

ANALYZING AND VISUALIZING TWEET DATA: A DATA CLEANSING APPROACH

FIQI FARADILLA
DATA SCIENCE (WAVE 21)



PENDAHULUAN

Media sosial merupakan salah satu platform untuk mengekspresikan pendapat dan berbagi informasi. Twitter adalah satu media sosial yang populer tentang banyaknya bentuk konten berupa ulasan komentar warga twitter dari berbagai kalangan usia mulai dari threads cerita, konten yang tidak relevan atau berbahaya, seperti ujaran kebencian dan kata-kata kasar.

Hal tersebut membuat ketertarikan untuk melakukan penelitian tentang ulasan komentar warga twitter. Salah satu langkah awalnya yaitu pembersihan data (data cleansing). Pembersihan data ini bertujuan untuk menganalisis panjang kata, panjang karakter, korelasi kata dengan ujaran kebencian dan kata - kata kasar serta kata yang sering muncul pada tweet.

METODE PENELITIAN

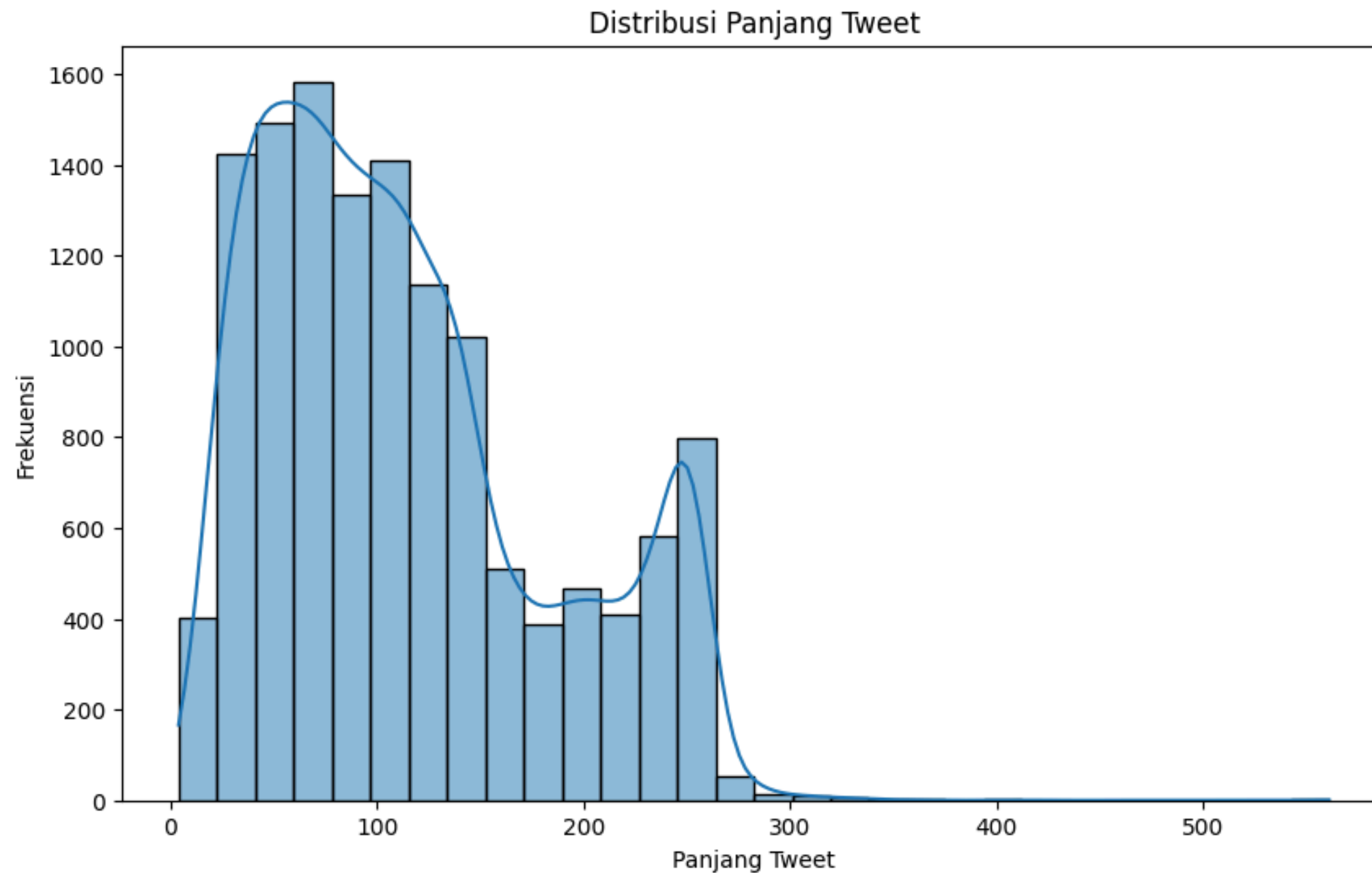
Data pada penelitian ini bersumber dari Indonesian Twitter.

Metode analisis yang dipakai dalam penelitian ini menggunakan Descriptive Analytics karena bertujuan mendeskripsikan pola dari data.

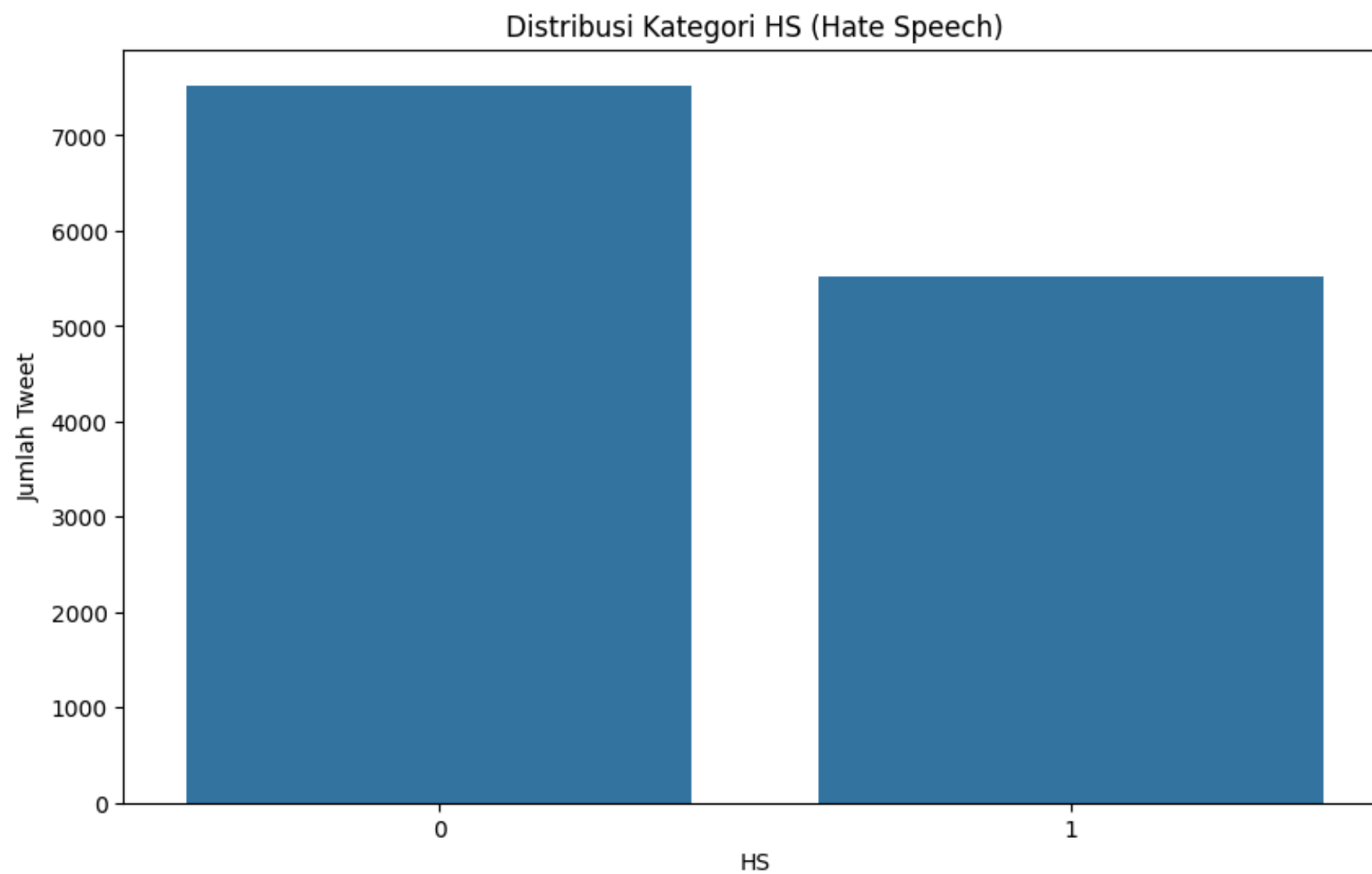
HASIL DAN KESIMPULAN

Tweet yang telah dibersihkan

	Tweet	cleaned_text
0	- disaat semua cowok berusaha melacak perhatian...	disaat semua cowok berusaha melacak perhatian ...
1	RT USER: USER siapa yang telat ngasih tau elu?...	user user siapa yang telat ngasih tau elu edan...
2	41. Kadang aku berfikir, kenapa aku tetap perc...	41 kadang aku berfikir kenapa aku tetap percay...
3	USER USER AKU ITU AKU\n\nKU TAU MATAMU SIPIT T...	user user aku itu aku n nku tau matamu sipit t...
4	USER USER Kaum cebong kapir udah keliatan dong...	user user kaum cebong kapir udah keliatan dong...

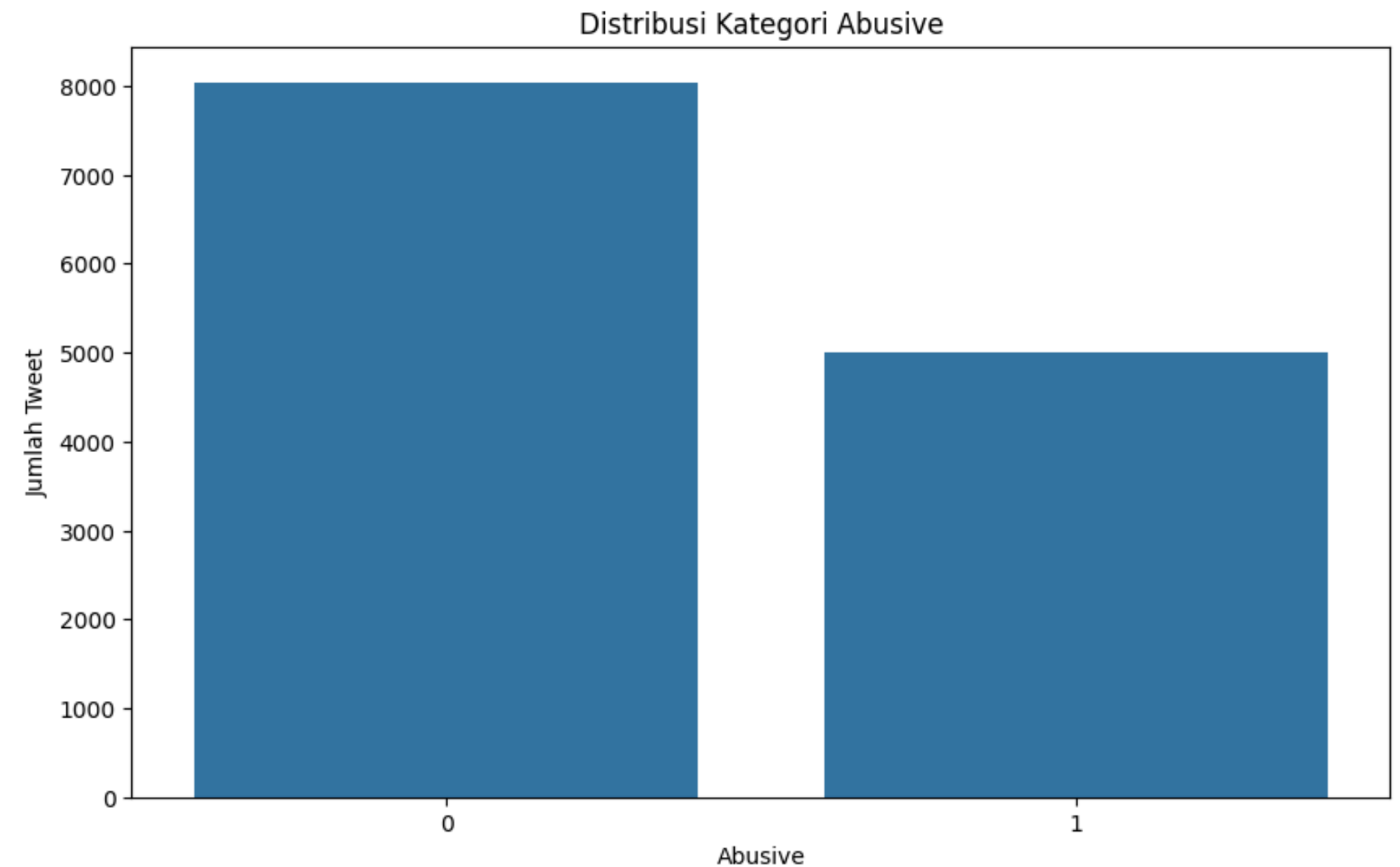


Berdasarkan hasil cleansing data
menunjukkan bahwa panjang
tweet terdiri dari 100 hingga
200 karakter



Tweet yang berisi hate speech (ujaran kebencian) bervariasi tetapi jumlah terbesar berada dalam kategori 0 (tidak mengandung ujaran kebencian)

Tweet yang termasuk dalam kategori abusive (kata kata kasar) jumlah terbesar berada dalam kategori 0 (tidak mengandung kata kata kasar)



Dari hasil diatas dapat disimpulkan bahwa panjang kata tweet sekitar 100 hingga 200 kata yang berisi tentang ulasan baik karena sebagian besar tidak mengandung hate speech dan kata kata kasar. Melainkan kata yang sering muncul berupa “user”, “yg”, “dan”, “di”, dan “itu” yang merupakan kata hubung.

