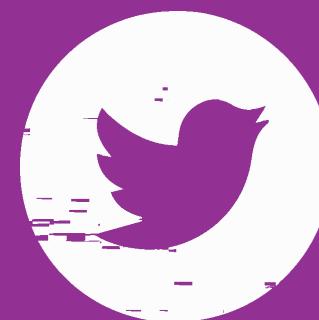


••• API

for Sentiment prediction



Acces API: 127.0.0.1:5000/docs/

API untuk prediksi sentimen komentar twitter di Indonesia



BINAR



HASIL API

Prediksi Sentimen oleh Model MLP Classifier



BINAR

● ● API *for Sentiment prediction*

Form - Input Teks MLP Classifier

Code	Details
200	<p>Response body</p> <pre>{ "data": { "sentiment": "positive", "text": "Terima kasih kepada para pendiri bangsa" }, "description": "Hasil Prediksi Sentiment dengan Model MLP Classifier melalui Form", "status_code": 200 }</pre> <p>Response headers</p> <pre>connection: close content-length: 183 content-type: application/json date: Sun 01 Sep 2024 20:47:53 GMT server: Werkzeug/3.0.3 Python/3.11.7</pre> <p>Responses</p>

Positif

Code	Details
200	<p>Response body</p> <pre>{ "data": { "sentiment": "negative", "text": "BABY ANJING SEMUA" }, "description": "Hasil Prediksi Sentiment dengan Model MLP Classifier melalui Form", "status_code": 200 }</pre> <p>Response headers</p> <pre>connection: close content-length: 161 content-type: application/json date: Sun 01 Sep 2024 20:49:24 GMT server: Werkzeug/3.0.3 Python/3.11.7</pre>

Negatif

● ● API *for Sentiment prediction*

Form - Input Teks MLP Classifier

Code	Details
200	<p>Response body</p> <pre>{ "data": { "sentiment": "neutral", "text": "jadilah pemimpin bangsa" }, "description": "Hasil Prediksi Sentiment dengan Model MLP Classifier melalui Form", "status_code": 200 }</pre> <p>Response headers</p> <pre>connection: close content-length: 166 content-type: application/json date: Sun 01 Sep 2024 20:50:50 GMT server: Werkzeug/3.0.3 Python/3.11.7</pre> <p>Copy Download</p>

Netral

● ● ● API *for Sentiment prediction*

Upload File MLP Classifier

Curl

```
curl -X POST "http://127.0.0.1:5000/mlp_file" -H "accept: application/json" -H "Content-Type: multipart/form-data" -F "file=@data.csv;type=text/csv"
```

Request URL

```
http://127.0.0.1:5000/mlp_file
```

Server response

Code Details

200

Response body

```
{ "data": [ { "sentiment": "positive", "text": "- disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego ! ! !"}, { "sentiment": "neutral", "text": "RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama siapa noh licew juga"}, { "sentiment": "positive", "text": "41. Kadang aku berfikir, kenapa aku tetap percaya pada Tuhan padahal aku selalu jatuh berkali-kali. Kadang aku merasa Tuhan itu meninggalkan aku sendirian. Ketika orangtuaku berencana berpisah, ketika kakakku lebih memilih jadi Kristen. Ketika aku anak ter"}, { "sentiment": "positive", "text": "USER USER AKU ITU AKU\\n\\nKU TAU MATAMU SIPIT TAPI DILIAH DARI MANA ITU AKU"}, { "sentiment": "negative", "text": "USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah"}, { "sentiment": "negative", "text": "USER Ya bani taplak dkk \\\xf0\\\x9f\\\x98\\\x84\\\xf0\\\x9f\\\x98\\\x84\\\xf0\\\x9f\\\x98\\\x84"}, { } ] }
```

Download

Response headers

```
connection: close
content-length: 1983350
content-type: application/json
date: Mon 02 Sep 2024 07:24:46 GMT
server: Werkzeug/3.0.3 Python/3.11.7
```

- API berhasil memprediksi sentiment di setiap baris data yang dimasukkan
- Sentiment diprediksi menggunakan model MLP Classifier



HASIL API

Prediksi Sentimen oleh Model LSTM

••• Hasil Prediksi API

Form LSTM - Input Text

Code Details

200 Response body

```
{  
  "data": {  
    "sentiment": "positive",  
    "text": "alhamdulillah baik syukur"  
  },  
  "description": "Hasil Prediksi Sentiment dengan Model LSTM melalui Form",  
  "status_code": 200  
}
```

Response headers

```
connection: close  
content-length: 187  
content-type: application/json  
date: Sun01 Sep 2024 10:44:22 GMT  
server: Werkzeug/3.0.3 Python/3.11.7
```

Positif

200 Response body

```
{  
  "data": {  
    "sentiment": "positive",  
    "text": "partai demokrat sejahtera demokrasi"  
  },  
  "description": "Hasil Prediksi Sentiment dengan Model LSTM melalui Form",  
  "status_code": 200  
}
```

Response headers

```
connection: close  
content-length: 169  
content-type: application/json  
date: Mon02 Sep 2024 08:07:56 GMT  
server: Werkzeug/3.0.3 Python/3.11.7
```

Negatif

••• Hasil Prediksi API

Form LSTM - Input Text

```
200
Response body
{
  "data": {
    "sentiment": "positive",
    "text": "ANJING BABI KALIAN SEMUAA!!"
  },
  "description": "Hasil Prediksi Sentiment dengan Model LSTM melalui Form",
  "status_code": 200
}
Download

Response headers
connection: close
content-length: 161
content-type: application/json
date: Mon 02 Sep 2024 08:09:00 GMT
server: Werkzeug/3.0.3 Python/3.11.7
```

Model masih belum bisa memprediksi secara akurat dan memberikan prediksi yang lebih banyak sentiment positif, bahkan untuk data yang jelas secara logika itu merupakan teks dengan sentiment negatif



● ● ● API *for Sentiment prediction*

Upload File LSTM

Request URL
`http://127.0.0.1:5000/lstm_file`

Server response

Code Details

200 Response body

```
{ "data": [ { "sentiment": "positive", "text": "- disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego ! ! !"}, { "sentiment": "positive", "text": "RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama siapa noh licew juga"}, { "sentiment": "negative", "text": "41. Kadang aku berfikir, kenapa aku tetap percaya pada Tuhan padahal aku selalu jatuh berkali-kali. Kadang aku merasa Tuhan itu meninggalkan aku sendirian. Ketika orangtuaku berencana berpisah, ketika kakaku lebih memilih jadi Kristen. Ketika anak ter"}, { "sentiment": "negative", "text": "USER USER AKU ITU AKU\\n\\nKU TAU MATAMU SIPIT TAPI DILIAT DARI MANA ITU AKU"}, { "sentiment": "positive", "text": "USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah"}, { "sentiment": "negative", "text": "USER Ya bani taplak dkk \\\xf0\\\x9f\\\x98\\\x84\\\xf0\\\x9f\\\x98\\\x84\\\xf0\\\x9f\\\x98\\\x84"} ]}
```

Response headers

```
connection: close
content-length: 1657101
content-type: application/json
date: Mon 02 Sep 2024 07:57:58 GMT
server: Werkzeug/3.0.3 Python/3.11.7
```

Responses

- API berhasil memprediksi sentiment di setiap baris data yang dimasukkan namun masih cukup banyak prediksi yang kurang akurat
- Sentiment diprediksi menggunakan model LSTM



Analisis Sentimen dari platform Twitter dengan Neural Network dan LSTM

Memuat prediksi sentimen dengan
menggunakan model Multilayer Perceptron
Classifier dan LSTM



●●● Sosial Media di Indonesia



#4

Indonesia merupakan negara dengan pengguna twitter terbanyak ke-4 di dunia pada Juli 2023 yaitu sebanyak 25,45 Juta pengguna (katadata.id)



••• Sosial Media di Indonesia

- Pengguna Twitter di Indonesia tergolong aktif dalam berkomentar, tidak jarang pengguna twitter melakukan ujaran kebencian dengan tujuan merendahkan dan mengolok - olok suatu individu, kelompok, gender, agama, dan ras.
- Biasanya pengguna twitter Indonesia juga sering membahas isu - isu yang ramai diperbincangkan seperti politik, agama, tren, produk dan lain - lain.



Netral

Negatif

Positif

Apa yang bisa kita lakukan?

ANALISIS SENTIMENT

Untuk mengetahui kecenderungan repons pengguna social media Twitter menjadi beberapa kategori



BINAR

● ● ● Manfaat Analisis Sentiment

Bagi Perusahaan

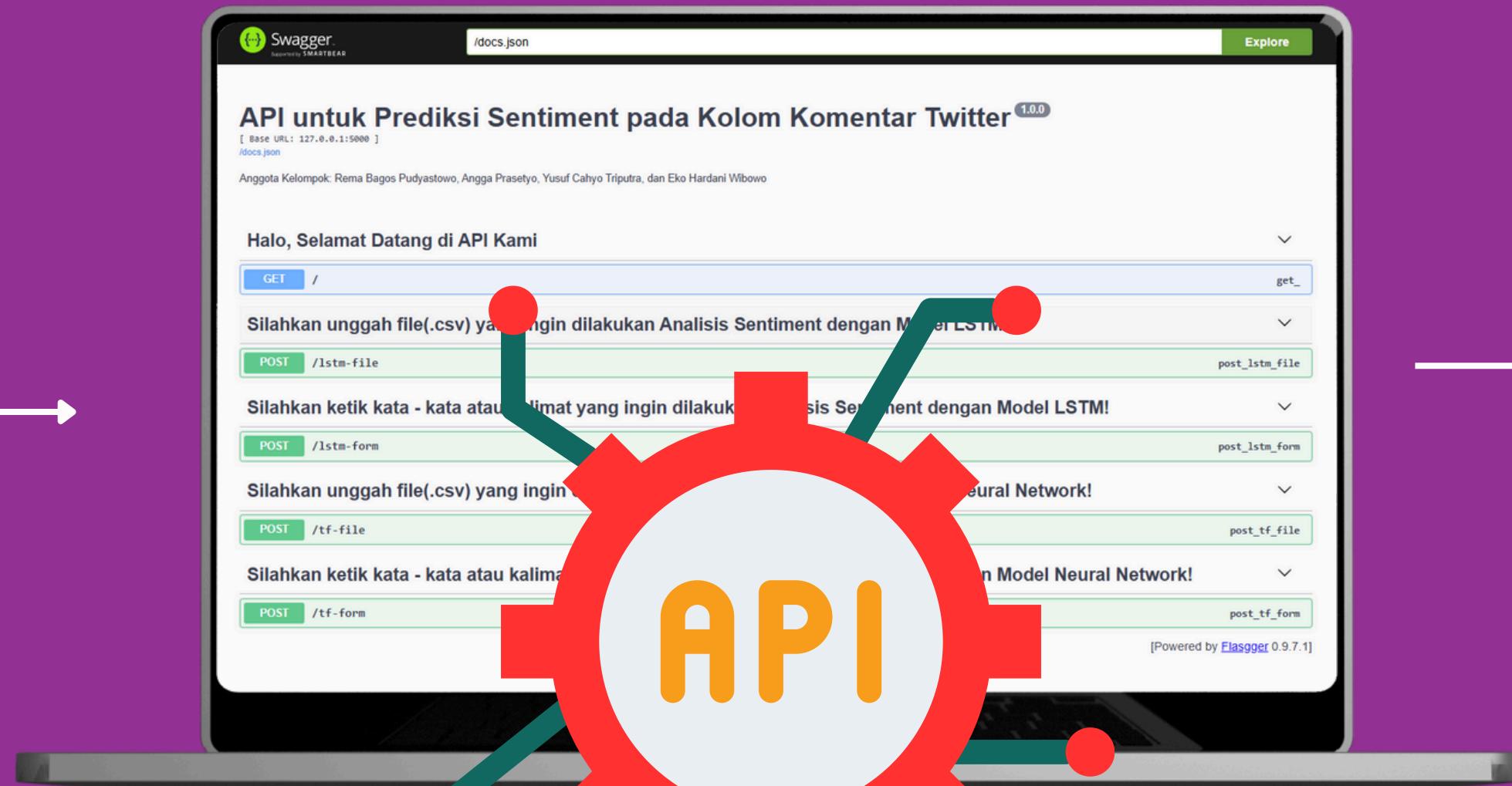
- Untuk membaca trend yang sedang hangat diperbincangkan untuk pengembangan produk
- Menentukan strategi pemasaran

Bagi Pemerintah

- Menjadi dasar pembuatan kebijakan baru dan evaluasi dari kebijakan yang lama
- Menggambarkan pandangan masyarakat terkait kebijakan pemerintah yang baru



API Sentiment Analysis



Positif

Netral

Negatif

ANGGOTA KELOMPOK

Kelompok kami terdiri dari 4 anggota sebagai berikut.



Rema Bagos P.
Ketua



Angga Prasetyo
Anggota



Eko Hardani W.
Anggota



Yusuf Cahyo T.
Anggota



METODE

Model untuk analisis sentimen

Tahap





Data yang digunakan

Data Model

**data_preprocess
ed.txt**



Data Tambahan

**kamusalay.csv
stopwords.csv**



Data Uji API dan Model

data.csv





Tools yang digunakan



Swagger™

API and Data
Processsing



RegEx



Model Building

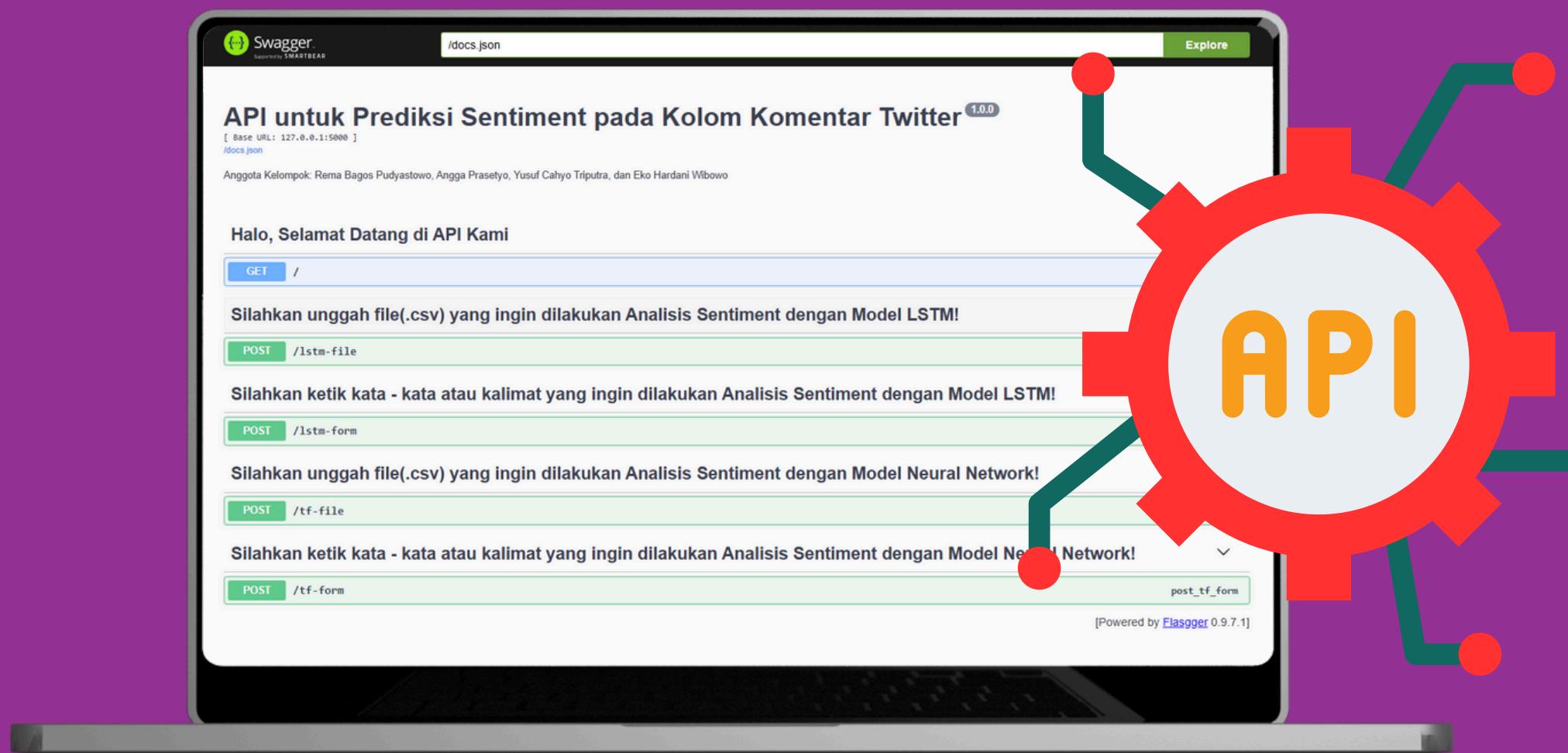


MODEL

Model untuk analisis sentimen



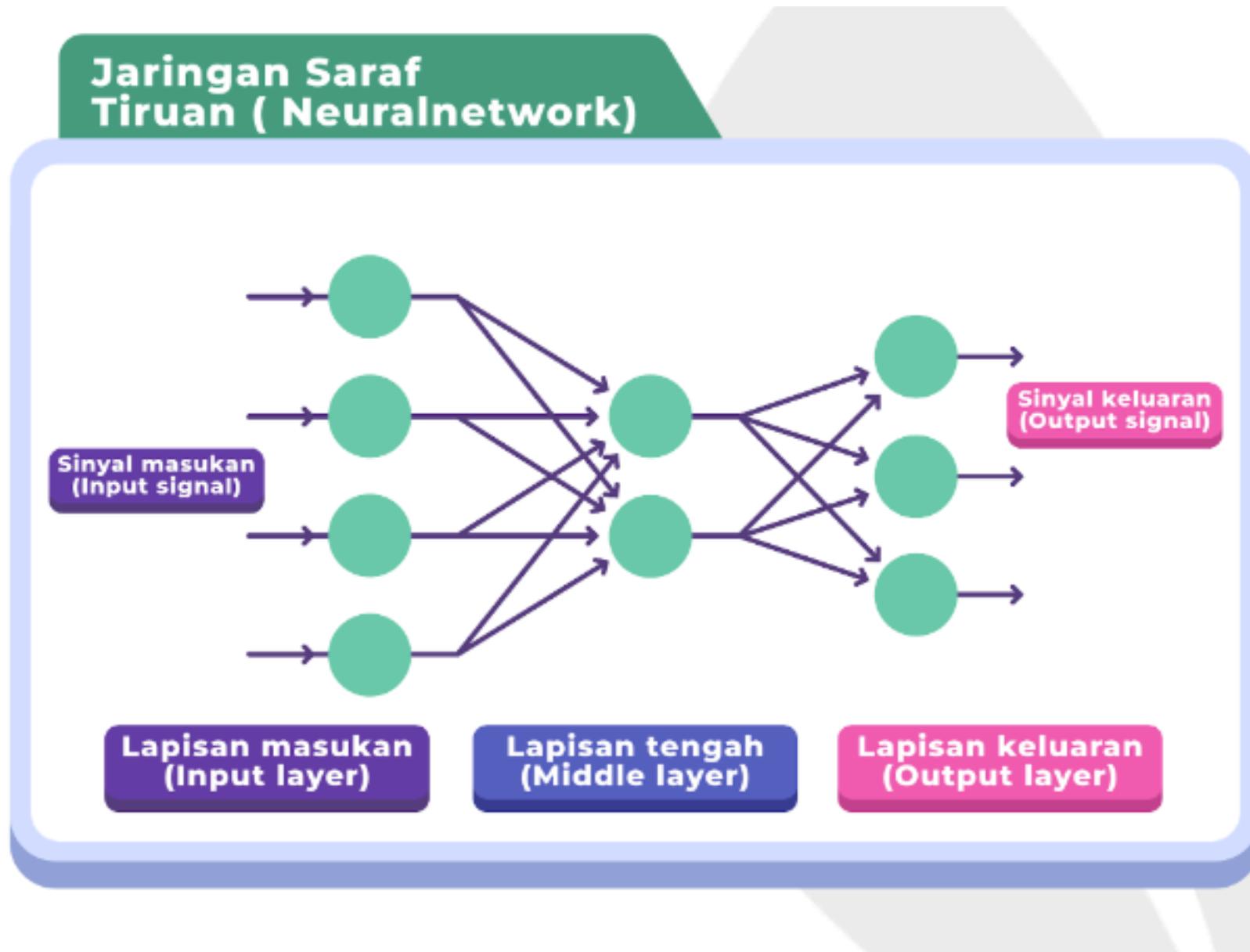
Model yang digunakan



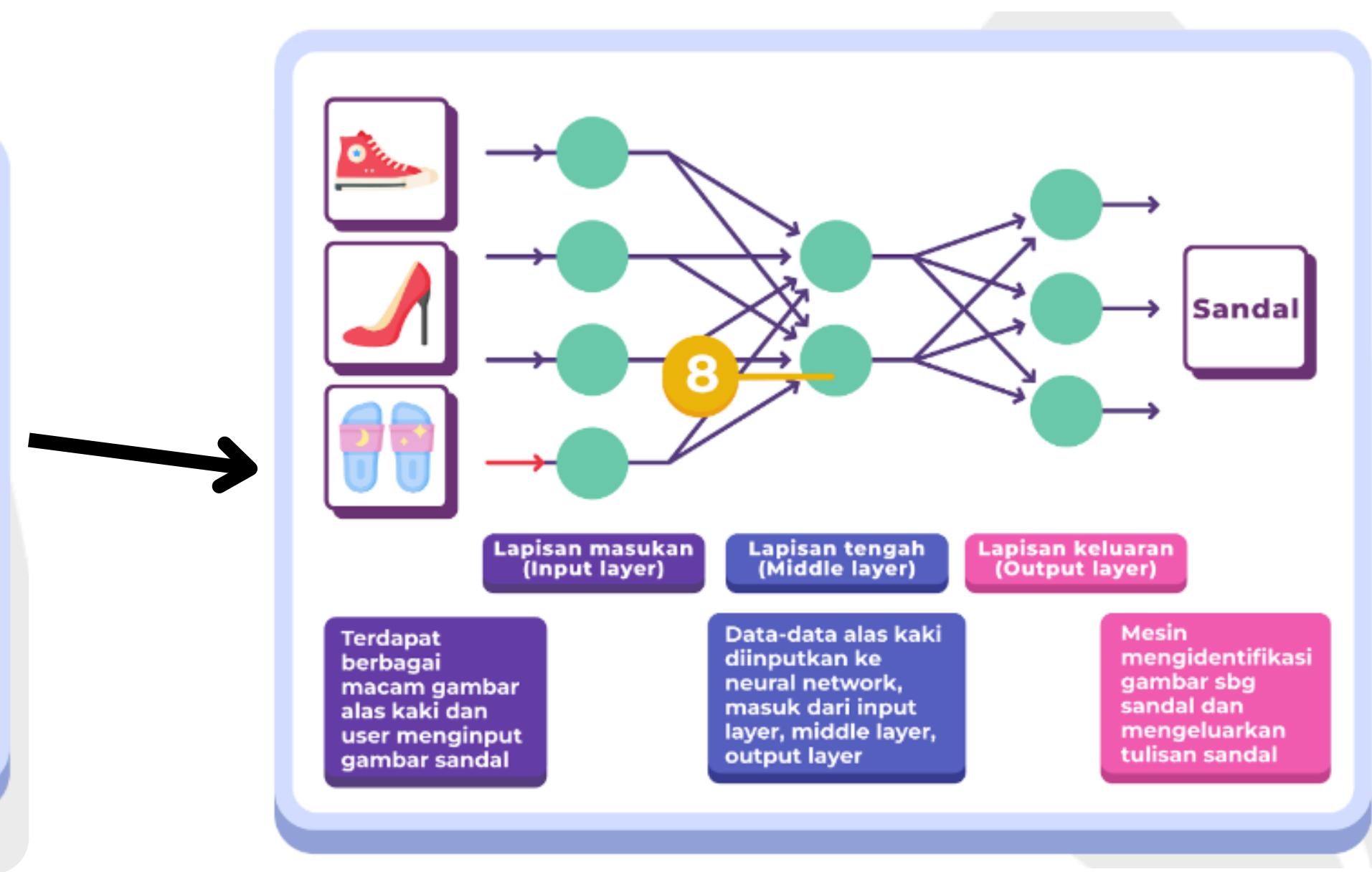
Model MLP C

Model LSTM

Model Neural Network



(1)



(2)

Sumber : Materi Bootcamp Data Sains Binar Academy (1, 2)

Model LSTM (Long Shorterm Memory)

Gerbang/Gate

Forget gate : Memilih informasi yang akan dilupakan atau dihapus

Input Gate : Memilih informasi baru yang akan ditambahkan ke dalam sel memori

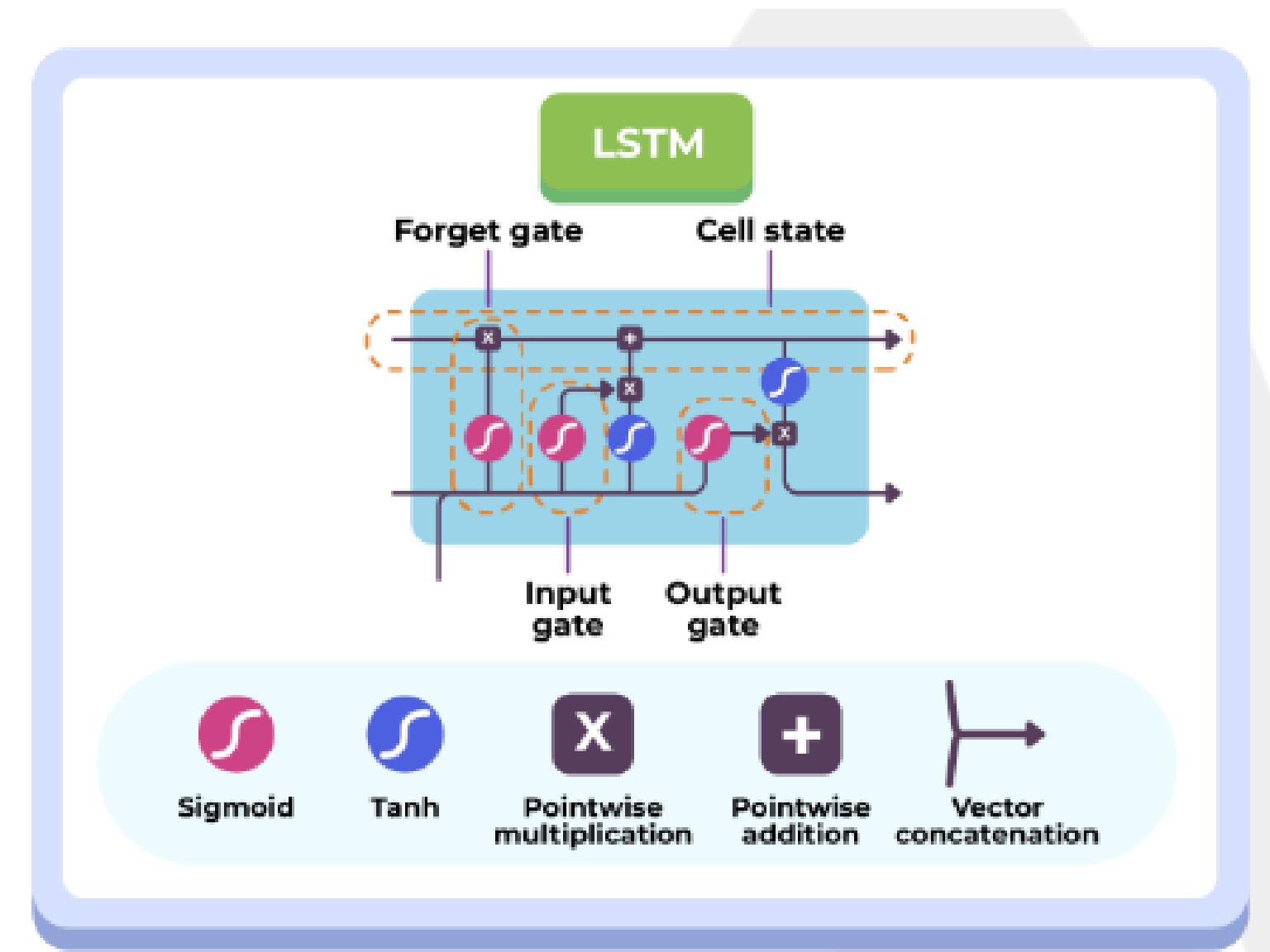
Output Gate : Menentukan output berdasarkan sel memory yang sudah di-update

Sel Memory

Menyimpan informasi jangka panjang dari waktu ke waktu

Hidden State

Ini adalah output dari sel LSTM pada suatu waktu tertentu dan berfungsi sebagai input untuk selanjutnya. Hidden state juga dapat dianggap sebagai representasi internal dari informasi yang relevan.



Gambar Model LSTM

Sumber : Materi Bootcamp Data Sains Binar Academy



Tahap Pengolahan Data

Data Preprocessing

Menghilangkan karakter, kata - kata yang tidak diperlukan seperti kata sambung dan imbuhan kata

Menjadikan kata - kata yang akan dilakukan analisis sentiment dalam bentuk paling dasarnya.



```
# Defining preprocess function
def preprocess_nya(text):
    text = lowercase(text)
    text = remove_nonaplhanumeric(text)
    text = remove_unnecessary_char(text)
    text = normalize_alay(text)
    text = remove_stopwords(text)
    text = stemming(text)
    return text
```



CodeImage



BINAR



Tahap Pengolahan Data

Feature Extraction

Menghilangkan karakter, kata - kata yang tidak diperlukan seperti kata sambung dan imbuhan kata
Menjadikan kata - kata yang akan dilakukan analisis sentiment dalam bentuk paling dasarnya dan melakukan ekstraksi fitur untuk diproses ke dalam model.

The screenshot shows a dark-themed code editor window titled 'feature_mlp.py'. The code itself is as follows:

```
X_MLP = df_token['Tweet_cleaned_token_2']
Y_MLP = df_token['Sentiment']

vectorizer_MLP = CountVectorizer(max_features=100000)
X_MLP_Bow = vectorizer_MLP.fit_transform(X_MLP)
```

In the bottom right corner of the code editor, there is a small watermark or logo for 'CodelImage'.



Tahap Pengolahan Data

Build Model dan Train Data

The screenshot shows a dark-themed code editor window with a Python script named 'feature_mlp.py'. The code implements a Machine Learning Pipeline for sentiment analysis using an MLPClassifier. It starts by importing necessary libraries (Pandas, CountVectorizer, and MLPClassifier), then defines X_MLP and Y_MLP from a DataFrame 'df_token'. It uses a CountVectorizer to convert text into numerical features (X_MLP_Bow). The data is split into training and testing sets (X_train_MLP, X_test_MLP, y_train_MLP, y_test_MLP) with a test size of 0.1. An MLPClassifier is trained on the training data, and the process is marked as completed with a print statement.

```
X_MLP = df_token['Tweet_cleaned_token_2']
Y_MLP = df_token['Sentiment']

vectorizer_MLP = CountVectorizer(max_features=100000)
X_MLP_Bow = vectorizer_MLP.fit_transform(X_MLP)

X_train_MLP, X_test_MLP, y_train_MLP, y_test_MLP = train_test_split(X_MLP_Bow, Y_MLP, test_size=0.1)
MLP = MLPClassifier(hidden_layer_sizes=(100,), activation = 'logistic', solver='adam', alpha = 0.2, batch_size = 'auto', max_iter=300, random_state=42)
MLP.fit(X_train_MLP, y_train_MLP)
print("Training Data with MLP model is Finished!")
```

 CodelImage

Menghilangkan karakter, kata - kata yang tidak diperlukan seperti kata sambung dan imbuhan kata

Menjadikan kata - kata yang akan dilakukan analisis sentiment dalam bentuk paling dasarnya.



Tahap Pengolahan Data

Model Evaluation

Melihat seberapa akurat model dalam memprediksi sentimen proses ini menjadi pertimbangan dalam layak atau tidak suatu model disimpan dan digunakan dalam API



The screenshot shows a dark-themed code editor window with a tab labeled "feature_mlp.py". The code itself is as follows:

```
y_pred_MLP = model.predict(X_test_MLP)

accuracy_MLP = accuracy_score(y_test_MLP, y_pred_MLP)
report_MLP = classification_report(y_test_MLP, y_pred_MLP)

print(f"Accuracy: {accuracy_MLP*100}%, '%")
print("Classification Report:\n", report_MLP)
```

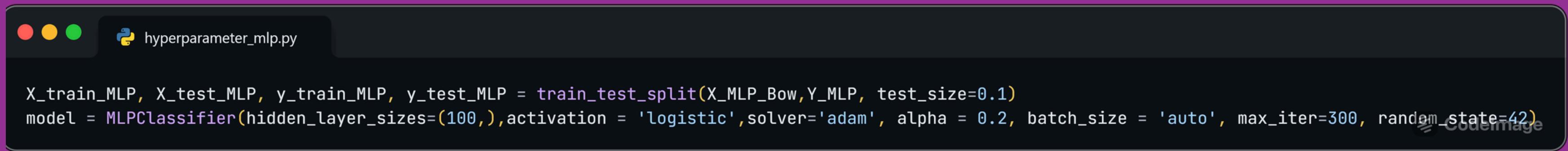
A watermark "CodeImage" is visible in the bottom right corner of the code editor.





Tahap Pengolahan Data

Setting parameter Model



The screenshot shows a terminal window with a dark theme. In the title bar, there are three small colored circles (red, yellow, green) followed by the file name "hyperparameter_mlp.py". The main area of the terminal contains the following Python code:

```
X_train_MLP, X_test_MLP, y_train_MLP, y_test_MLP = train_test_split(X_MLP_Bow, Y_MLP, test_size=0.1)
model = MLPClassifier(hidden_layer_sizes=(100,), activation = 'logistic', solver='adam', alpha = 0.2, batch_size = 'auto', max_iter=300, random_state=42)
```

A watermark "CodeImage" is visible in the bottom right corner of the terminal window.

Memperbarui parameter model supaya mendapatkan akurasi performa model yang lebih tinggi

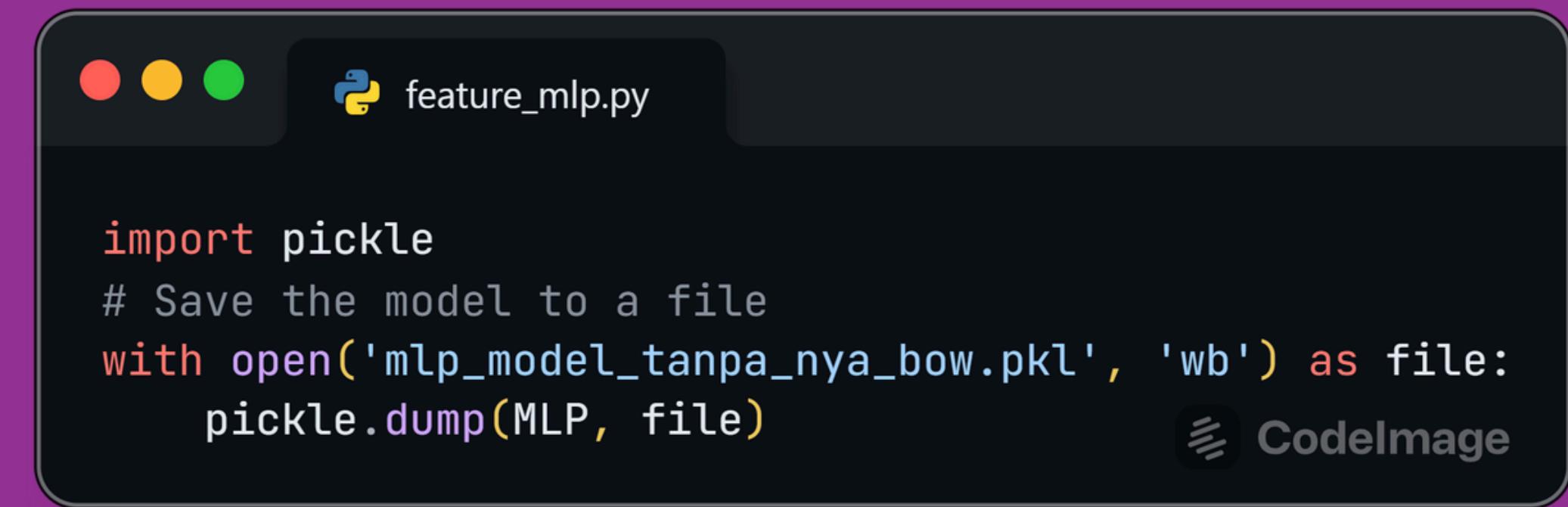




Tahap Pengolahan Data

Menyimpan Model

Menyimpan model MLP dan LSTM
Model yang sudah disimpan
nantinya akan dimasukkan ke
dalam API



A screenshot of a terminal window titled "feature_mlp.py". The window has three colored status indicators (red, yellow, green) at the top left. The file name "feature_mlp.py" is displayed at the top right. The terminal content shows Python code for saving a machine learning model:

```
import pickle
# Save the model to a file
with open('mlp_model_tanpa_nya_bow.pkl', 'wb') as file:
    pickle.dump(MLP, file)
```

In the bottom right corner of the terminal window, there is a watermark that says "CodeImage" next to a small icon.



Tahap Pengolahan Data

Menyusun API

- Mendefinisikan packages
- Mendefinisikan endpoint
- Memasukkan Model
- Memasukkan x_pad dan y_pad
- Memasukkan proses cleaning data
- Menampilkan hasil prediksi sentimen

```
...
Alur API
Menerima data .csv, cleaning tweet column, tokenisasi feature, tf-idf atau
bag of word, proses oleh model dan munculkan dengan .json_response
'''

#=====
# Import API packages =====
import re
import pandas as pd
import numpy as np
from flask import Flask,jsonify, request,send_file
import io
from flask import request
from flasgger import Swagger, LazyString, LazyJSONEncoder
from flasgger import swag_from
```

 CodeImage



BINAR



HASIL

Berikut hasil prediksi sentimen dari kedua model



Hasil Perhitungan

Neural Network

Initial Value					
x_1	x_2	x_3	α	Threshold	$Y_{d,6}$
0,7	0,8	0,9	0,1	-1	0

Initial Random										
W_{14}	W_{15}	W_{24}	W_{25}	W_{34}	W_{35}	W_{46}	W_{56}	θ_4	θ_5	θ_6
0,5	0,6	0,3	1,1	-1,0	0,1	-1,1	-0,7	0,2	0,3	0,4





Hasil Perhitungan

Neural Network

Y_4	Y_5	Y_6	e
0,3751	0,7483	0,208	-0,208

δ_6	∇_{46}	∇_{56}	$\nabla \theta_6$
-0,034	0,0012	0,0025	0,0034

δ_4	δ_5
0,0088	0,0045





Hasil Perhitungan

Neural Network

Y_4	Y_5	Y_6	e
0,3751	0,7483	0,208	-0,208

δ_6	∇_{46}	∇_{56}	$\nabla \theta_6$
-0,034	0,0012	0,0025	0,0034

δ_4	δ_5
0,0088	0,0045





Hasil Perhitungan

Neural Network

∇w_{14}	∇w_{24}	∇w_{34}	$\nabla \theta_4$	∇w_{15}	∇w_{25}	∇w_{35}	$\nabla \theta_5$
0,000616	0,000704	0,000792	-0,00088	0,000315	0,00036	0,000405	-0,00045

w_{14}	w_{15}	w_{24}	w_{25}	w_{34}	w_{35}	θ_3	θ_4	θ_5
0,5006	0,6003	0,3007	1,1003	-0,9992	0,1004	0,199	0,299548	0,403



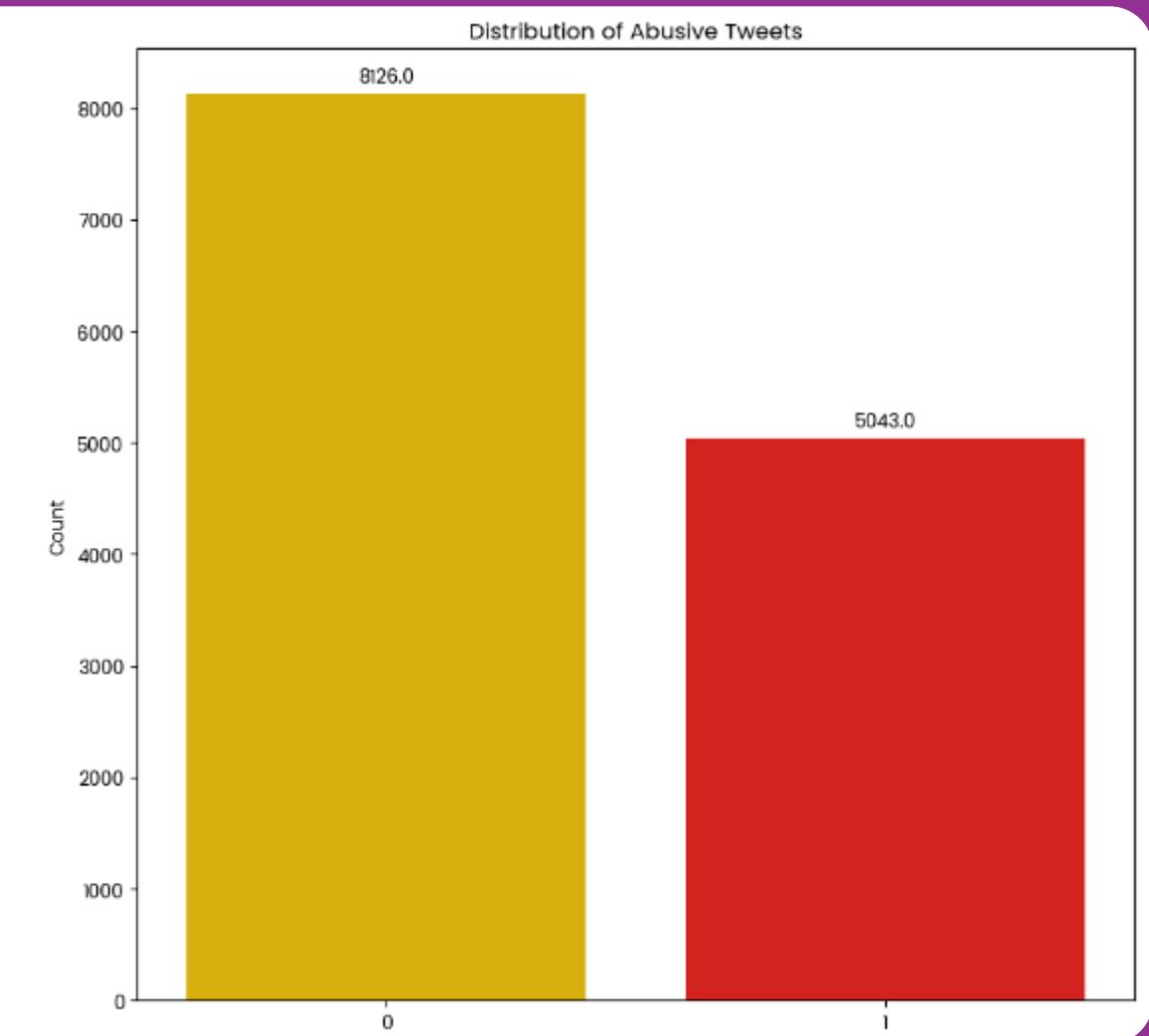
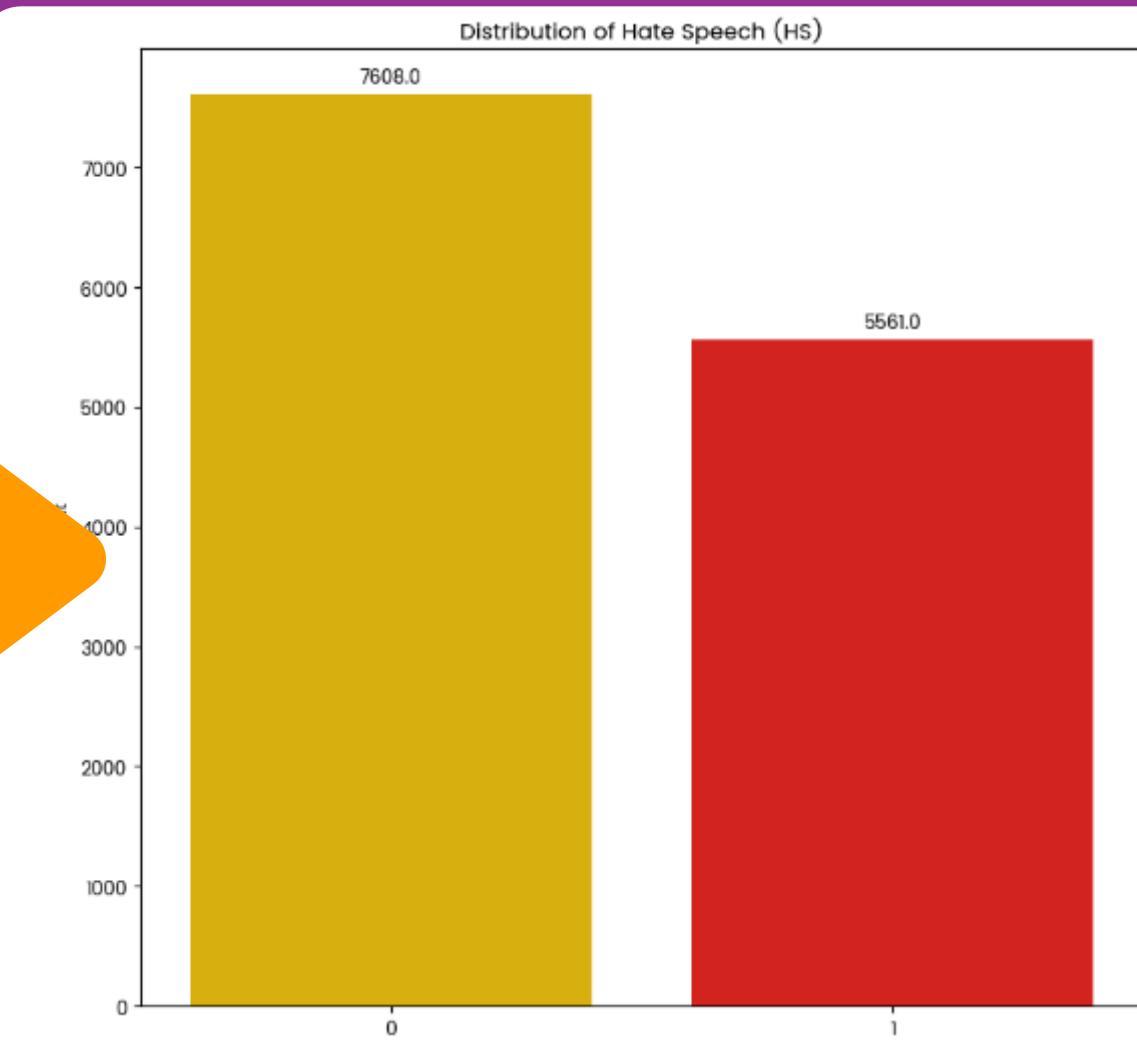
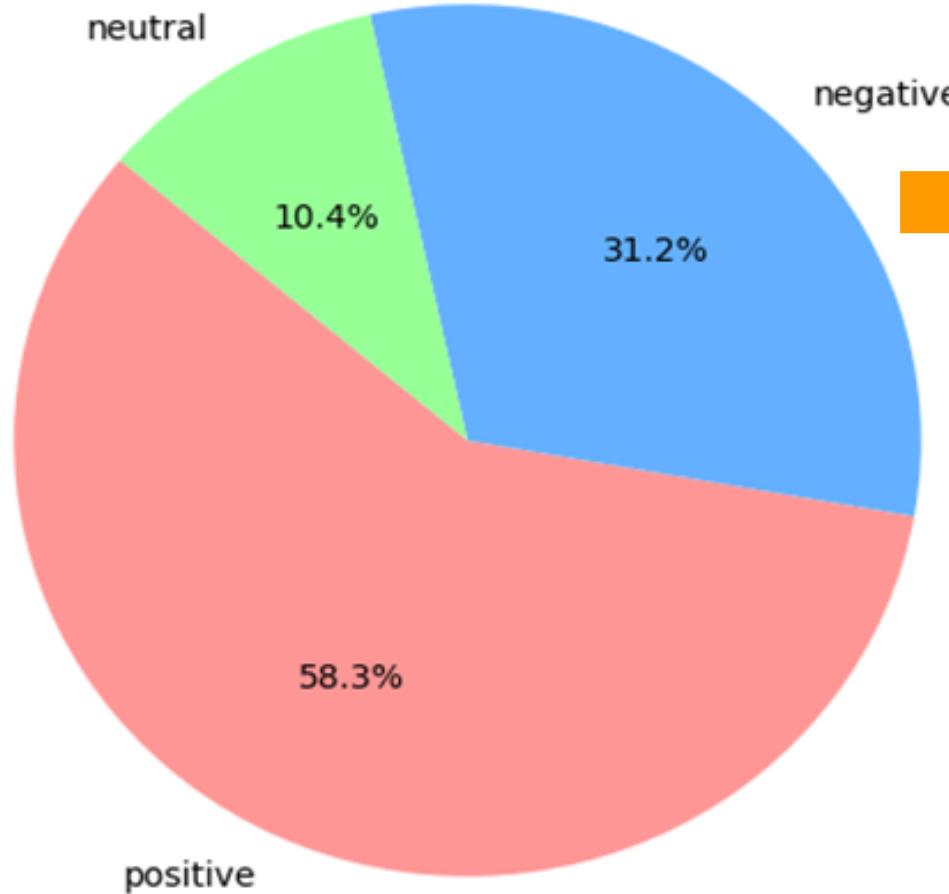
Hasil Perhitungan

Neural Network

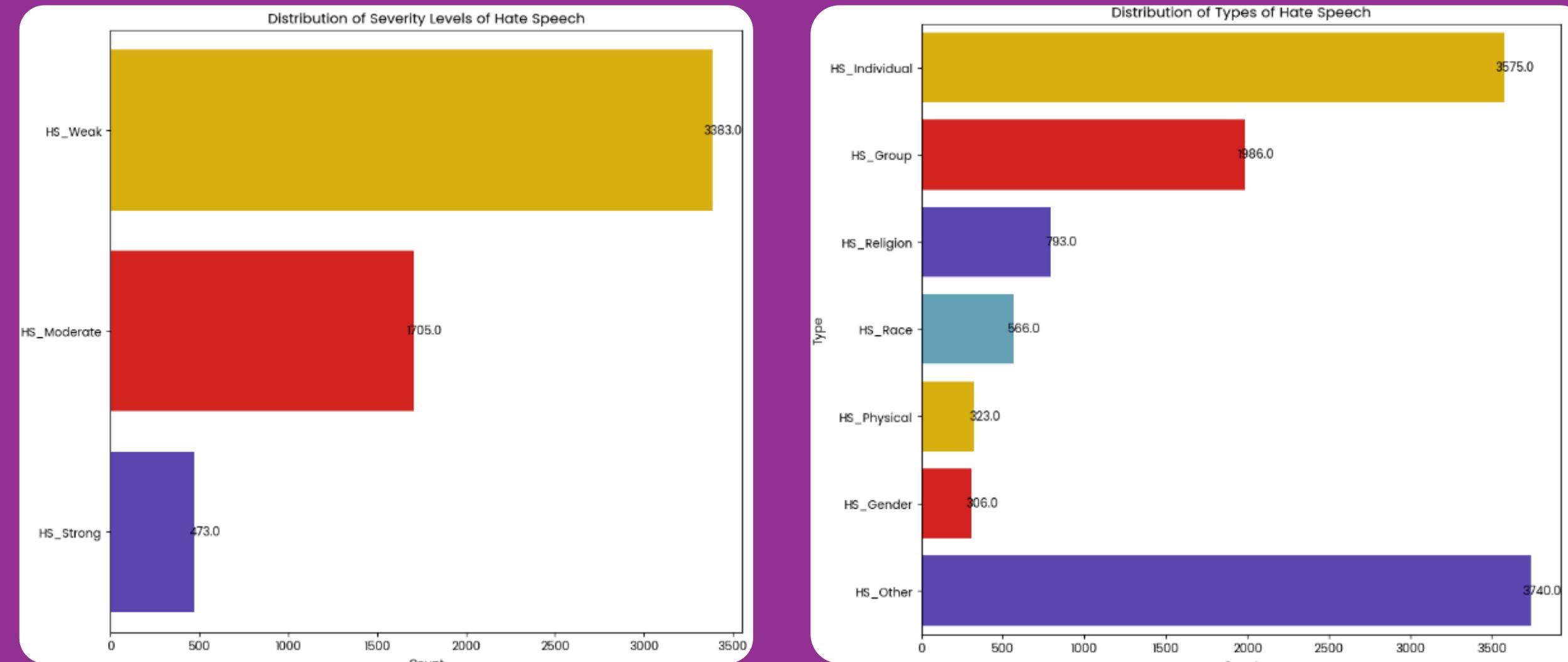
○○○ Tentang data input

Real data

Sentiment Distribution



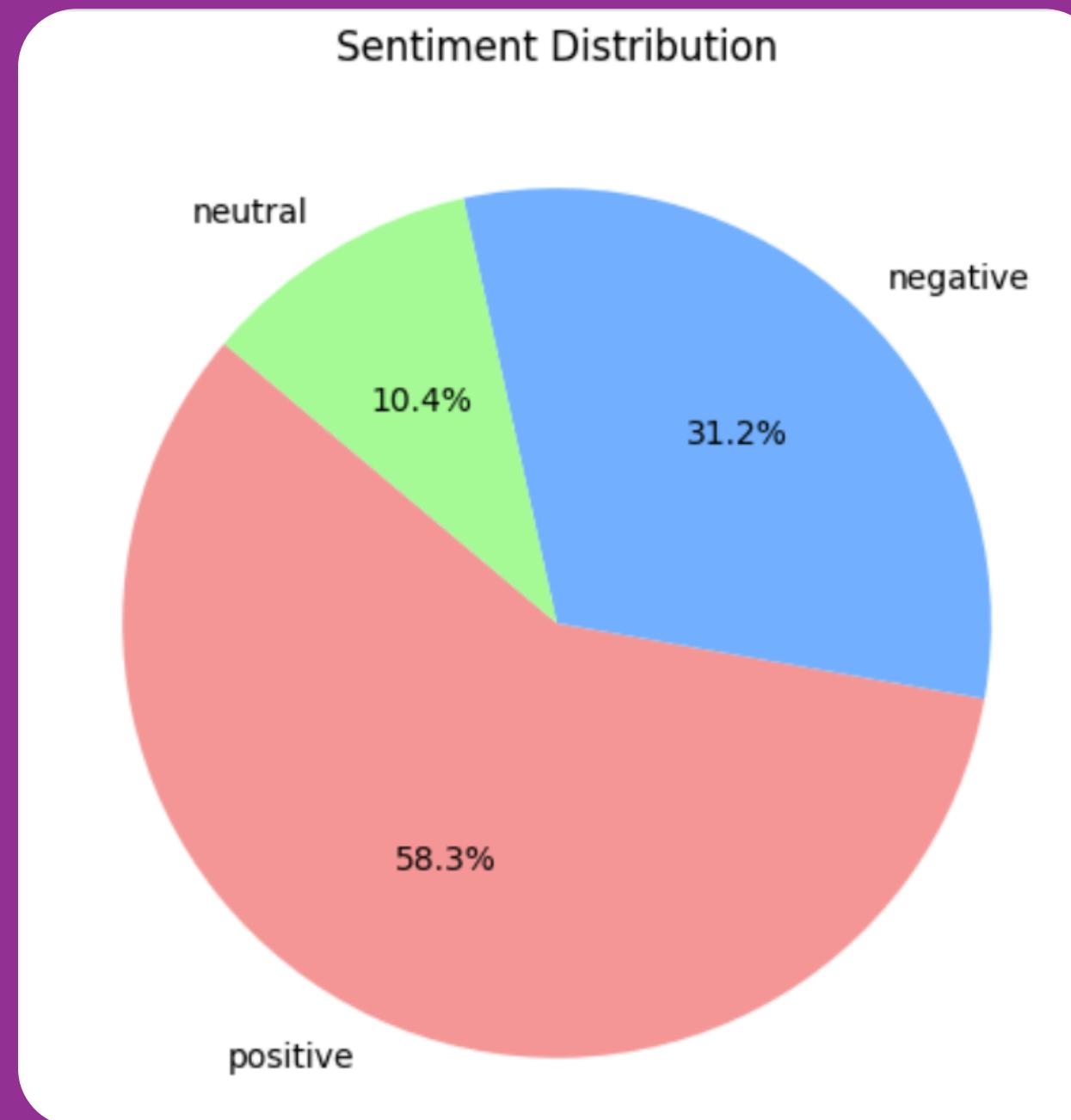
○ ○ ○ Visualisasi Data



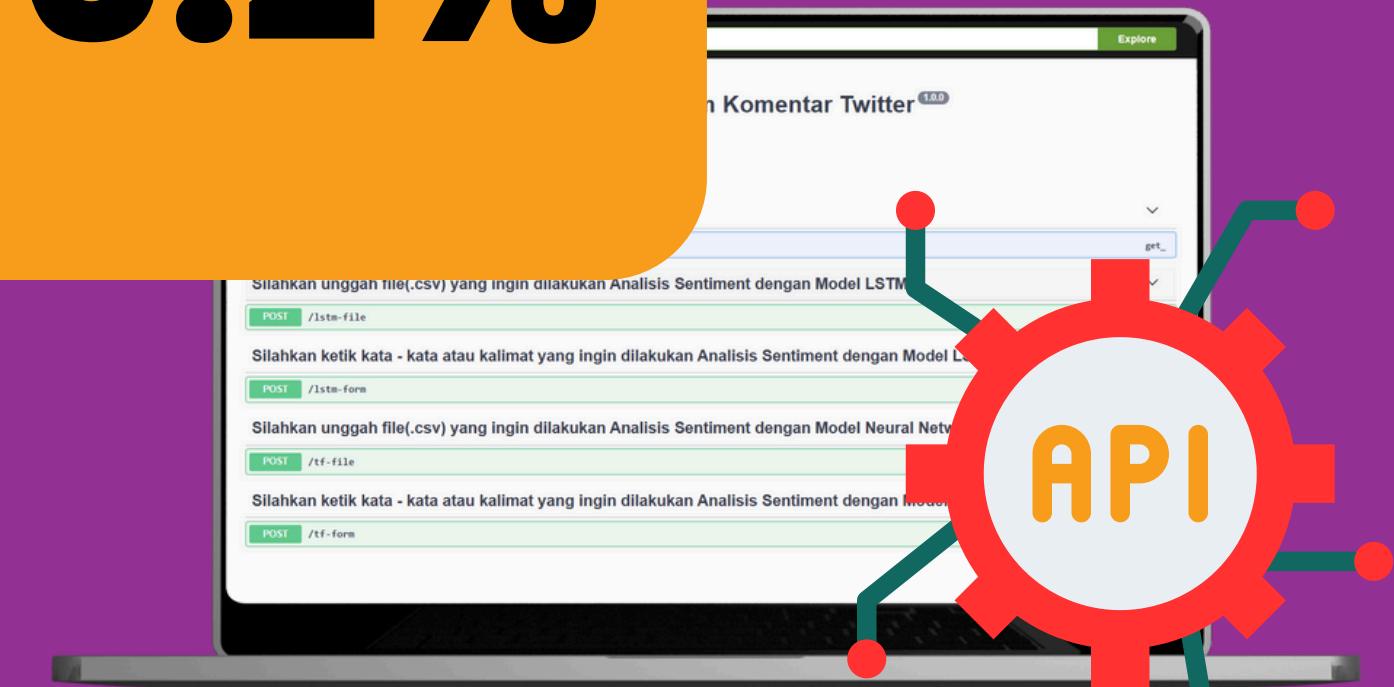
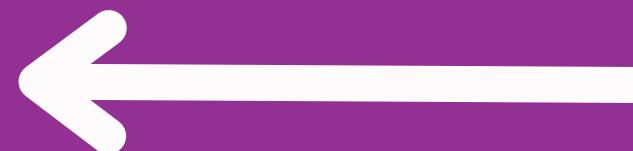
- Sentimen positif memiliki jumlah yang paling tinggi, sekitar lebih dari 6000, dibandingkan dengan sentimen lainnya
- Terdapat sekitar 4000 sentimen negatif
- Untuk sentimen netral mendapat paling sedikit yaitu 2000

Hasil Prediksi Sentiment – MLPC

Data Training



86.2%



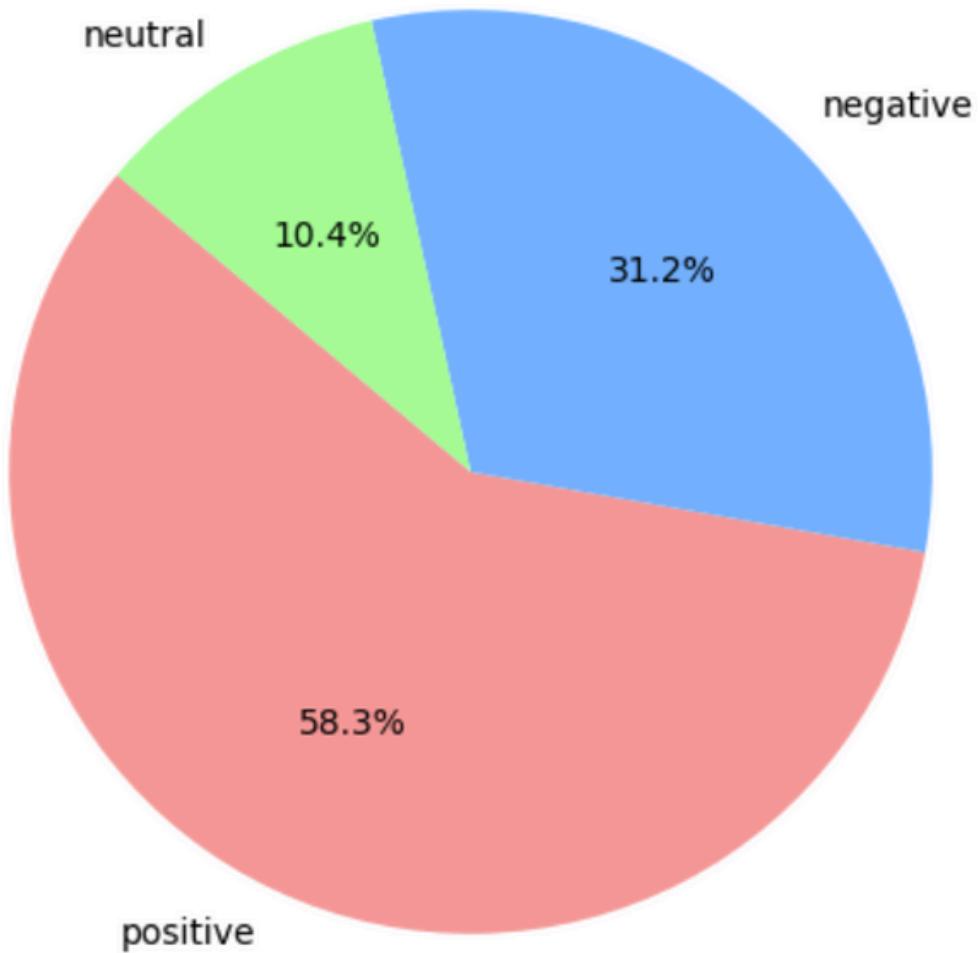
- Data training yang digunakan memiliki sentimen positif sebanyak 58.3%, negatif 31.2%, dan netral 10.4
- Data testing yang dimasukkan dalam API memiliki sentimen positif, negatif, dan netral



Hasil Prediksi Sentiment – MLPC

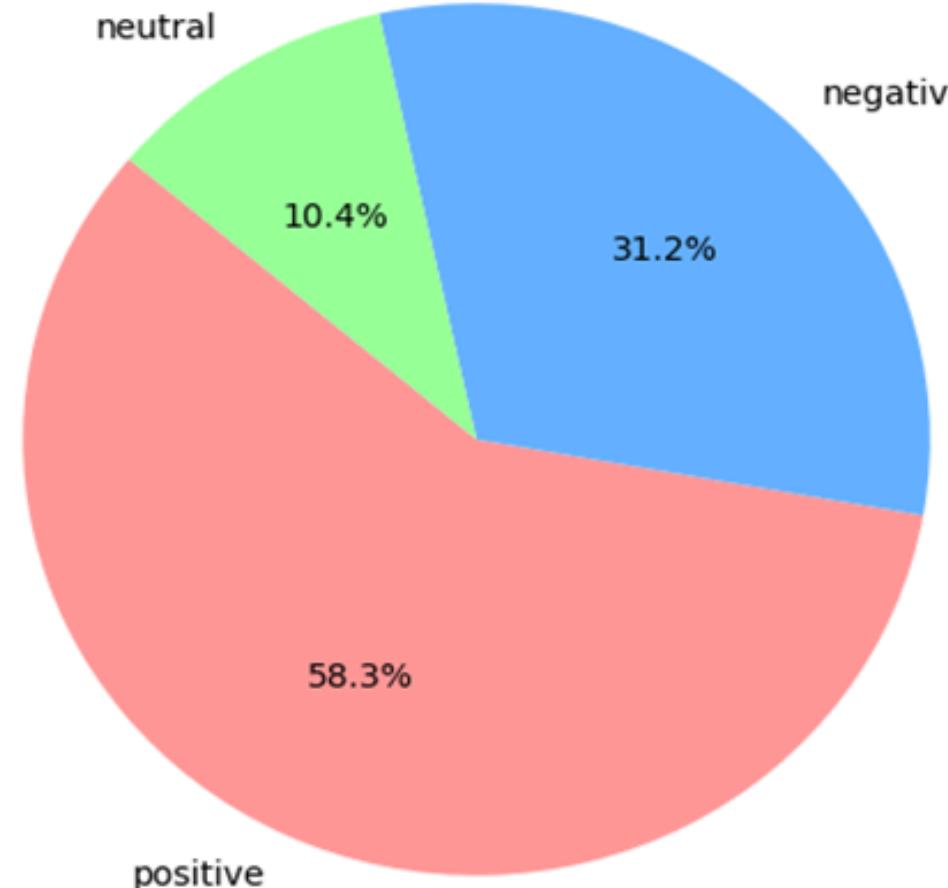
Data Training

Sentiment Distribution



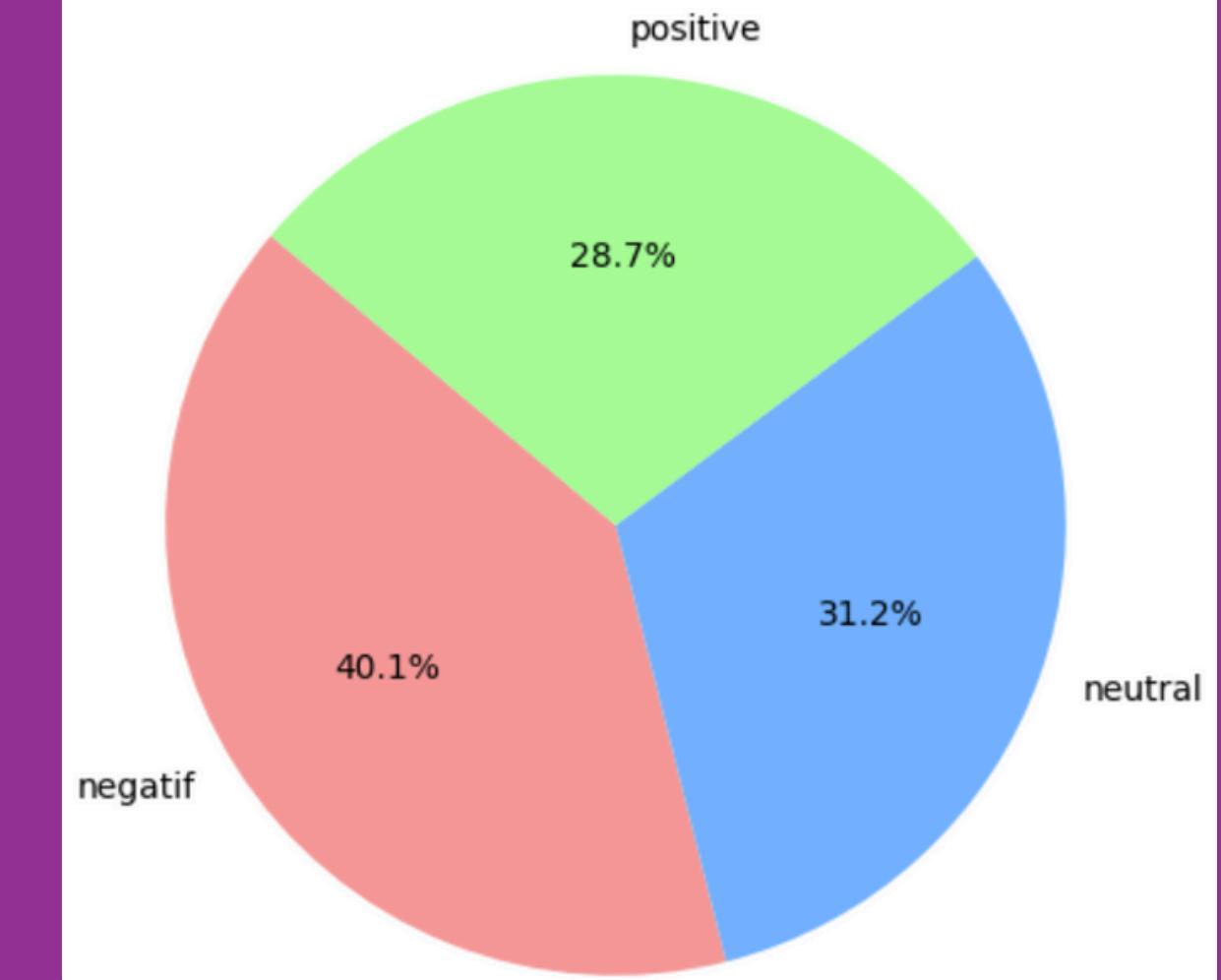
Real data

Sentiment Distribution



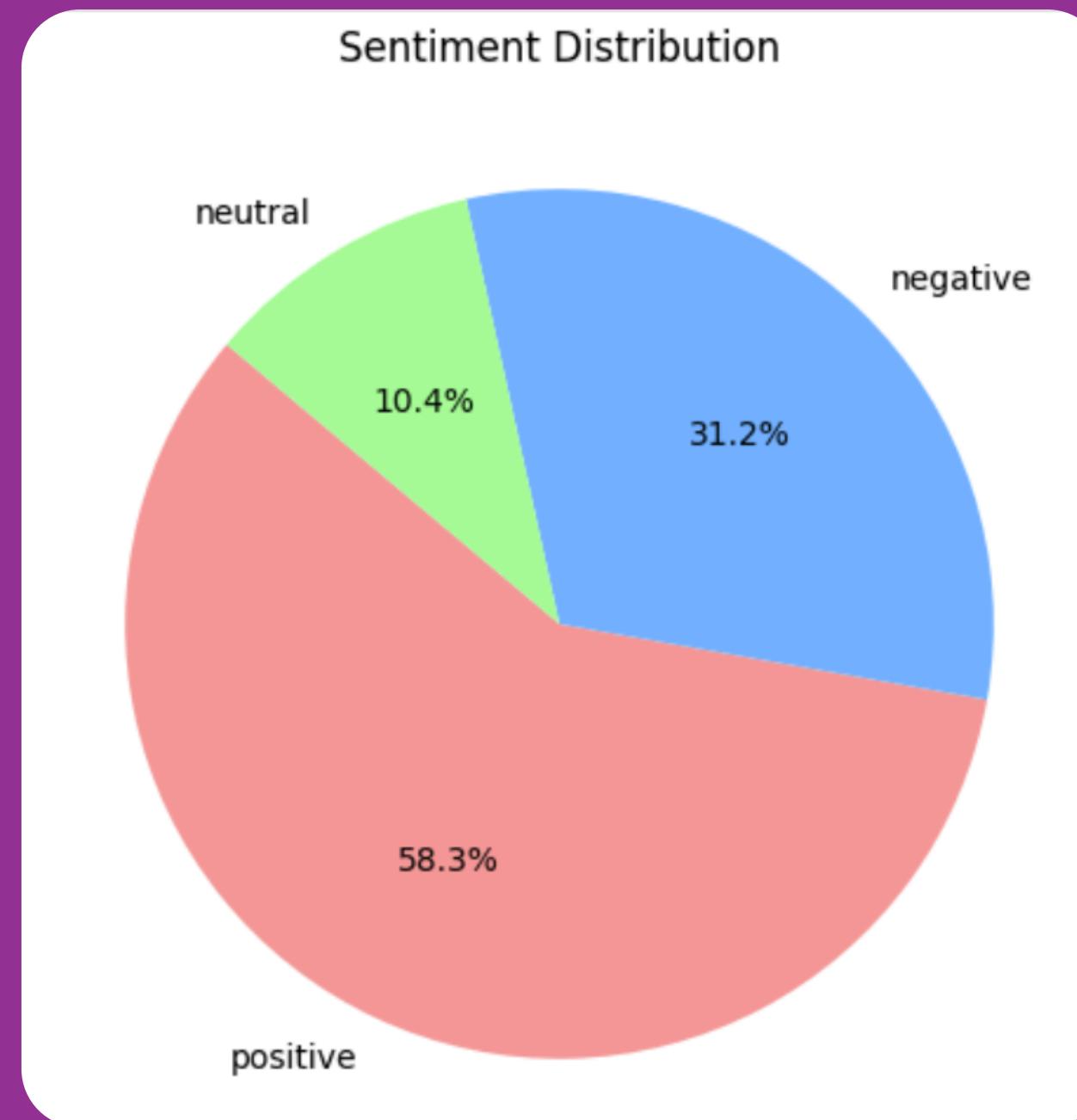
Sentimen prediksi

Sentiment Distribution



Hasil Prediksi Sentiment – LSTM

Data Training

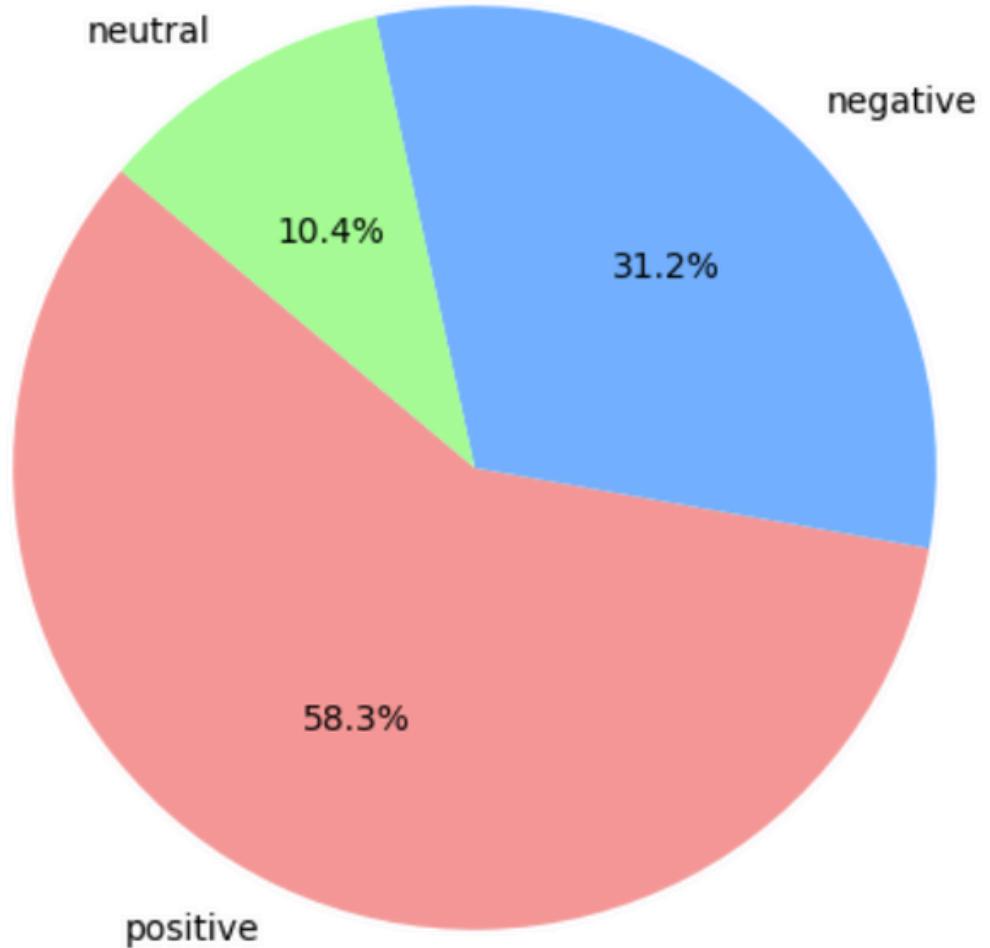




Hasil Prediksi Sentiment – LSTM

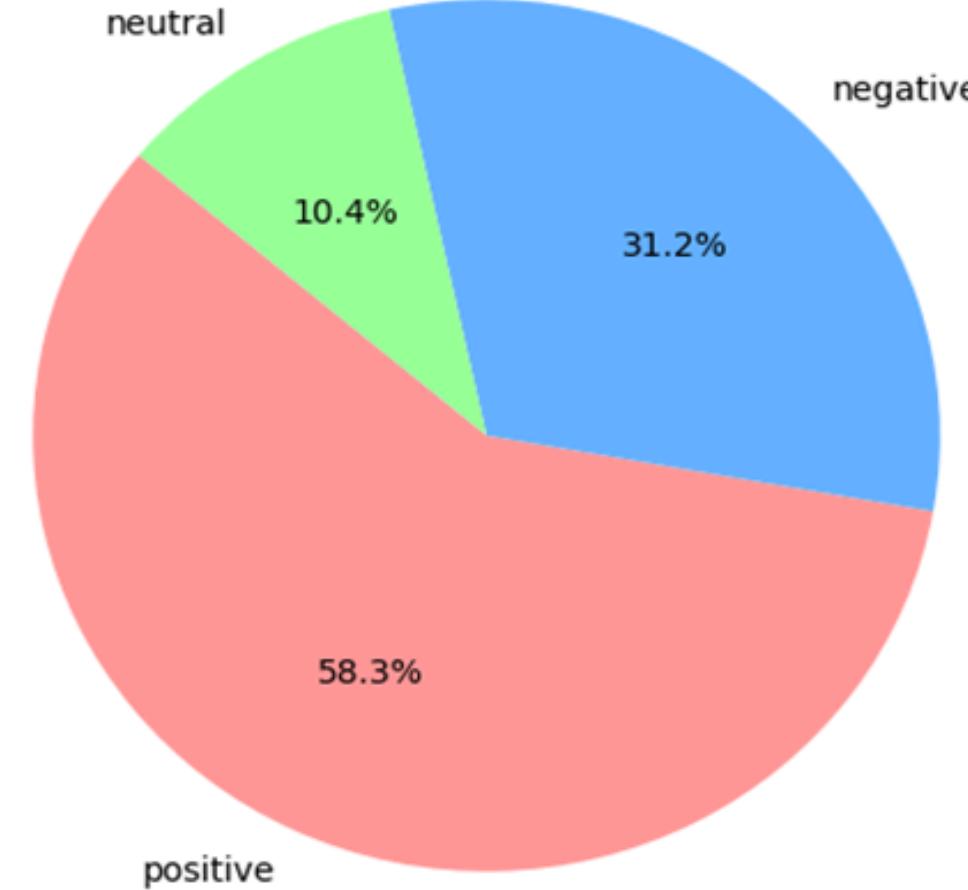
Data Training

Sentiment Distribution



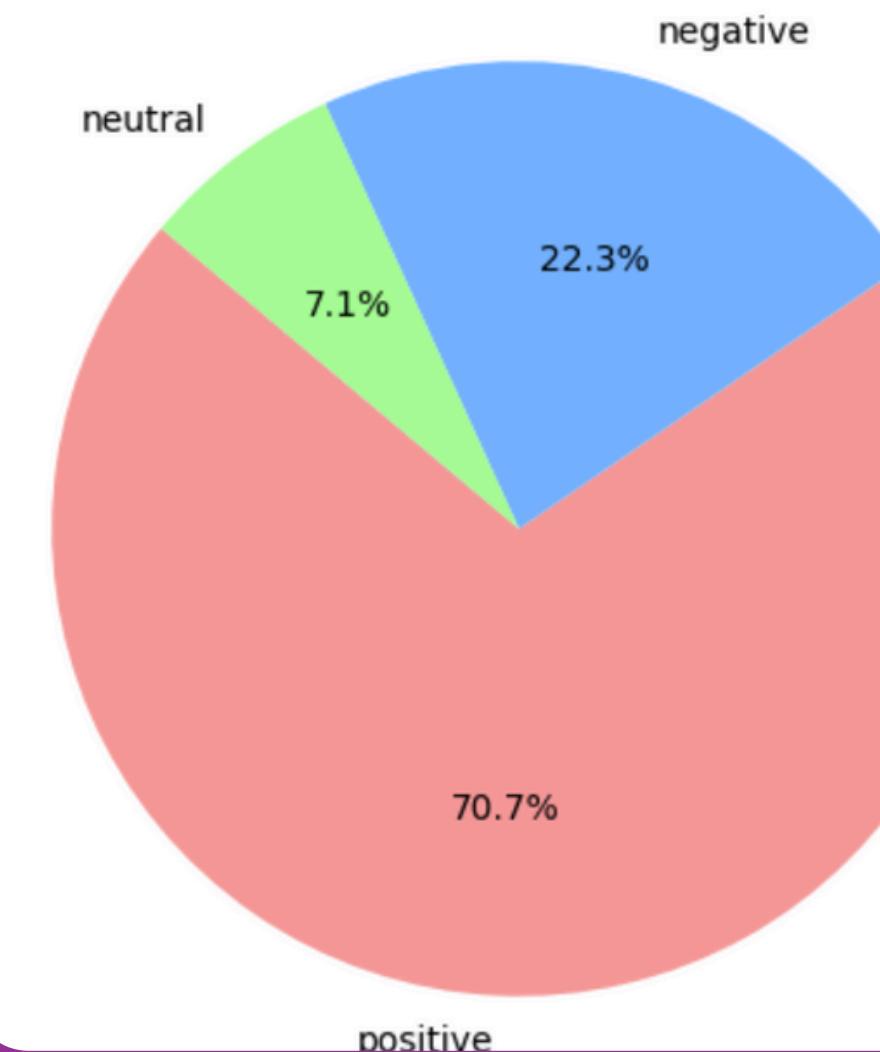
Real data

Sentiment Distribution



Sentimen prediksi

Sentiment Distribution



● o o Visualisasi Data

Hasil Atau Manfaat Analisis Sentimen Pada Pemerintah & Perusahaan

- Memberi dampak positif bagi pemerintah dalam membuat keputusan lebih baik guna perumusan kebijakan yang lebih responsif terhadap kebutuhan dan kekhawatiran masyarakat.
- Menambah wawasan pemerintah untuk merumuskan kebijakan yang lebih baik dan lebih sesuai dengan kebutuhan dan aspirasi masyarakat, serta untuk meningkatkan layanan publik.
- Mendukung keputusan bisnis, seperti peluncuran produk baru, strategi pemasaran, atau respon terhadap kritik publik.
- Mengungkapkan kebutuhan atau keluhan yang diungkapkan oleh konsumen, yang dapat menjadi input berharga untuk pengembangan produk atau peningkatan layanan. Ini juga bisa membantu dalam mengenali tren baru yang mungkin belum disadari





Kesimpulan

Berikut hasil prediksi sentimen dari kedua model

● ○ ○ Kesimpulan

- Bag of word lebih cocok untuk digunakan dalam model MLP Classifier, memiliki akurasi lebih tinggi pada model MLP Classifier yaitu 86% dan 83 % untuk TF-IDF
- LSTM lebih cocok menggunakan TF-IDF dengan mempertimbangkan waktu training data
- Model MLP C berhasil memprediksi data baru dengan akurasi sebesar 86% dan cukup bisa menggambarkan sentimen komentar di twitter
- Model LSTM memiliki kemampuan yang lebih rendah dalam memprediksi senitment berdasarkan hasil prediksi dari data masukkan (data.csv) dibanding dengan model MLP Classifier dengan feature extraction bag of words

● ○ ○ Saran

- Melakukan cleaning dan pre-processing data dengan lebih detail
- Melakukan research lebih mendalam untuk mendapatkan formulasi hyperparamatertuning yang lebih optimal sehingga menghasilkan prediksi yang lebih akurat

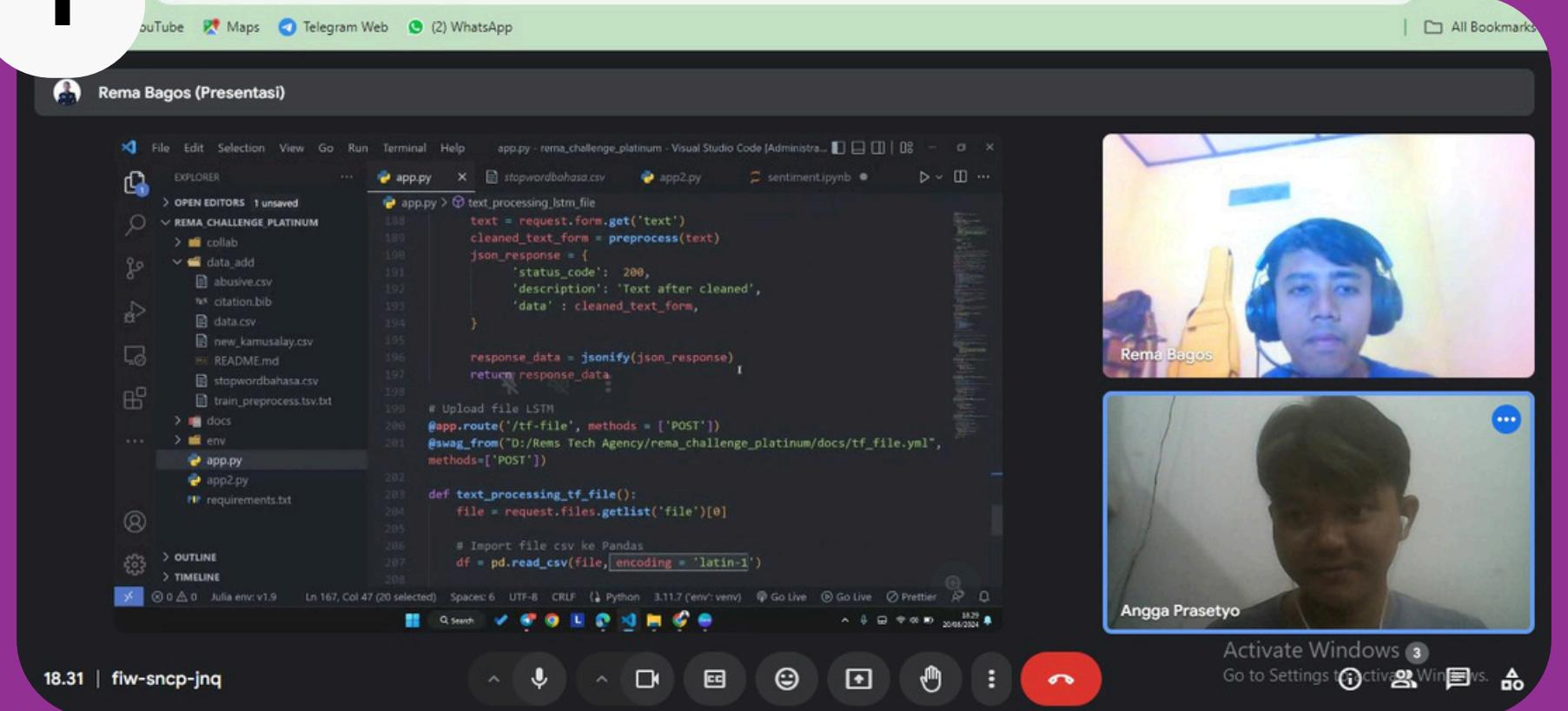


Terima kasih!

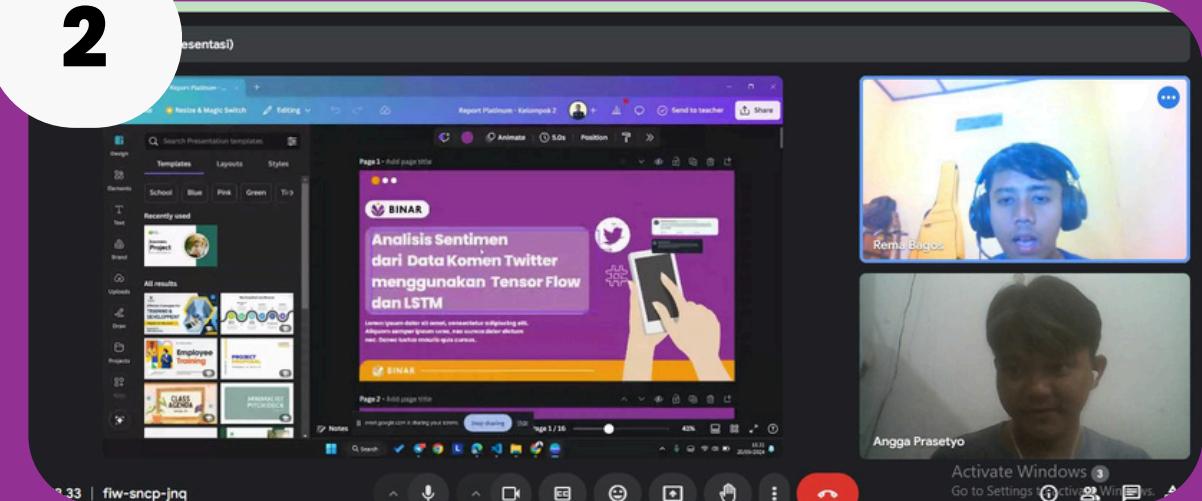


Dokumentasi

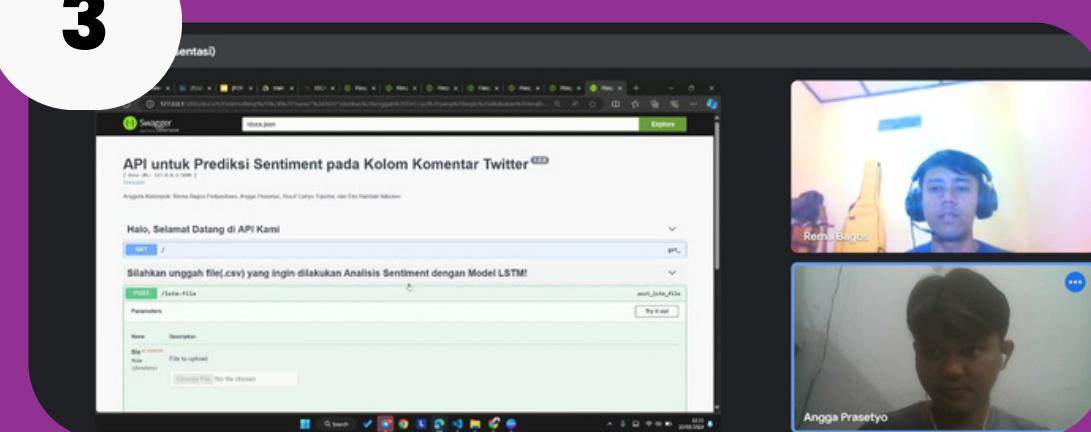
1



2



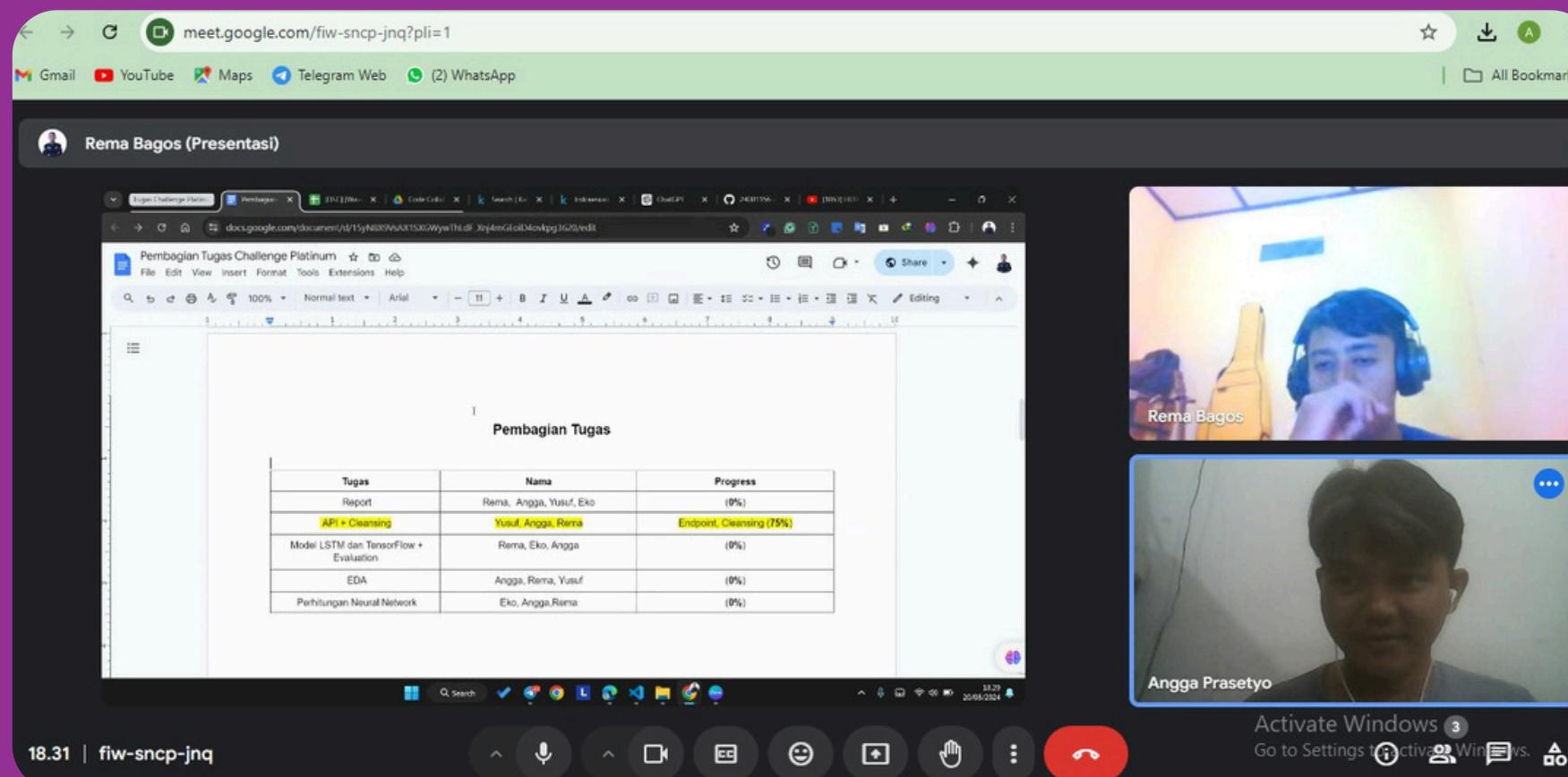
3



- Pembuatan API
- Penyusunan Report
- Testing API



Dokumentasi



No	Pekerjaan	Status
1	Perhitungan Neural Network Manual	In Progress
2	Konsep statistika dalam cleansing data	Not Started
3	Konsep statistika dalam evaluasi Model	Not Started
4	API dengan 4 endpoint	Completed
5	API dengan Swagger UI dan Flask	Completed
6	API input form dan file	Completed
7	API menghasilkan output dengan label positive, negative dan netral	Not Started
8	API dapat menganalisis sentimen menggunakan LSTM	Not Started
9	Report (judul, pendahuluan, metode (termasuk script python), hasil)	In Progress

- Diskusi tentang pembagian tugas untuk setiap anggota kelompok
- Rekap Progress pengerjaan dari setiap anggota dan progress sesuai dengan penilaian challenge platinum



Wordcloud

Sentimen Positif



Sentimen Negatif



Sentimen Netral

