



BINAR GOLD CHALLENGE

Tweet

ooo

API UNTUK CLEANSING DATA DAN LAPORAN ANALISIS DATA

Trends



Hatespeech

Presented by Yaya Hidayana



Feedback

LATAR BELAKANG

1

PERTUMBUHAN INTERNET

Pertumbuhan internet di Indonesia sangat pesat dalam beberapa tahun terakhir. Menurut data dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2021, jumlah pengguna internet di Indonesia mencapai 196,7 juta orang atau sekitar 71,7% dari total populasi. Hal ini membuat Indonesia menjadi salah satu negara dengan pengguna internet terbesar di dunia.

2

TWITTER

Salah satu platform media sosial yang sangat populer di Indonesia adalah Twitter. Menurut data StatCounter, pada bulan Maret 2023, Twitter merupakan platform media sosial terpopuler kedua di Indonesia setelah Facebook, dengan pangsa pasar sebesar 12,5%.

3

DAMPAK NEGATIF


Penggunaan sosial media khususnya Twitter juga membawa dampak negatif, seperti meningkatnya hate speech atau ujaran kebencian yang dapat memicu konflik sosial dan memperburuk situasi keamanan di masyarakat.


METODE PENELITIAN


DATA


Name


↑

 abusive.csv


 citation.bib


 data.csv

 new_kamusalay.csv

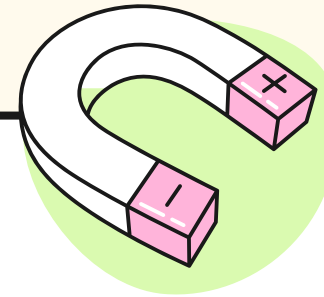
 README.md

TOOLS

 + colab

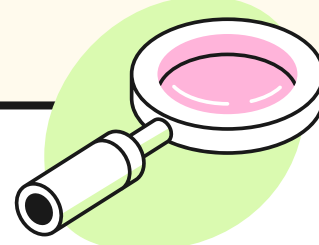


python



EDA

- import pandas as pd
- import seaborn as sns
- import matplotlib.pyplot as plt
- from collections import Counter
- import numpy as np
- from wordcloud import WordCloud, STOPWORDS

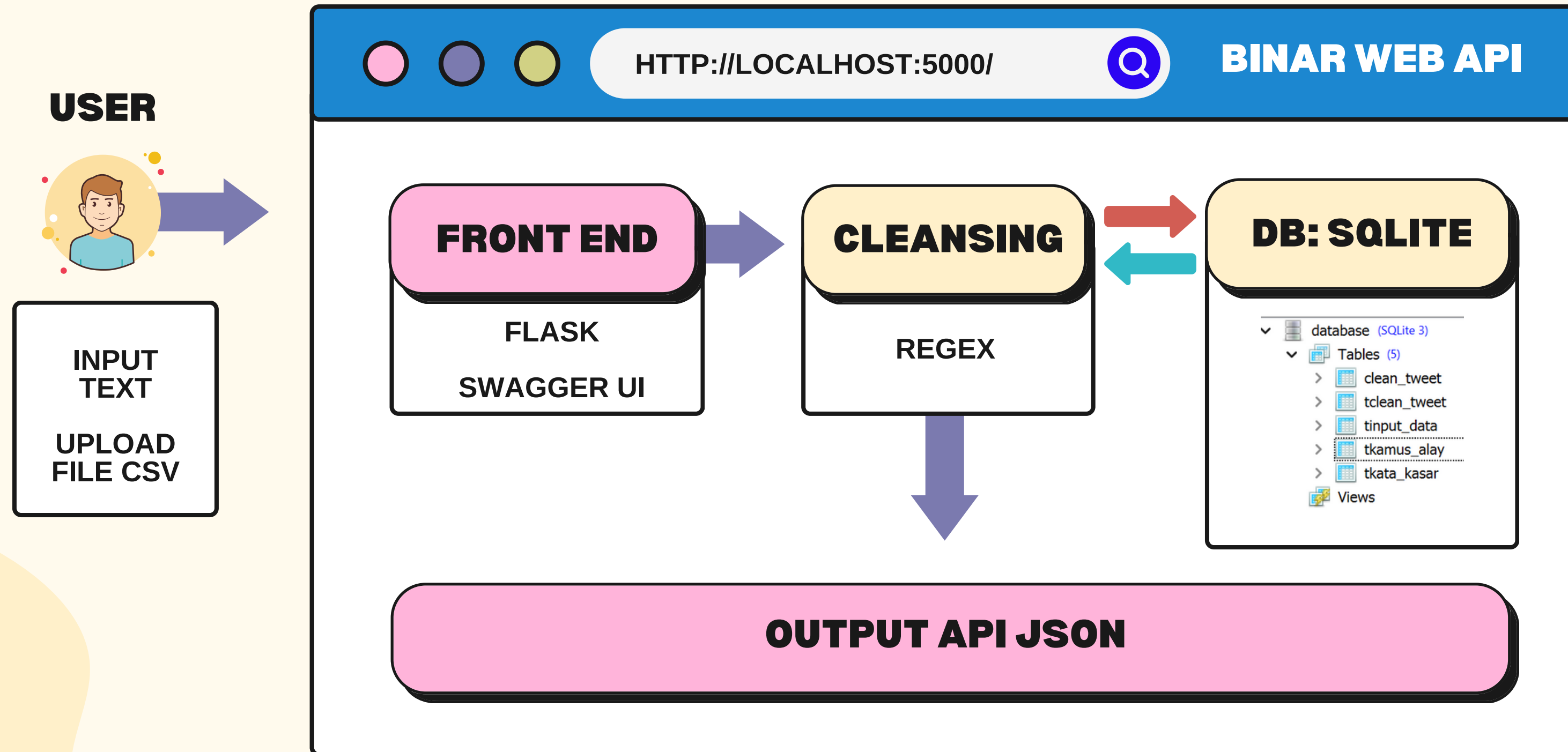


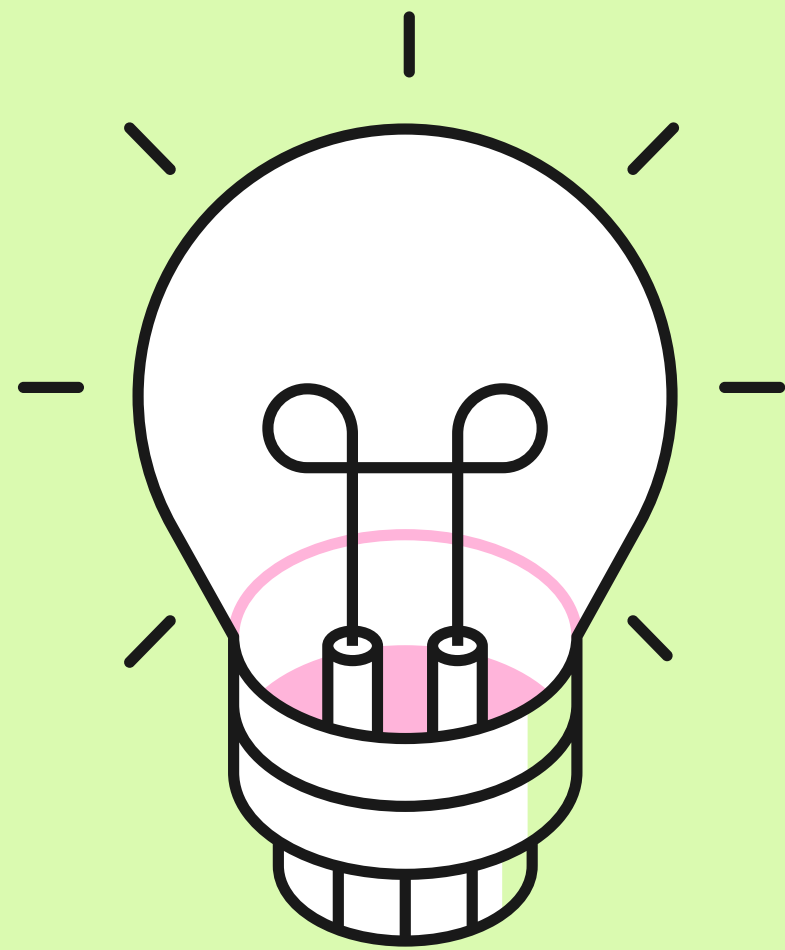
CLEANSING

- Cleansing dengan Regular Expression (Regex)
- Server API dibuat dengan Flask dan Swagger UI
- Penyimpanan data menggunakan SQLite (SQLite 3)



CLEANSING





CONCLUSION

API data cleansing otomatis dibutuhkan dalam postingan Twitter untuk membantu mempercepat dan menyederhanakan proses pembersihan data dari Twitter API. Dalam jumlah data yang besar, tugas membersihkan data secara manual akan memakan waktu dan memerlukan upaya yang cukup besar. Oleh karena itu, penggunaan API data cleansing otomatis dapat membantu mempercepat proses pembersihan data dan mengurangi beban kerja pengguna.

Selain itu, dengan menggunakan API data cleansing otomatis, pengguna dapat memastikan bahwa data yang diperoleh dari Twitter API selalu bersih dan terstruktur dengan baik. Hal ini dapat meminimalkan kesalahan dan memastikan kualitas data yang diperoleh konsisten dan akurat, sehingga mempermudah dalam proses analisa.

LAMPIRAN

INPUT

```
127 # Mendefinisikan dokumentasi API yang digunakan
128 @swag_from("docs/input_data.yml", methods=['POST'])
129 # Mendefinisikan route atau URL endpoint pada web server
130 @app.route('/input_data', methods=['POST'])
131 def test():
132     # Mengambil data input yang diterima melalui form dengan nama "input_data" pada request POST
133     input_txt = str(request.form["input_data"])
134     # Memanggil fungsi text_cleansing() dengan parameter input_txt dan menyimpan hasilnya ke dalam variabel output_txt
135     output_txt = text_cleansing(input_txt)
136
137     # Membuat koneksi ke database dan membuat tabel baru dengan nama "tinput_data" pada database jika belum ada
138     # Menggunakan with statement untuk memastikan bahwa koneksi database ditutup setelah digunakan
139     with sqlite3.connect("database.db") as baca_db:
140         # Mengeksekusi query untuk membuat tabel baru dengan nama tinput_data jika tabel tersebut belum ada
141         baca_db.execute('create table if not exists tinput_data (input_text varchar(255), output_text varchar(255))')
142         # Membuat query SQL untuk memasukkan nilai input_txt dan output_txt ke dalam tabel tinput_data
143         query_txt = 'insert into tinput_data (input_text , output_text) values (?,?)'
144         val = (input_txt,output_txt)
145         # Menjalankan query untuk memasukkan nilai input_txt dan output_txt ke dalam tabel tinput_data
146         baca_db.execute(query_txt,val)
147         # Menyimpan perubahan pada database
148         baca_db.commit()
149
150     # Membuat dictionary baru bernama return_txt yang berisi input_txt dan output_txt
151     return_txt = { "input" :input_txt, "output" : output_txt}
152     # Mengembalikan nilai dalam format JSON dari dictionary return_txt menggunakan fungsi jsonify dari library Flask
153     return jsonify (return_txt)
154
```

Code

Details

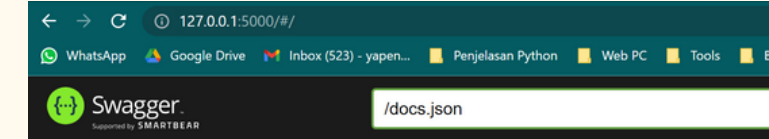
200

Response body

```
{
  "input": "lo anjing",
  "output": "kamu"
}
```

Response headers

```
connection: close
content-length: 38
content-type: application/json
date: Sun, 26 Mar 2023 14:43:03 GMT
server: Werkzeug/2.2.3 Python/3.11.2
```



API Documentation BINAR CHALLENGE

[Base URL: 127.0.0.1:5000]
/docs.json

API Documentation for Text Processing

Please Input the Word

POST /input_data

Please Upload the File

POST /upload_data

UPLOAD

```
163 # Menggunakan decorator untuk mendefinisikan route atau URL endpoint pada web server
164 @app.route('/upload_data', methods=['POST'])
165 def upload_file():
166
167     # Mengambil file yang diupload melalui form dengan nama "upload_data" pada request POST
168     file = request.files["upload_data"]
169
170     # Membaca file CSV menggunakan Pandas dan menyimpan hasilnya ke dalam variabel df_csv
171     df_csv = (pd.read_csv(file, encoding="latin-1"))
172
173     # Mengaplikasikan fungsi text_cleansing() pada kolom 'Tweet' pada dataframe df_csv dan menyimpan hasilnya ke dalam kolom 'new_tweet'
174     df_csv['new_tweet'] = df_csv['Tweet'].apply(text_cleansing)
175
176     # Menyimpan dataframe df_csv ke dalam tabel 'clean_tweet' pada database menggunakan fungsi to_sql dari Pandas
177     df_csv.to_sql("clean_tweet", con=baca_db, index=False, if_exists='append')
178
179     # Menutup koneksi ke database
180     baca_db.close()
181
182     # Mengambil semua nilai dari kolom 'new_tweet' pada dataframe df_csv dan menyimpannya ke dalam variabel cleansing_tweet
183     cleansing_tweet = df_csv.new_tweet.to_list()
184
185     # Membuat dictionary baru bernama return_file yang berisi cleansing_tweet
186     return_file = {
187         'output': cleansing_tweet}
188
189     # Mengembalikan nilai dalam format JSON dari dictionary return_file menggunakan fungsi jsonify dari library Flask
190     return jsonify(return_file)
191
```

Code

Details

200

Response body

```
{
  "output": [
    "di saat semua cowok berusaha melacak perhatian saya kamu lantas remehkan perhatian yang saya kas",
    "siapa yang telat memberi tau kamu saya bergaul dengan cigax jifla calis sama siapa itu licew ju",
    "41 kadang aku berpikir kenapa aku tetap percaya pada tuhan padahal aku selalu jatuh berkali kali",
    "cana berpisah ketika kakakku lebih memilih jadi kristen ketika aku anak ter",
    "aku itu aku dan ku tau matamu tapi dilihat dari mana itu aku",
    "kaum sudah kelihatan dongoknya dari awal tambah lagi haha",
    "ya dan kawan kawan",
    "deklarasi pilihan kepala daerah 2018 aman dan anti hoaks warga dukuh sari jabon",
    "saya baru saja selesai re watch aldnnoah zero paling memang akhirnya 2 karakter utama cowoknya ke",
    "nah admin belanja satu lagi po terbaik nak makan ais kepal milo ais kepal horlicks atau cendol",
    "mantika bank islam senawang",
    "anak lagi kalau sambil",
    "setidaknya gue punya jari tengah buat kamu sebelum gue ukur nyali sama kamu",
    "kaleng malu tidak bisa jawab pe anyaan kami dari 2 hari lalu nyungsep koe uniform resource loca",
    "kalau belajar ekonomi mestinya jago memprivatisasi hati orang aduh ironi",
    "aktor huru hara 98 prabowo si ingin lengserkan pemerintahan jokowi nyata",
    "bu guru enakan jadi atau guru sekolah dasar sih kayaknya menikmati jadi ini guru",
    "lawan bicara gue tidak intelek kayak kamu yang otak tidak punya tentang kencing gue mengakui ha",
    "belakangan ini kok pikiran banget ya",
    "ari sama beki adalah rapi",
    "jadi cowok itu harus gantle kalau tidak gantle itu namanya",
    "alga mnr bom",
    "ya tapi gue jarang mengambek takut wkwk gue kan budak cinta",
    "kalau kamu pasti peluang disakiti nya lebih gede sih",
    "joko widodo dinilai sebagai presiden terlemah dalam sejarah indonesia hal ini terjadi bukan saja",
    "karena ketidakmampuan pemerintahannya menghadapi situasi ekonomi tidak",
  ]
}
```

Response headers

```
connection: close
content-length: 1442827
content-type: application/json
date: Sun, 26 Mar 2023 14:44:21 GMT
server: Werkzeug/2.2.3 Python/3.11.2
```