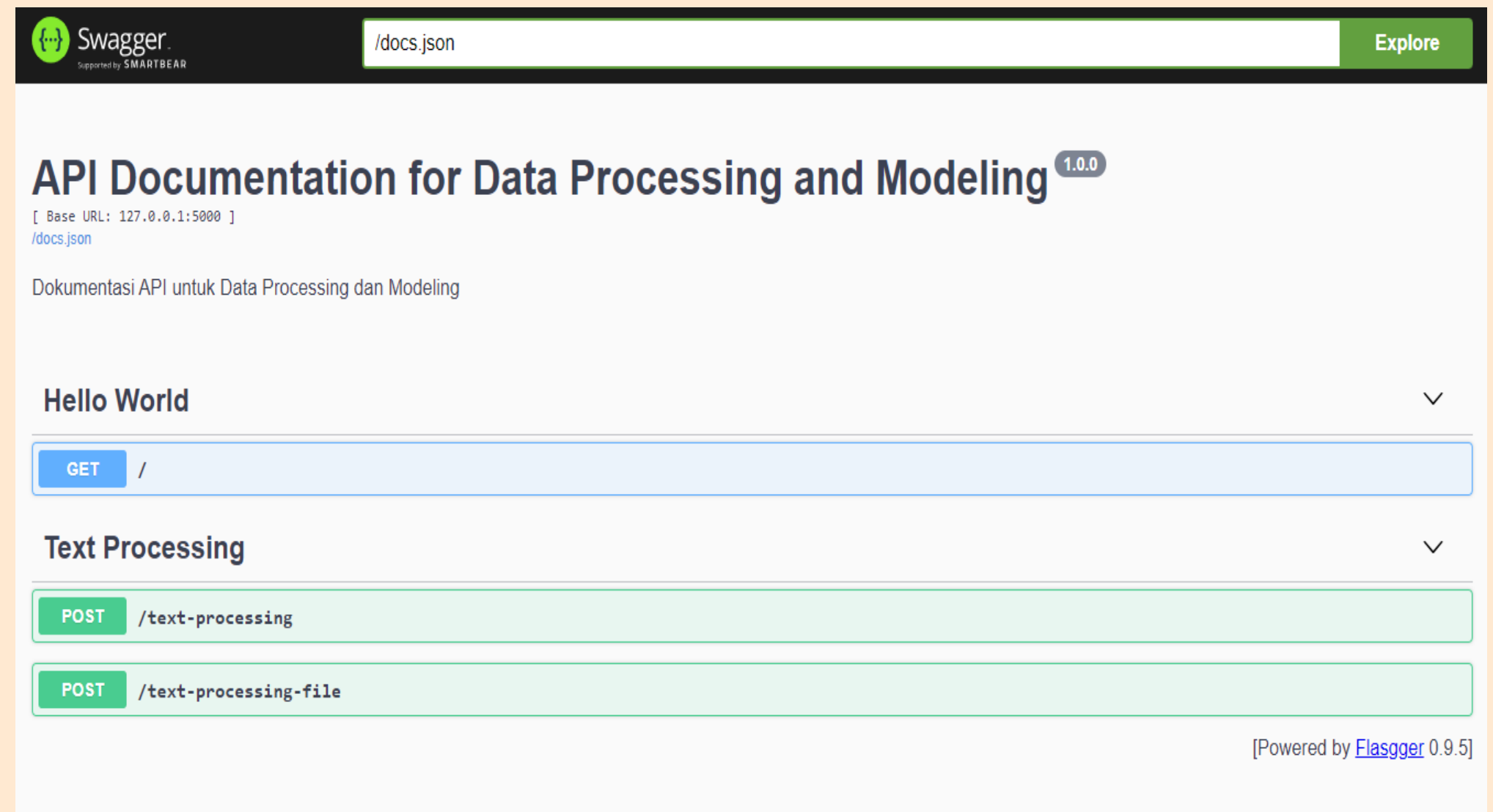


# Challenge Gold Binar Academy

## (API Untuk Cleansing Data)

Nama :  
Ambika Arya Maheswara

Facilitator :  
Bill Clinton Sianipar



# Latar Belakang

Dapat dilihat dari data berikut terdapat banyak kata dari rows data mentah yang sangat berantakan dan tidak nyaman untuk dibaca sehingga perlu dibersihkan agar dapat dianalisis lebih lanjut. Pada latar belakang ini peneliti tertarik untuk membuat suatu API point untuk membersihkan kata-kata dari data tersebut dan menganalisis lebih lanjut.

13164	USER jangan asal ngomong ndasmu. congor lu yg ...	1	1	1	0	0
13165	USER Kasur mana enak kunyuk'	0	1	0	0	0
13166	USER Hati hati bisu :( .g\n\nlagi bosan huft \n...	0	0	0	0	0
13167	USER USER USER USER Bom yang real mudah terdet...	0	0	0	0	0
13168	USER Mana situ ngasih('': itu cuma foto ya kuti...	1	1	1	0	0

13169 rows × 13 columns

# Metode Penelitian

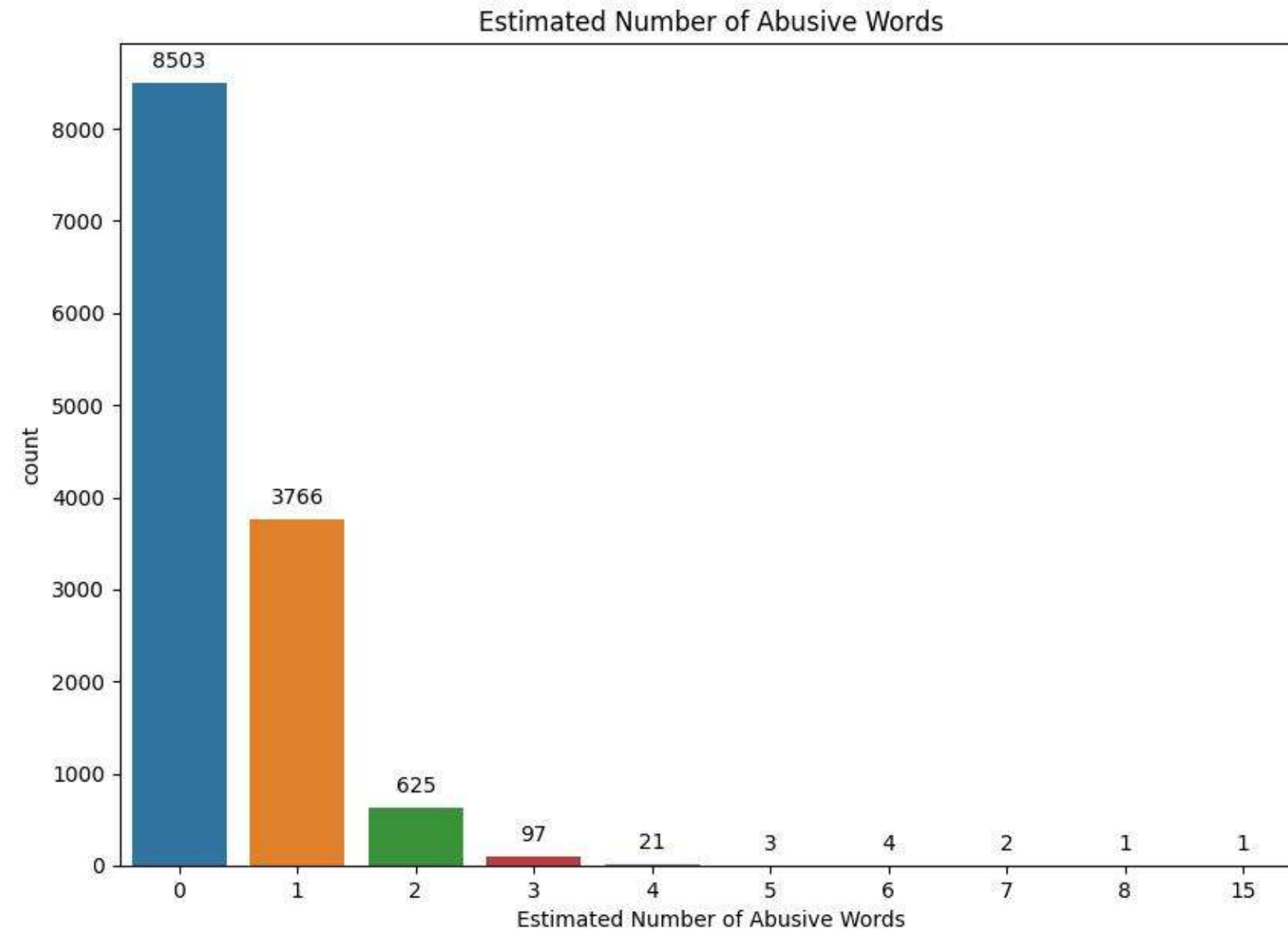
1

Proses code function dapat diakses melalui [https://github.com/mhswra/23001012\\_13\\_Maheswara\\_hate-speech\\_Gold](https://github.com/mhswra/23001012_13_Maheswara_hate-speech_Gold) pada file Maheswara\_Gold\_Challenge.ipynb

2

Exploratory Data Analysis (EDA) menggunakan countplot untuk melihat berapa banyak abusive words pada data secara keseluruhan dan boxplot untuk melihat adanya outlier pada data.

# Hasil Dan Kesimpulan



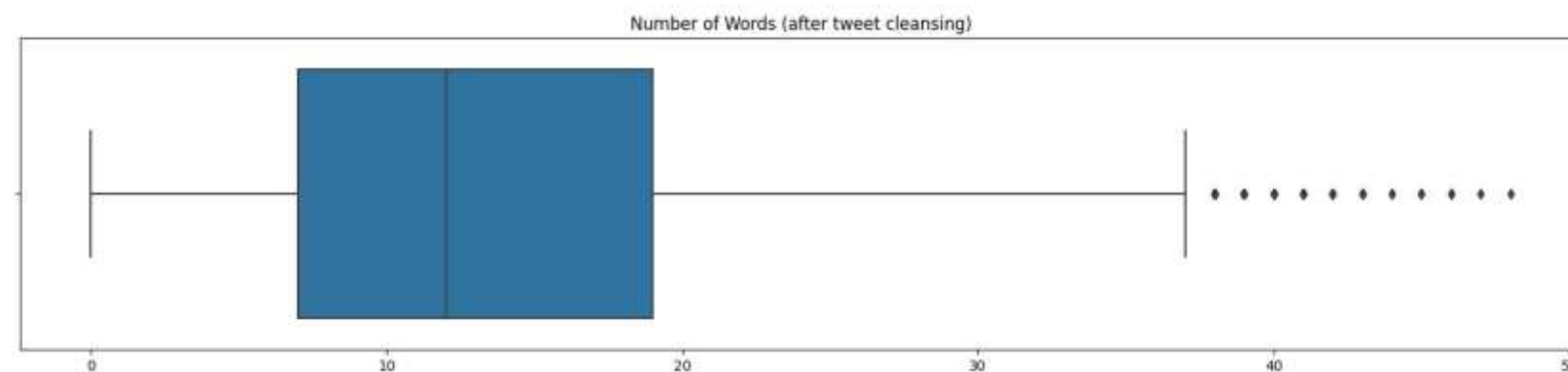
1. Pada tabel countplot, mayoritas tweet terdapat 0 abusive words.

2. Pada tabel countplot, terdapat paling banyak 15 kata abusive dalam 1 tweet.

3. Pada tabel boxplot, rentang tweet yang sudah dibersihkan berada pada estimasi 8 - 18 kata.

4. Pada tabel boxplot, terdapat outlier sebanyak 11 tweet.

5. Kesimpulan, dalam data tersebut mayoritas pengguna twitter di Indonesia bijak dalam menggunakan aplikasi Twitter terbukti dari mayoritas dalam tweet terdapat 0 abusive words dari grafik countplot dan mayoritas masyarakat Indonesia menulis tweet dengan banyak kata terbukti dari boxplot Q3 lebih panjang dari Q1 (postive skewness).



**TERIMA KASIH**