

---

# ANALISIS SENTIMEN DENGAN MACHINE LEARNING: **NN & LSTM**

Data Science Wave 2 Binar Academy

- Dhitya Ovim Barkley Perdana
  - Rai Ikhsan Wiratama
  - Ngurah Pradipta Surya Mandala
  - Sabillah Sakti
-

# PENDAHULUAN

Pada *challenge* kali ini, akan dibuat sebuah *machine learning* (ML) yang akan memprediksi kalimat berdasarkan kecenderungan sentimennya, apakah kalimat tersebut bersifat **positif**, **negatif**, atau **netral**. Untuk memprediksi sentimen suatu kalimat, akan dibuat dua model prediksi yakni model **Neural Network** (NN) dan model **Long Short Term Memory** (LSTM). Di akhir penelitian, akan ditentukan model ML mana yang lebih efektif dalam memprediksi sentimen.



---

# METODE

---

## ALUR PENELITIAN



### CLEANSING

Pembersihan dataset terhadap huruf kecil, simbol-simbol, dan mengubah kata-kata alay menjadi formal



### MACHINE LEARNING

Pembuatan model machine learning dengan metode NN dan LSTM



### DEPLOYMENT

Peluncuran model machine learning ke API Flask & Swagger



### ANALYSIS

Analisis mengenai perbandingan prediksi sentimen dengan metode NN dan LSTM

## PACKAGE PYTHON YANG DIGUNAKAN

Package	Kegunaan
Pandas	Mengolah data dalam bentuk tabel (dataframe)
Regex	Melakukan cleansing kata-kata/karakter
Scikit Learn	Membuat model machine learning NN
Tensorflow	Membuat model machine learning LSTM
Pickle	Mengekstrak model untuk diproses di API
Matplotlib	Menampilkan grafik
Flask & Swagger	Membuat API

---

# **HASIL & KESIMPULAN**

---

# MODEL NEURAL NETWORK

## Evaluasi Nilai F1 Score

Data Split	Akurasi Rata-Rata
Data Test	0.85
Data Train	0.91

### Insight:

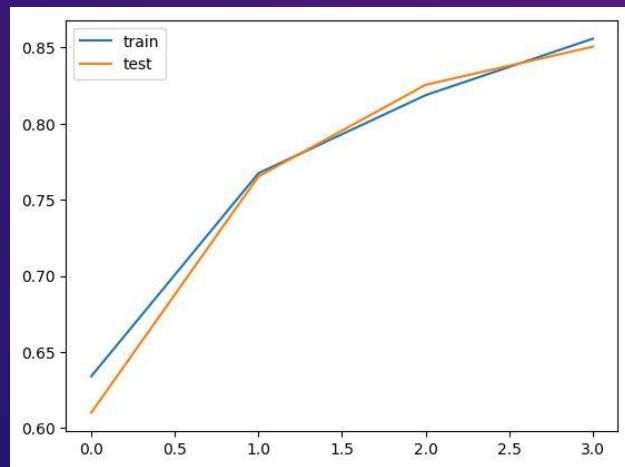
- Nilai F1 score mendekati 1 → Good Fit
- Data Test < Data Train → Good Fit Underfit
- Akurasi validasi silang  $\pm 0.85$  → modelnya stabil

## Cross Validation

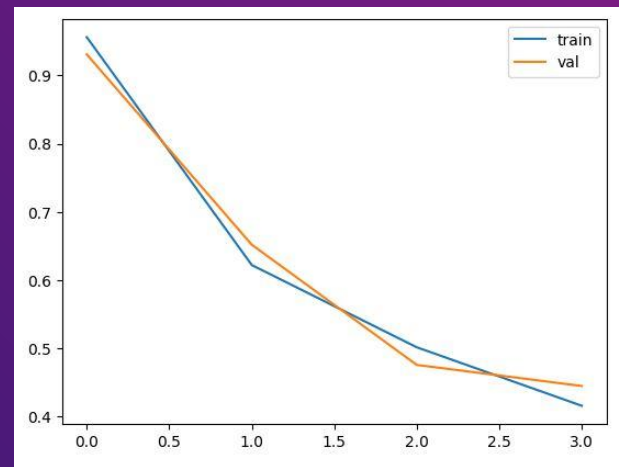
Data Split	Akurasi Rata-Rata
Training 1	0.85
Training 2	0.91
Training 3	0.85
Training 4	0.84
Training 5	0.84
<b>Average</b>	<b>0.8452727272727273</b>

# MODEL LONG SHORT TERM MEMORY

Kurva Akurasi



Kurva Loss





# MODEL LONG SHORT TERM MEMORY

## Evaluasi Model

```
16/16 [=====] - 1s 25ms/step - loss: 0.2600 - categorical_accuracy: 0.9493
16/16 [=====] - 0s 15ms/step - loss: 0.4447 - categorical_accuracy: 0.8506
16/16 [=====] - 0s 16ms/step - loss: 0.4650 - categorical_accuracy: 0.8486

Train: 0.949, val: 0.851, test: 0.849,
```

### Insight:

- Nilai Akurasi Train dan Val → Tinggi
- Nilai Akurasi Test → Cukup Tinggi
- Nilai Akurasi Test < Nilai Akurasi Train → Cenderung Overfit Pada Data Latih (Train)

# TAMPILAN API

## API Data Science Platinum Challenge <sup>1.0.0</sup>

[ Base URL: 127.0.0.1:5000 ]  
[/docs.json](#)

Dokumentasi Data Science Class Platinum Challenge

### LSTM



POST [/post\\_file\\_lstm](#)

POST [/post\\_text\\_lstm](#)

### NN



POST [/post\\_file\\_nn](#)

POST [/post\\_text\\_nn](#)

[Powered by [Flasgger](#) 0.9.5]

# TAMPILAN API

**POST** /post\_text\_lstm

Parameters

Name	Description
<b>text</b> * required string (formData)	USER USER Kaum cebong kapir udah keliat

Execute

Request URL

http://127.0.0.1:5000/post\_text\_lstm

Server response

Code	Details
200	<b>Response body</b> Hasil adalah negative

**POST** /post\_file\_lstm

Parameters

Name	Description
<b>file</b> * required file (formData)	<div>Choose File data_dummy_1.csv</div>

Execute

Request URL

http://127.0.0.1:5000/post\_file\_lstm

Server response

Code	Details
200	<b>Response body</b> DONE

# TAMPILAN API

**POST** /post\_text\_nn

Parameters

Name	Description
text * required string (formData)	- disaat semua cowok berusaha melacak perh

Execute

Request URL

http://127.0.0.1:5000/post\_text\_nn

Server response

Code	Details
200	<div>Response body</div> <div>Hasil adalah negative</div> <div>Response headers</div>

**POST** /post\_file\_nn

Parameters

Name	Description
file * required file (formData)	<div>Choose File</div> data_dummy_1.csv

Execute

Request URL

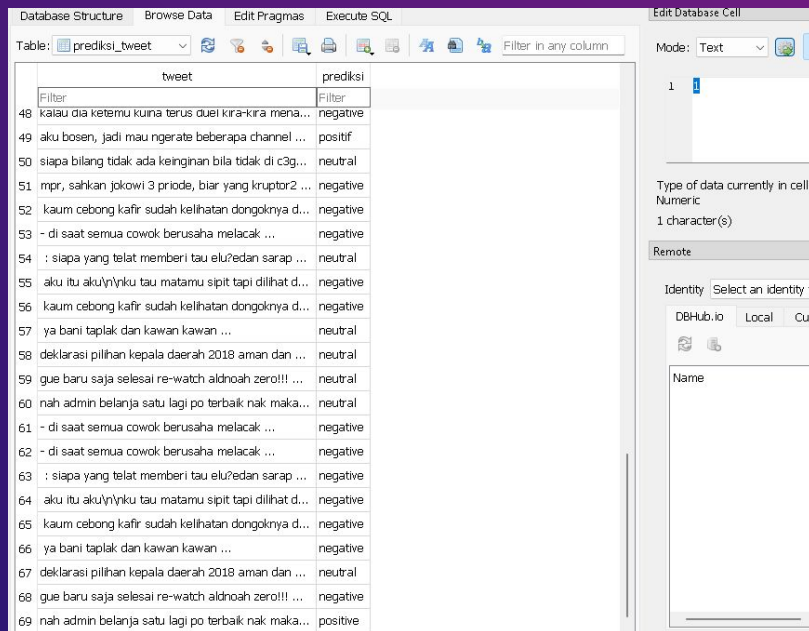
http://127.0.0.1:5000/post\_file\_nn

Server response

Code	Details
200	<div>Response body</div> <div>DONE</div>

# TAMPILAN DATABASE

Semua text yang diinput maupun yang diupload di API akan dimasukkan kedalam database.



tweet	prediksi
Filter	Filter
48 kalau dia ketemu kuina terus duet kira-kira mena...	negative
49 aku bosan, jadi mau ngerate beberapa channel ...	positif
50 siapa bilang tidak ada keinginan bila tidak di c3g...	neutral
51 mpr, sahkan Jokowi 3 priode, biar yang kruptor2 ...	negative
52 kaum cebong kafir sudah kelihatan dongoknya d...	negative
53 - di saat semua cowok berusaha melacak ...	negative
54 : siapa yang telat memberi tau elu?edan sarap ...	neutral
55 aku itu aku\ryku tau matamu sipit tapi dilihat d...	negative
56 kaum cebong kafir sudah kelihatan dongoknya d...	negative
57 ya bani taplak dan kawan kawan ...	neutral
58 deklarasi pilihan kepala daerah 2018 aman dan ...	neutral
59 gue baru saja selesai re-watch aldnoah zero!!! ...	neutral
60 nah admin belanja satu lagi po terbaik nak maka...	neutral
61 - di saat semua cowok berusaha melacak ...	negative
62 - di saat semua cowok berusaha melacak ...	negative
63 : siapa yang telat memberi tau elu?edan sarap ...	negative
64 aku itu aku\ryku tau matamu sipit tapi dilihat d...	negative
65 kaum cebong kafir sudah kelihatan dongoknya d...	negative
66 ya bani taplak dan kawan kawan ...	negative
67 deklarasi pilihan kepala daerah 2018 aman dan ...	neutral
68 gue baru saja selesai re-watch aldnoah zero!!! ...	negative
69 nah admin belanja satu lagi po terbaik nak maka...	positive

# ANALISIS SENTIMEN

Berikut hasil dari analisis dari data text sebelum diprediksi dengan NN dan LSTM.

```
[91] data.groupby(['Abusive']).agg({'Tweet':['count']})
```

Abusive	
0	1
11697	7207

```
data.groupby(['HS']).agg({'Tweet':['count']})
```

HS	
0	1
10928	7976

## Keterangan

HS : Hate Speech

0 : tidak

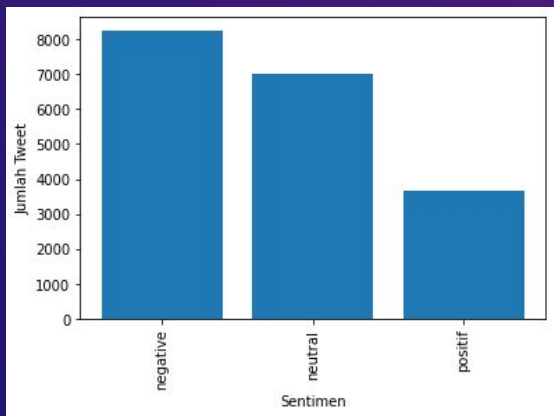
1 : ya

Dari data di sampling, terdapat sekitar 7000 teks dengan kata abusive dan hampir 8000 teks hate speech di mana kedua data tersebut saling beririsan. Artinya, terdapat **±7000** teks dengan kemungkinan hasil prediksi **negatif**.

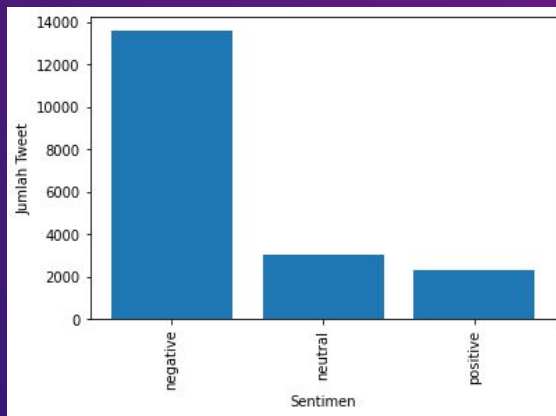
# ANALISIS SENTIMEN

Berikut hasil dari analisis dari data text dengan menggunakan model Neural Network dan LSTM:

## LSTM



## NN



Pada grafik, data kalimat negatif dengan model NN mencapai hampir 14.000 buah, sedangkan dengan model NN hanya mencapai sekitar 8.000 buah

# KESIMPULAN

Model prediksi machine learning dengan metode Neural Network dan LSTM berhasil dibuat dan diluncurkan ke API dengan perbandingan akurasi sebagai berikut:

	Neural Network	LSTM
Data Test	0.85	0.85
Data Train	0.91	0.95

Nilai akurasi kedua model hampir sama besar, namun model LSTM menunjukkan angka yang sedikit lebih tinggi pada data train. Hal ini mengindikasikan bahwa **metode LSTM** adalah metode machine learning yang lebih cocok digunakan untuk memprediksi sentimen pada data teks.

Hal ini turut diperkuat dengan adanya ketidakcocokan antara data mentah dengan hasil prediksi model neural network. Pada data mentah terdapat 7000 lebih tweet yang mengandung kata Abusive dan Hate speech sedangkan model Neural Network menghasilkan hampir 14000 tweet dengan sentimen negatif. Artinya, model Neural Network kurang akurat dalam mengklasifikasi sentimen karena pada kasus ini sentimen negatif yang dikeluarkan melebihi hampir 2 kali lipat dari data yang seharusnya.