

# **Kriteria Pengumpulan Challenge Gold**

Fahmi Indra Setiawan

# 1. API 2 endpoint (input text dan file)

app.py

```
52 @swag_from('docs/text_processing.yml', methods=['POST'])
53 @app.route('/text-processing', methods=['POST'])
54 def text_processing():
55
56     text = request.form.get('text')
57
58     conn = sqlite3.connect('database.db')
59     cursor = conn.cursor()
60     cursor.execute('''INSERT INTO text (before, after)
61     | | | | | VALUES (?, ?)''', (text, text_cleansing(text)))
62     conn.commit()
63     conn.close()
64
65     json_response = {
66         'status': 200,
67         'description': 'Teks yang sudah diproses',
68         'before': text,
69         'after': text_cleansing(text),
70     }
71
72     response_data = jsonify(json_response)
73     return response_data
```

```
75 @swag_from('docs/text_processing_file.yml', methods=['POST'])
76 @app.route('/text-processing-file', methods=['POST'])
77 def text_processing_file():
78
79     file = request.files.getlist('file')[0]
80     df = pd.read_csv(file)
81     text = df.iloc[:, 0].tolist()
82
83     conn = sqlite3.connect('database.db')
84     cursor = conn.cursor()
85
86     cleaned_text = []
87     for t in text:
88         cleaned_text.append(text_cleansing(t))
89         cursor.execute('''INSERT INTO text (before, after)
90         | | | | | VALUES (?, ?)''', (t, text_cleansing(t)))
91
92     conn.commit()
93     conn.close()
94
95     json_response = {
96         'status': 200,
97         'description': 'Teks yang sudah diproses',
98         'before': text,
99         'after': cleaned_text,
100     }
101
102     response_data = jsonify(json_response)
103     return response_data
```

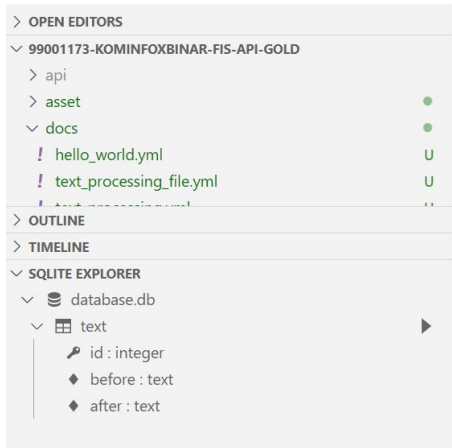
## 2. API dibuat dengan Flask dan Swagger UI

app.py

```
5  from flask import Flask, jsonify
6
7  app = Flask(__name__)
8
9  from flask import request
10 from flasgger import Swagger, LazyString, LazyJSONEncoder
11 from flasgger import swag_from
12
13 app.json_encoder = LazyJSONEncoder
14 swagger_template = dict(
15     info={
16         'title': LazyString(lambda: 'API Documentation for Data Processing and Modelling'),
17         'version': LazyString(lambda: '1.0.0'),
18         'description': LazyString(lambda: 'Dokumentasi API untuk Data Processing and Modelling.'),
19     },
20     host=LazyString(lambda: request.host)
21 )
22 swagger_config = {
23     "headers": [],
24     "specs": [
25         {
26             "endpoint": 'docs',
27             "route": '/docs.json',
28         }
29     ],
30     "static_url_path": "/flasgger_static",
31     "swagger_ui": True,
32     "specs_route": "/docs/"
33 }
34 swagger = Swagger(app, template=swagger_template, config=swagger_config)
```

### 3. SQLite menggunakan modul SQLite 3

database.py







```
database.py > ...
1 import sqlite3
2 import pandas as pd
3 import csv
4
5 conn = sqlite3.connect('database.db')
6 cursor = conn.cursor()
7
8 conn.execute('''CREATE TABLE IF NOT EXISTS text(
9     id INTEGER PRIMARY KEY AUTOINCREMENT NOT NULL,
10    before TEXT NOT NULL,
11    after TEXT NOT NULL)
12 ''')
13
14 conn.commit()
15 conn.close()
```



SQL ▼

< 1 / 1 > 1 - 7 of 7



id	
1	T USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama siap
2	- disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus
3	RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama sia
4	41. Kadang aku berfikir, kenapa aku tetap percaya pada Tuhan padahal aku selalu jatuh berkali-kali. Ka
5	USER USER AKU ITU AKU\n\nKU TAU MATAMU SIPIIT TAPI DILIAT DARI MANA ITU AKU'
6	USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah'
7	USER Ya bani taplak dkk \xf0\x9f\x84\xf0\x9f\x98\x84\xf0\x9f\x98\x84\xf0\x9f\x98\x84'

4

## 4. Output berupa teks yang sudah di-cleansing

### Response body

```
{
  "after": "siapa yang telat ngasih edan sarap bergaul dengan cigax jifla calis sama siapa licew juga",
  "before": "RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama siapa noh licew juga',0,1,0,0,0,0,0,0,0,0,0",
  "description": "Teks yang sudah diproses",
  "status": 200
}
```

[Download](#)

### Response headers

```
connection: close
content-length: 328
content-type: application/json
date: Sun, 04 Aug 2024 12:12:08 GMT
server: Werkzeug/2.2.0 Python/3.8.10
```

### Response body

```
{
  "after": [
    "disaat semua cowok berusaha melacak perhatian lantas remehkan perhatian kasih khusus basic cowok bego",
    "siapa yang telat ngasih edan sarap bergaul dengan cigax jifla calis sama siapa licew juga",
    "kadang berfikir kenapa aku tetap percaya pada tuhan padahal selalu jatuh berkali kali kadang merasa tuhan ninggalkan sendirian ketika orangtuaku berencana berpisah ketika kakakku lebih memilih jadi kristen ketika anak ",
    "matamu sipit tapi dilihat dari mana ",
    "kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah",
    "bani taplak "
  ],
  "before": [
    "disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego ! ! !",
    "RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama siapa noh licew juga",
    "41. Kadang aku berfikir, kenapa aku tetap percaya pada Tuhan padahal aku selalu jatuh berkali-kali. Kadang aku merasa Tuhan itu ninggalkan aku sendirian. Ketika orangtuaku berencana berpisah, ketika kakakku lebih memilih jadi Kristen. Ketika aku anak ter",
    "USER USER AKU ITU AKU\\n\\nKU TAU MATAMU SIPIT TAPI DILIAT DARI MANA ITU AKU",
    "USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah",
    "USER Ya bani taplak dkk \\xf0\\x9f\\x98\\x84\\xf0\\x9f\\x98\\x84\\xf0\\x9f\\x98\\x84"
  ],
  "description": "Teks yang sudah diproses",
  "status": 200
}
```

[Download](#)

### Response headers

```
connection: close
content-length: 1444
content-type: application/json
date: Sun, 04 Aug 2024 12:12:59 GMT
server: Werkzeug/2.2.0 Python/3.8.10
```

## 5. API bisa memproses text cleansing

Name

Description

text \* required  
string  
(formData)

RT USER: USER siapa yang telat ngasih tau

File to upload

Execute

Responses

Curl

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

Request URL

```
http://127.0.0.1:5000/text-processing
```

Server response

Code	Details
200	<div>Response body</div> <pre>{   "after": "siapa yang telat ngasih edan sarap bergaul dengan cigax jifla calis",   "before": "RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul",   "description": "Teks yang sudah diproses",   "status": 200 }</pre>

Name

Description

file \* required  
file  
(formData)

File to upload

Execute

Clear

Responses

Response content type 

application/json

Curl

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

Request URL

```
http://127.0.0.1:5000/text-processing-file
```

Server response

Code	Details
200	<div>Response body</div> <pre>{   "after": [     "disaat semua cowok berusaha melacak perhatian lantas remehkan perhatian kasih khusus basic cowok bego",     "siapa yang telat ngasih edan sarap bergaul dengan cigax jifla calis sama siapa licew juga",     "kadang berfikir kenapa tetap percaya pada tuhan padahal selalu jatuh berkali-kali kadang merasa tuhan meninggalkan sendirian ketika orangtua berencana berpisah ketika kakaku lebih memilih jadi kristen ketika aku anak ten",     "matamu sipit tapi diliat dari mana ",     "kaum cebong kapir udah kelihatan dongkanya dari awal tambah dongkok lagi hahahah",     "bani taplak"   ],   "before": [     "- disaat semua cowok berusaha melacak perhatian gue, loe lantas remehkan perhatian yg gue kasih khusus ke elo, basic elo cowok bego ! ! !",     "RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama siapa noh licew juga",     "Zi. Kadang aku berfikir, kenapa aku tetap percaya pada tuhan padahal aku selalu jatuh berkali-kali. Kadang aku merasa tuhan itu meninggalkan aku sendirian. Ketika orangtua berencana berpisah, ketika kakaku lebih memilih jadi kristen ketika aku anak ten",     "USER USER AKU ITU AKU\\n\\nRU TAU MATAMU SIPIIT TAPI DILIAT DARI MAMA ITU AKU",     "USER User Kaum cebong kapir udah kelihatan dongkanya dari awal tambah dongkok lagi hahahah",     "USER Ya bani taplak dik \\xrf(\\x9f\\x98\\x84\\xr(f)\\x9f\\x98\\x84\\xr(f)\\x9f\\x98\\x84"   ],   "description": "Teks yang sudah diproses",   "status": 200 }</pre>

# 6. report (laporan) dari hasil analisis data teks

## Laporan Hasil Analisis Data

Fahmi Indra Setiawan

### Pendahuluan

#### Latar Belakang

- Dalam era digital saat ini, media sosial menjadi platform utama untuk berinteraksi dan menyebarkan informasi secara cepat dan luas.
- Namun, di balik manfaatnya, media sosial juga menjadi lahan subur bagi penyebaran ujaran kebencian dan bahasa kasar yang dapat merugikan individu maupun kelompok.

#### Tujuan

- Mengidentifikasi kata yang sering menjadi perbincangan oleh pengguna twitter.
- Melakukan identifikasi lebih lanjut terkait kata kasar atau ujaran kebencian yang ada dalam perbincangan tersebut.

### Metode Penelitian

#### Proses Code Function Cleansing

- Mengubah menjadi huruf kecil semua
- Menghilangkan karakter khusus
- Menghapus kata khusus yang sering muncul dalam twitter (RT, User, Mention, Emoji)
- Menghilangkan spasi berlebih
- Mengganti kata alay dengan kata formal
- Menghapus stopwords dalam bahasa Indonesia

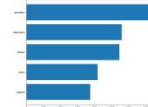
#### Metode statistika dan EDA

- Menggunakan univariate analysis
- Mencari kata umum yang paling sering muncul
- Mencari kata kasar yang paling sering muncul
- Menampilkan dengan visualisasi wordcloud dan horizontal bar chart

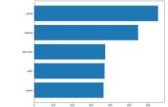
### Visualisasi



Kata yang sering muncul secara umum



Kata kasar yang sering muncul



### Hasil dan Kesimpulan

- Secara umum, masyarakat twitter sering membahas tentang presiden Indonesia pada saat itu yaitu Jokowi. Hal ini disebabkan karena dataset tersebut diambil pada saat tahun politik 2019.
- Selain itu, kata yang sering dibahas juga merupakan kata-kata kasar yang digunakan untuk menunjukkan kebencian kepada presiden Indonesia.
- Kata-kata ujaran kebencian dan bahasa kasar yang sering muncul juga berkaitan dengan sentimen politik kata itu yaitu asing, cebong, komunis, kafir, dan rezim. Dengan ini dapat diambil kesimpulan bahwa ujaran kebencian dan bahasa kasar sering muncul dan perlu dilakukan penanganan lebih lanjut.

Terima Kasih

**Terima Kasih**