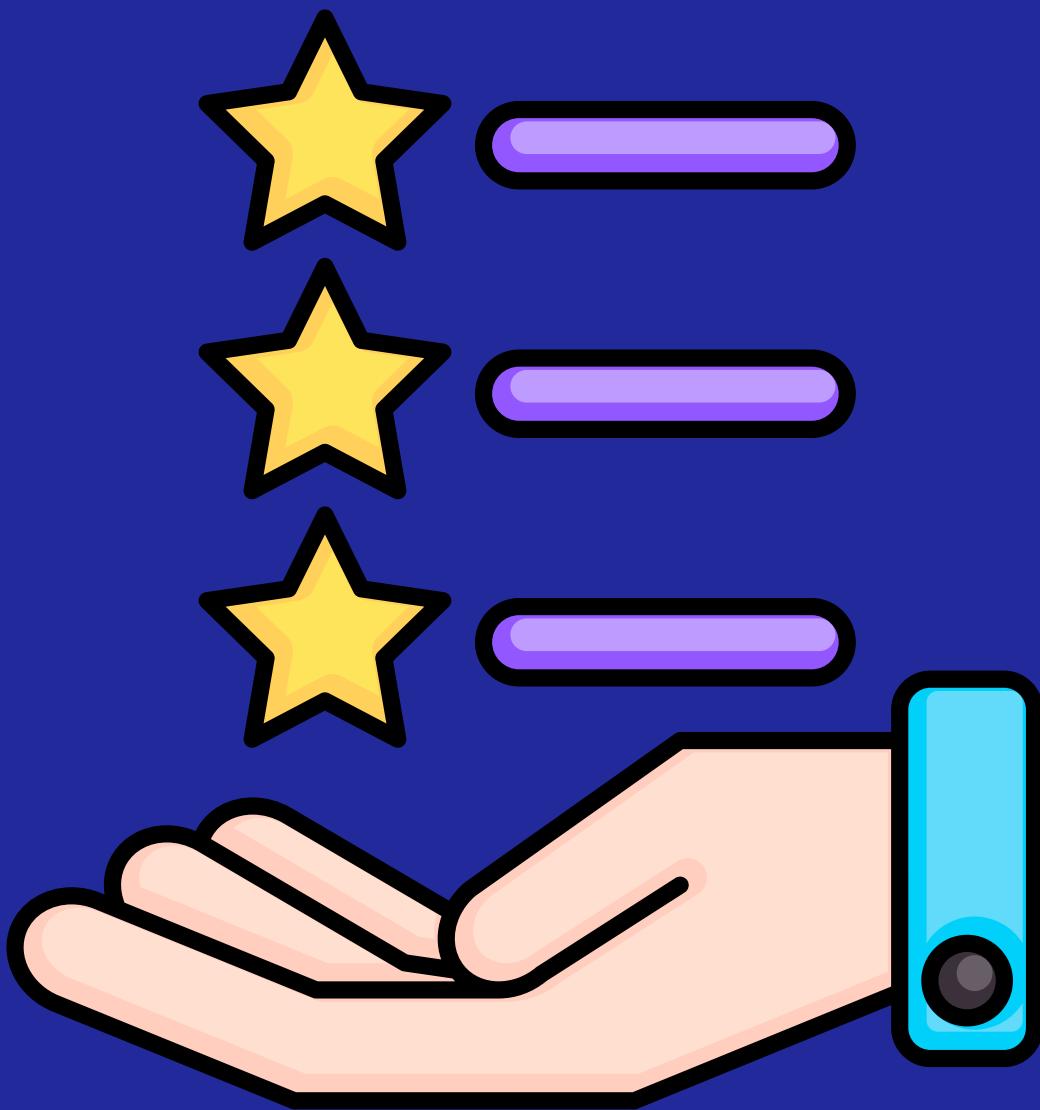
df['Sentence_Vector_Avg'] = df['Indonesian_Text'].apply(lambda x: get_sentence_vector(x, model_ft, weighted=True))
```

Kode ini bertujuan untuk merubah *word vector* yang merupakan *output* dari *FastText model* menjadi *sentence vector*.

Dalam proses ini, *weighted average* digunakan untuk mengonversi vektor kata-kata dari model *FastText* ke vektor kalimat. Bobot (*weight*) setiap kata ditentukan oleh frekuensinya dalam kalimat. Secara singkat, setiap vektor kata dikalikan dengan bobotnya (frekuensi kemunculan dalam kalimat) dan kemudian dijumlahkan untuk mendapatkan vektor kalimat.

Proses ini memberikan representasi kalimat yang mempertimbangkan kontribusi relatif setiap kata, sehingga lebih akurat mencerminkan makna keseluruhan kalimat. Hasilnya adalah vektor kalimat yang dapat digunakan untuk analisis atau pemrosesan lebih lanjut.

Modeling



Langkah Modelling Machine Learning



| | | |
|---|--------|--|
| ✓ | Step 1 | <i>Splitting data train dan data test</i> |
| | | Train : Test = 80 : 20. |
| ✓ | Step 2 | <i>Train Model for Predict</i> |
| | | Melatih model <i>machine learning</i> untuk melakukan prediksi. |
| ✓ | Step 3 | <i>Validation</i> |
| | | Memeriksa apakah model <i>machine learning</i> berjalan dengan baik. |
| ✓ | Step 4 | <i>Predict New Data</i> |
| | | Proses prediksi data baru. |
| ✓ | Step 5 | <i>Prediction</i> |
| | | Hasil prediksi dan contohnya. |

Splitting Data Train & Data Test

```
[ ] from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report
from sklearn.model_selection import train_test_split

X = np.vstack(df['Sentence_Vector_Avg'].values)
y = df['Sentiment']

# Bagi dataset menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Kode ini bertujuan untuk memisahkan *data train* menjadi *data train* dan *data test* untuk proses *validation*.

Train Model for Predict

```
[ ] # Train Model Machine Learning

from xgboost import XGBClassifier

# Definisikan pipeline
pipeline_xgb = Pipeline([
    ('algo', XGBClassifier())
])

# Lakukan RandomizedSearchCV untuk mencari parameter terbaik
model_xgb = RandomizedSearchCV(pipeline_xgb, rsp.xgb_params, cv=4, n_iter=50, scoring='accuracy', n_jobs=-1, verbose=1, random_state=42
model_xgb.fit(X_train, y_train)

# Tampilkan parameter terbaik dan performa model
print(model_xgb.best_params_)
print(model_xgb.score(X_train, y_train), model_xgb.best_score_, model_xgb.score(X_test, y_test))
```

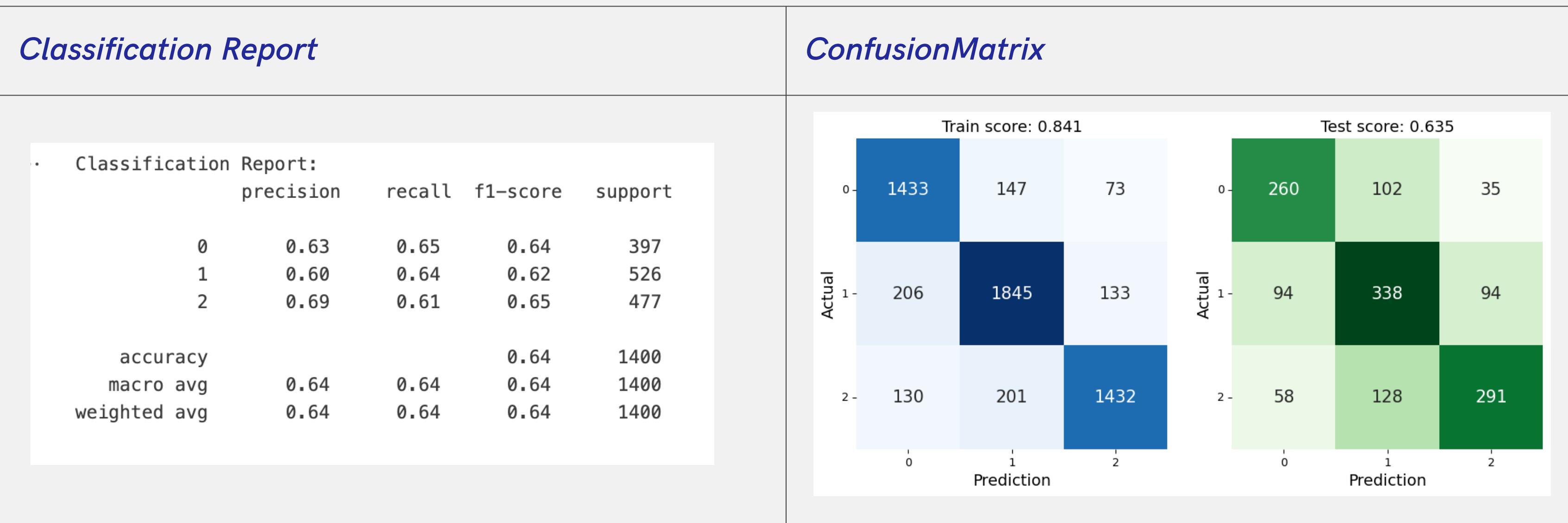
Dalam proses ini, *RandomizedSearchCV* digunakan untuk menentukan parameter optimal untuk model *XGBClassifier*. *XGBClassifier* sendiri merupakan implementasi dari algoritma *Gradient Boosting* yang kuat. *RandomizedSearchCV* membantu mencari kombinasi parameter terbaik dengan cara mencoba sejumlah kombinasi secara acak dari ruang parameter yang didefinisikan.

Singkatnya, *RandomizedSearchCV* mempercepat pencarian parameter terbaik untuk meningkatkan kinerja model *XGBClassifier*, sehingga memastikan bahwa model yang dihasilkan memiliki parameter yang optimal.

Validation

```
[ ] # Classification Report  
  
y_pred = model_xgb.predict(X_test)  
report = classification_report(y_test, y_pred)  
print("Classification Report:\n", report)  
  
[ ] from jcopml.plot import plot_confusion_matrix  
plot_confusion_matrix(X_train, y_train, X_test, y_test, model_xgb)
```

Kode ini memberikan *output* seperti pada tabel di bawah, Berdasarkan *output* yang diperoleh dapat disimpulkan bahwa model berjalan baik dengan akurasi 0.64



Predict New Data

```
[ ] # Load model Fast Text
model_ft = FastText.load('/Users/nikitaandrea/Documents/LOMBA/GELAR RASA/model_ft.bin')

[ ] # Teks baru yang akan diprediksi
new_text = "pemain itu selalu hebat"

normalized_text = normalize_tweet(new_text)

new_text_feature = get_sentence_vector(normalized_text, model_ft, weighted=True)

predicted_sentiment = model_xgb.predict(new_text_feature.reshape(1, -1))
print(f"Prediksi sentimen untuk teks '{new_text}': {predicted_sentiment}")

[ ] # Data for predict

df_crawl = pd.read_csv("/Users/nikitaandrea/Documents/LOMBA/GELAR RASA/pildun-2.csv", sep=';')
df_crawl1 = pd.read_csv("/Users/nikitaandrea/Documents/LOMBA/GELAR RASA/piala-2.csv", sep=';')
df_crawl2 = pd.read_csv("/Users/nikitaandrea/Documents/LOMBA/GELAR RASA/timnasindo.csv", sep=';')
df_crawl3 = pd.read_csv("/Users/nikitaandrea/Documents/LOMBA/GELAR RASA/garuda-2.csv", sep=';')

df_crawl = pd.concat([df_crawl['full_text'], df_crawl1['full_text'], df_crawl2['full_text'], df_crawl3['full_text']], ignore_index=True)
df_crawl = pd.DataFrame(df_crawl).astype(str)

[ ] # Normalisasi Data untuk dimasukkan ke model
df_crawl['text'] = [normalize_tweet(text) for text in df_crawl['full_text']]

df_crawl['text_vec'] = df_crawl['text'].apply(lambda x: get_sentence_vector(x, model_ft, weighted=True))
predicted = np.vstack(df_crawl['text_vec'].values)

[ ] # Prediksi model
predicted_sentiment = model_xgb.predict(predicted)

sentiment = pd.Series(predicted_sentiment)
sentiment = sentiment.replace({1: 'Neutral', 0: 'Negatif', 2: 'Positif'})
sentiment.value_counts()
```

Dalam proses ini, data dikumpulkan, digabungkan, dan dinormalisasi untuk memudahkan interpretasi oleh model *machine learning*. Proses normalisasi memastikan bahwa data memiliki skala yang seragam, sehingga model dapat memberikan prediksi yang lebih akurat.

Setelah tahap ini, data hasil *crawling* digunakan untuk melakukan prediksi oleh model *machine learning*, sehingga dapat memungkinkan analisis yang lebih efektif dan informasi yang lebih berharga.

Prediction

```
pd.set_option('display.max_colwidth', None)

textr = df_crawl['text']
hasil = pd.DataFrame({'text': textr, 'sentiment': sentiment})
hasil.set_index('sentiment', inplace=True)
hasil.sample(5)
```

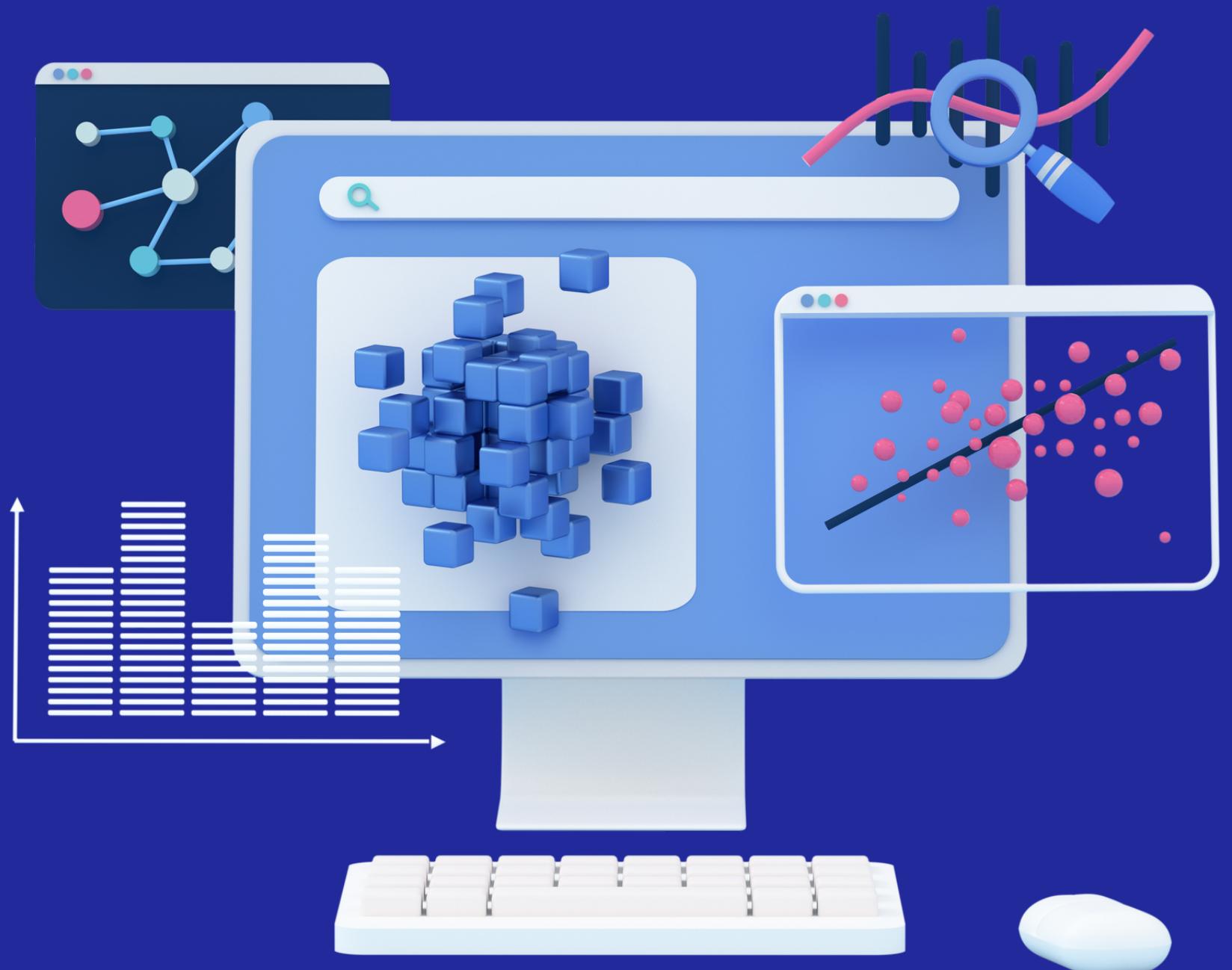
Python

| sentiment | text |
|-----------|--|
| Negatif | piala dunia 1966 menjadi mimpi buruk dalam karier pele legenda brasil itu gagal mempertahankan titel piala dunia di inggris dan pulang dengan kaki pincang |
| Positif | komisi e dprd sumut menyoroti dua proyek pemprov sumut yaitu pembangunan wisma atlet dan lapangan sepakbola di siosar kemudian pembangunan mess di parapat via detiksumut_ |
| Positif | piala dunia 2022 tuntas sudah ada tujuh pemain yang ahli dalam lakukan teknik paling sedikit 10 teknik sukses paling banyak 17 teknik sukses siapa saja juaraboladunia |
| Negatif | mochamad iriawan turut merasakan kekecewaan para pemain karena dibatalkannya indonesia menjadi tuan rumah piala dunia u20 2023 |
| Positif | pelatih timnas indonesia u22 indra sjafri menyambut aturan baru sea games cabang sepakbola tak ada lagi aturan pemain senior khusus untuk pemain 22 tahun |

Kode di atas bertujuan untuk menampilkan prediksi menggunakan model *machine learning* yang telah dibuat dengan menggunakan *data predict*, berdasarkan *output* dari sampel yang diambil dapat disimpulkan bahwa *prediction* benar.



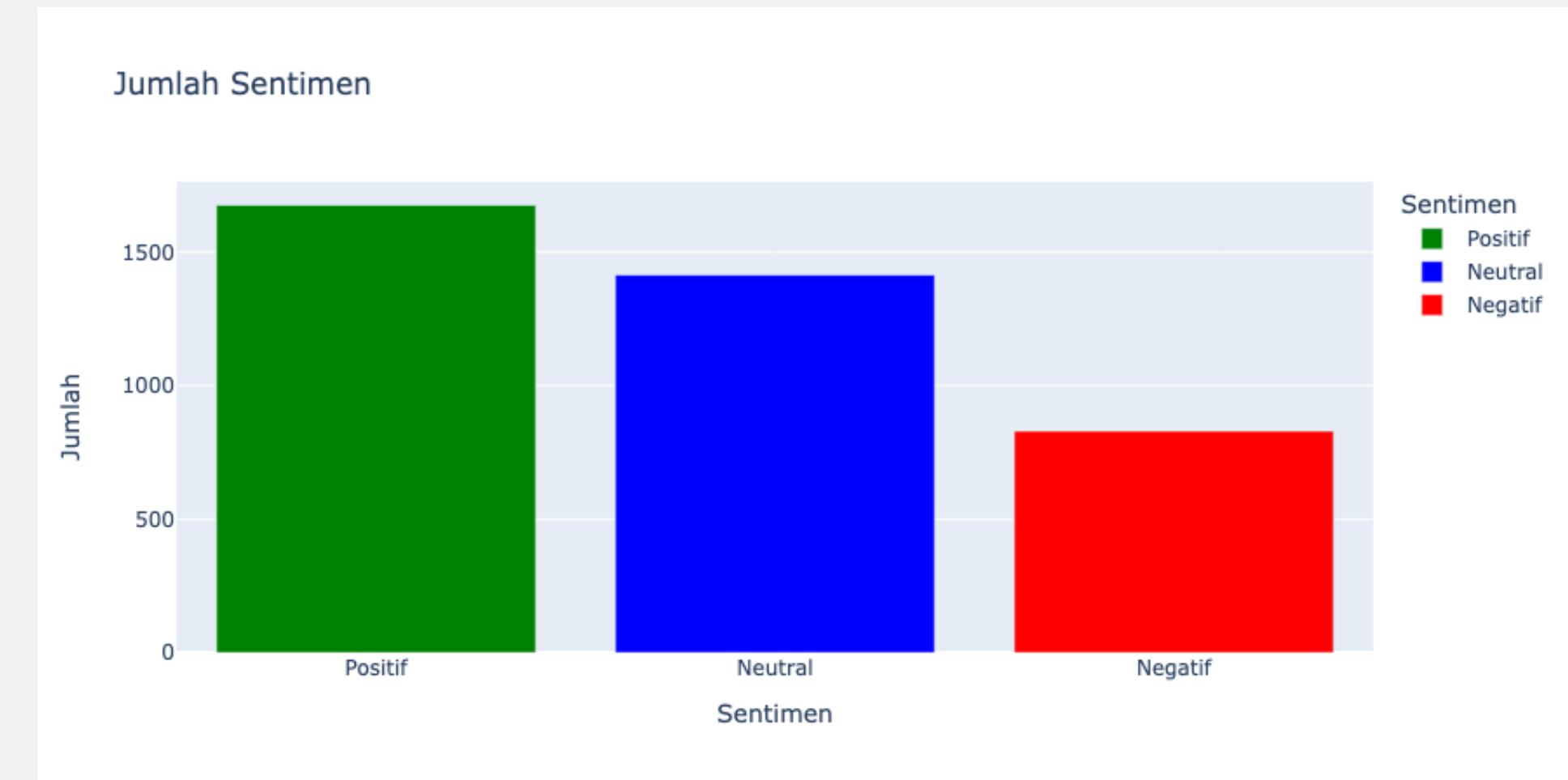
Visualization



Bar Plot

Berdasarkan plot di samping,

dapat disimpulkan bahwa jumlah sentimen positif lebih banyak daripada sentimen neutral dan negatif



Word Cloud

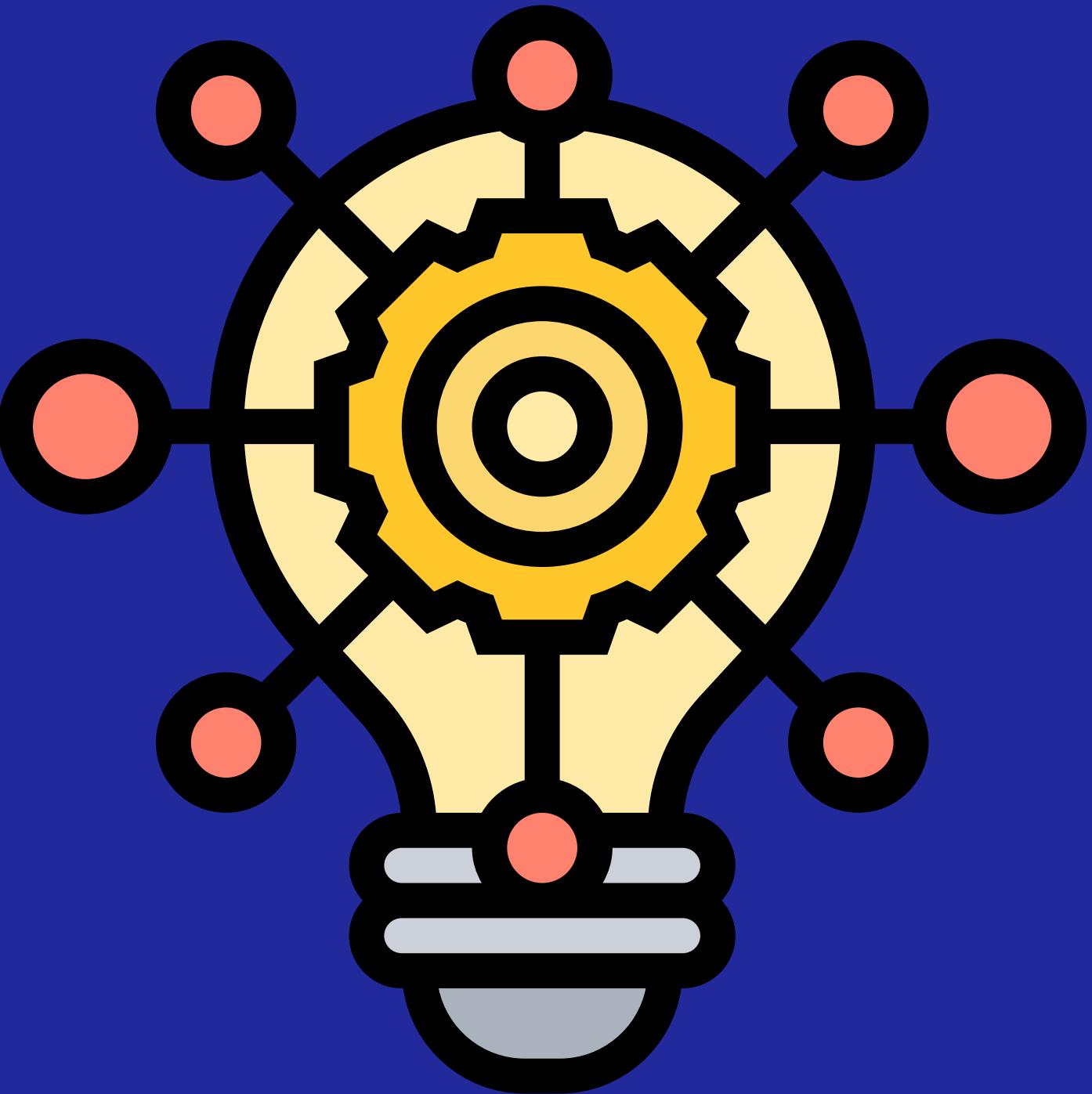
Berdasarkan word cloud di samping,

dapat disimpulkan bahwa kata yang sering muncul pada kalimat yang di prediksi adalah piala dunia.

Word Cloud dari full_text



Insights



Insight

Pengguna Twitter sangat antusias terhadap Piala Dunia.

Pengguna Twitter Indonesia memiliki ketertarikan yang besar terhadap sepak bola.

Indonesia memiliki potensi untuk menjadi tuan rumah *event* sepak bola besar lainnya.

Business Recomendation



Business Recommendation

Pemerintah Indonesia dapat menggunakan insights untuk meningkatkan promosi pariwisata di mata dunia.

PSSI dapat menggunakan insight untuk meningkatkan kualitas pemain.

Penyedia layanan *streaming* olahraga dapat menggunakan insight untuk meningkatkan kualitas konten sepak bola.