



BINAR GOLD CHALLENGE

Tweet

ooo

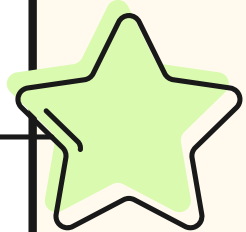
API UNTUK CLEANSING DATA DAN LAPORAN ANALISIS DATA

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

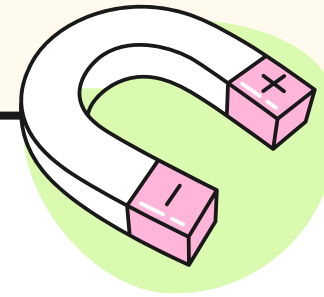
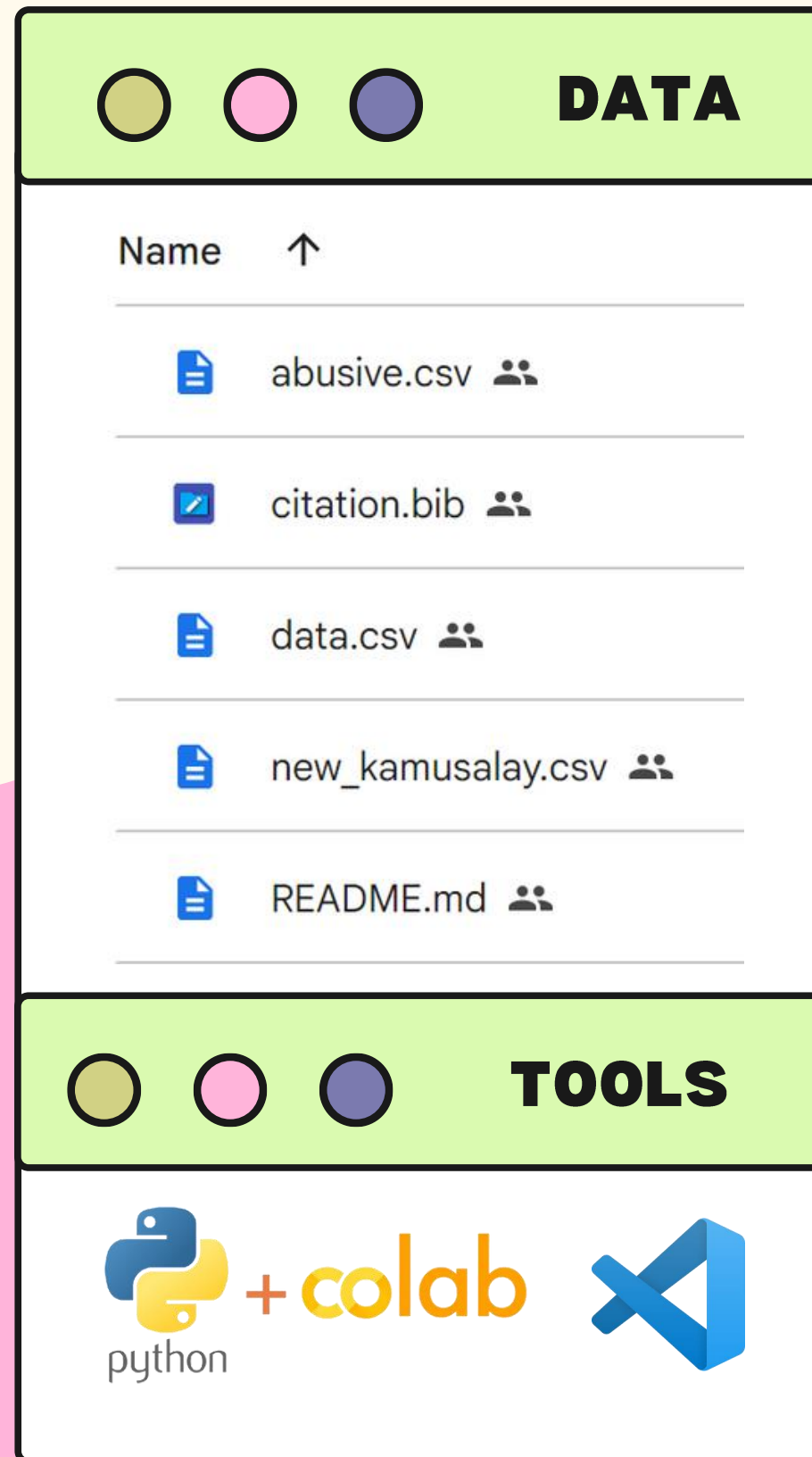
Salah satu platform media sosial yang sangat populer di Indonesia adalah Twitter. Menurut data StatCounter, pada bulan Maret 2023, Twitter merupakan platform media sosial terpopuler kedua di Indonesia setelah Facebook, dengan pangsa pasar sebesar 12,5%.

3

DAMPAK NEGATIF

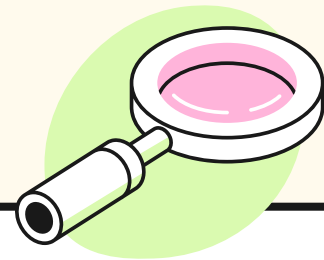
Penggunaan sosial media khususnya Twitter juga membawa dampak negatif, seperti meningkatnya hate speech atau ujaran kebencian yang dapat memicu konflik sosial dan memperburuk situasi keamanan di masyarakat.

METODE PENELITIAN



EDA

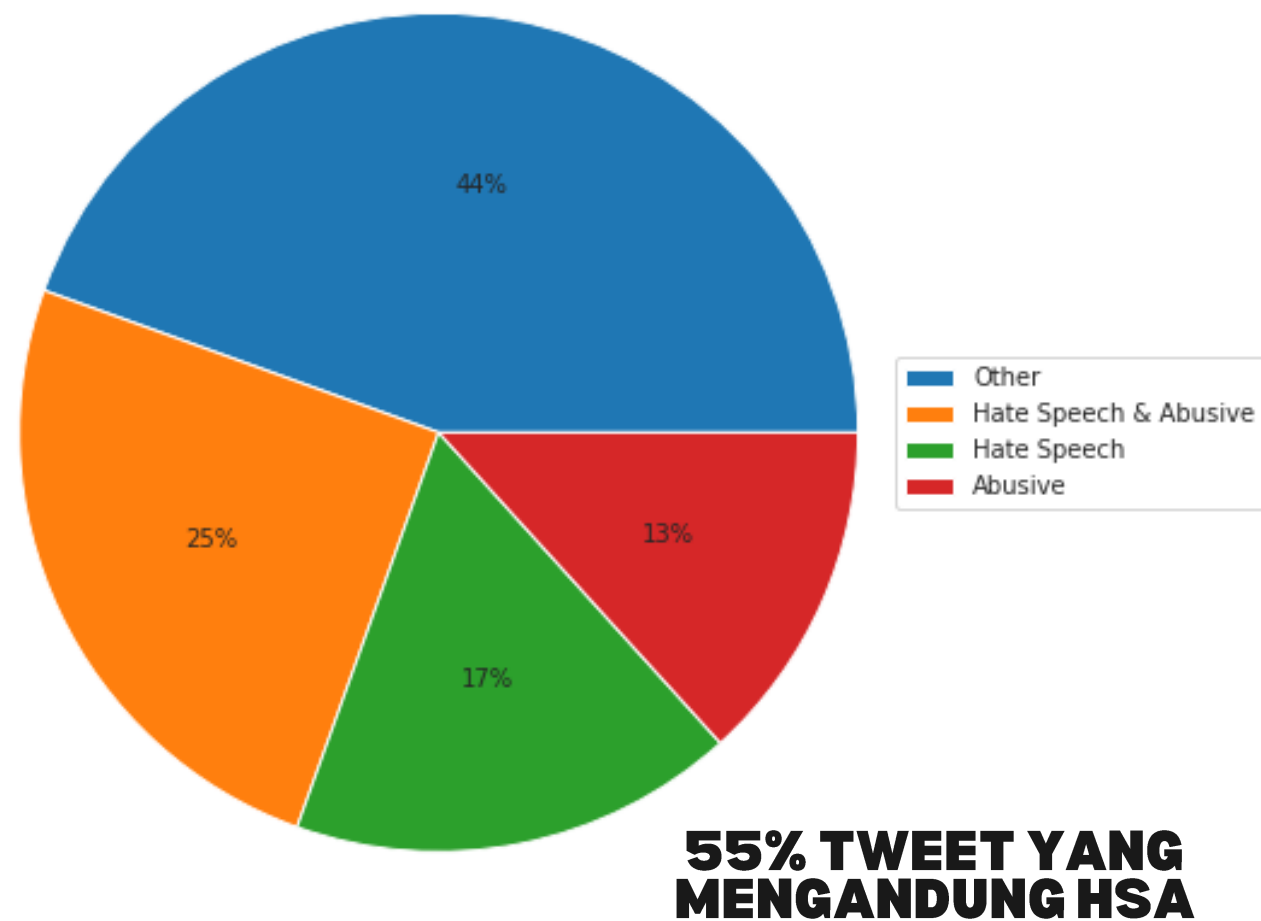
- Memanipulasi data dalam bentuk DataFrame menggunakan pandas.
- Memvisualisasi data statistik dalam bentuk grafik dengan seaborn.
- Memvisualisasi data statistik dalam bentuk grafik dengan matplotlib.
- menghitung frekuensi elemen dalam sebuah list atau tuple dengan collections.
- Menghitung operasi numerik pada data dengan numpy.
- Menggambarkan kata-kata yang sering muncul dalam teks dengan wordcloud.



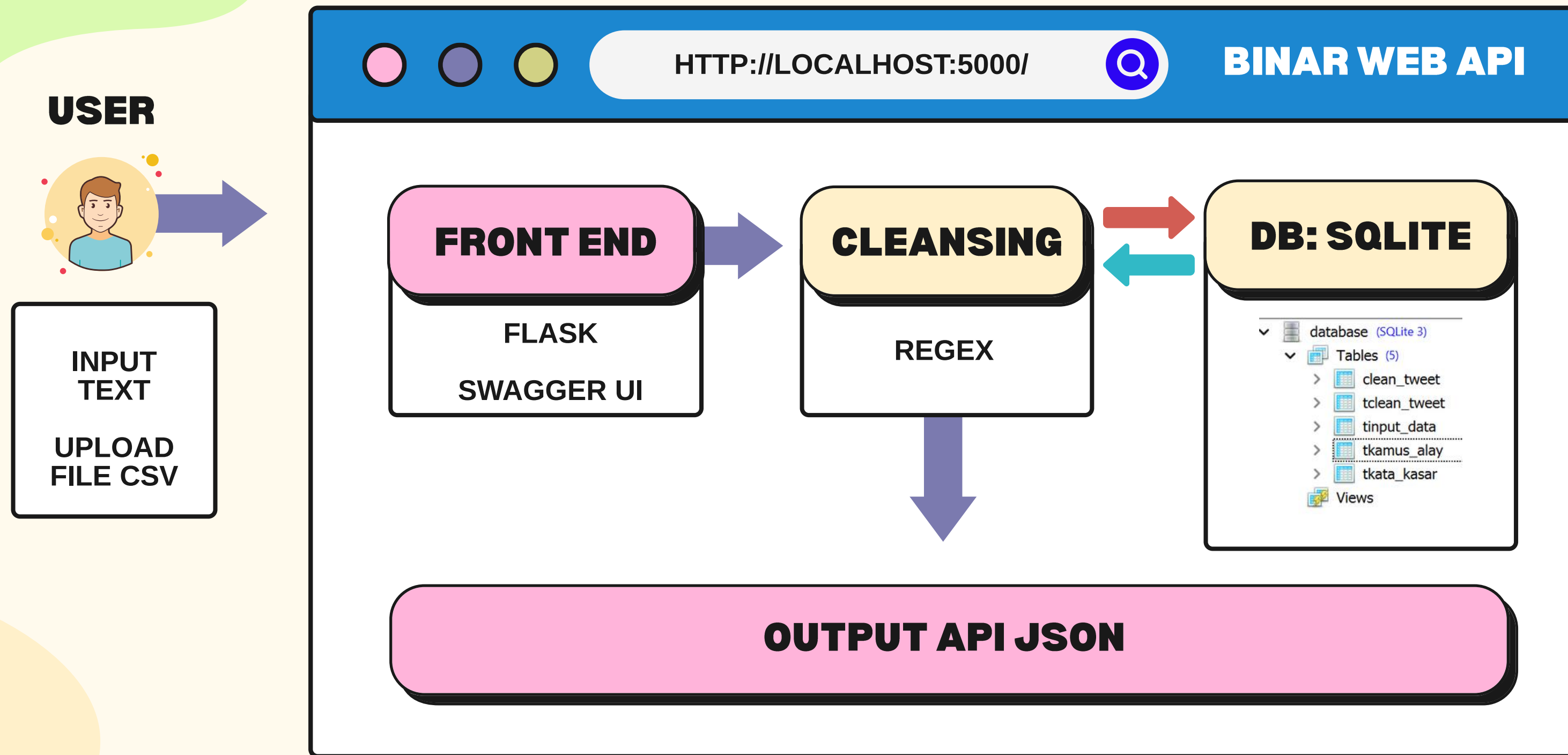
CLEANSING

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

VISUALIZATION



CLEANSING

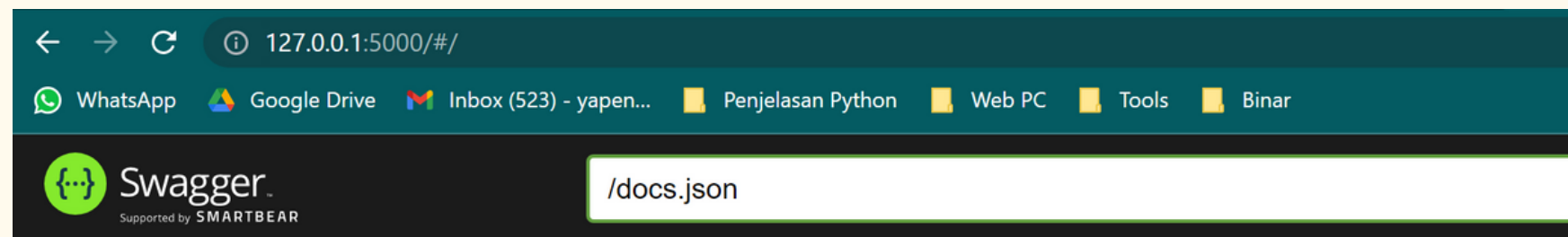


Ket:

Pada proses cleansing, input yang dimasukan pada fontend akan dibersihkan dari simbol, gambar, link, dll menggunakan modul regex, kemudian secara otomatis menghilangkan kata kasar dan kata yang tidak sesuai dengan ejaan baku (alay) sebagaimana data yang tersimpan pada tabel **tkamus_alay** dan **tkata_kasar**, kemudian hasilnya disimpan ke dalam database.

HASIL

Tampilan Frontend Web API



API Documentation BINAR CHALLENGE 1.0.0 // BETA

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

API Documentation for Text Processing

Please Input the Word

POST /input_data

INPUT TEXT

Please Upload the File

POST /upload_data

UPLOAD FILE CSV

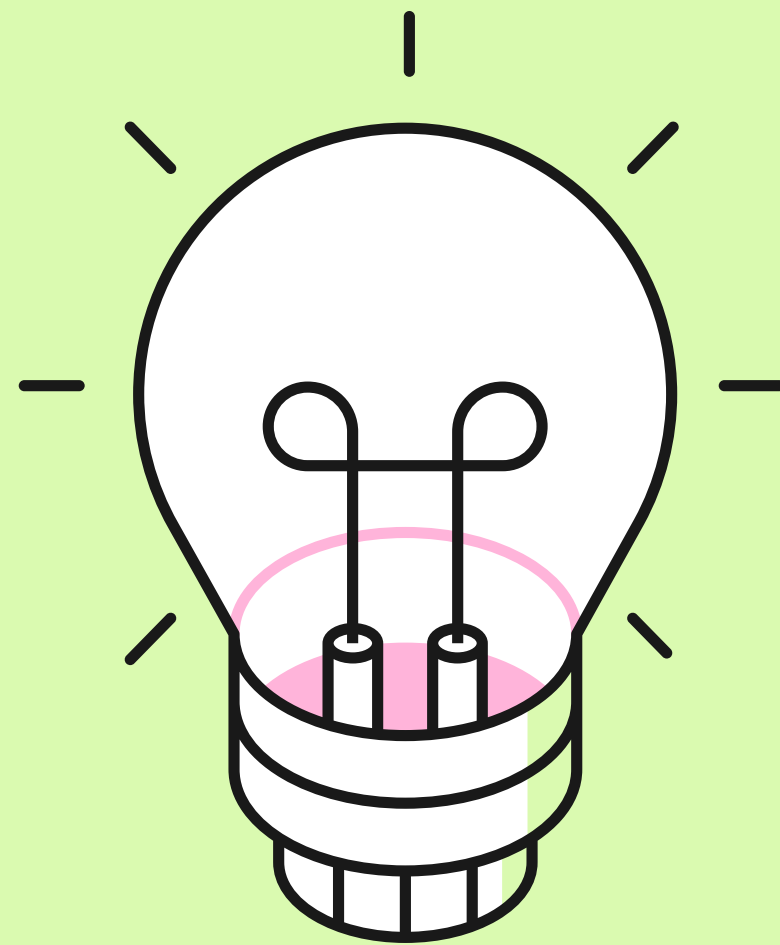
Ket: Pada contoh kasus sederhana, user memasukkan input text secara langsung berisi kata: “Lo Anjing” maka output akan menjadi “Kamu”, karena kata “Lo” akan diubah menjadi bentuk baku yaitu “Kamu”, sedangkan kata “Anjing” akan dihapus karena termasuk kata kasar pada kalimat itu.

OutPut

Code	Details
200	<div><div>Response body</div><div><pre>{ "input": "lo anjing", "output": "kamu" }</pre></div><div>Response headers</div><div><pre>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</pre></div></div>

OutPut

Code	Details
200	<div><div>Response body</div><div><pre>{ "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 j "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 aldoah zero paling memang akhirnya 2 karakter utama cowoknya k "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 loc "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 h "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 w Widodo dinilai sebagai presiden terlemah dalam sejarah indonesia hal ini terjadi bukan sa "karena ketidakmampuan pemerintahannya menghadapi situasi ekonomi tidak",] }</pre></div><div>Response headers</div><div><pre>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</pre></div></div>



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.