

---

# Analisis Tweet Ujaran Kebencian dan Abusif pada Twitter Indonesia

---

Algaza Geofarry Susanto

# Pendahuluan

Berdasarkan laporan dari perusahaan agensi media sosial yang berpusat di Inggris, We Are Social, jumlah pengguna Twitter di Indonesia pada 2022 mencapai 18,45 juta dari total 436 Juta pengguna dari seluruh dunia. Sebuah pesan pada Twitter atau lebih dikenal sebagai 'tweet' memiliki berbagai macam variasi, salah satunya adalah dalam bentuk teks. *Tweet* berupa teks yang dikirim oleh masyarakat Indonesia tidak jarang mengandung unsur ujaran kebencian dan abusif.

Penelitian ini dilakukan untuk mengetahui secara mendetail tentang *tweet* dengan unsur ujaran kebencian dan abusif. Analisis secara mendalam dilakukan untuk mengetahui kisaran kata dan huruf pada *tweet*. Kata yang sering muncul dalam *tweet* dengan ujaran kebencian dan abusif juga dapat ditemukan untuk mengetahui karakteristik secara lebih mendalam.

*Tweet* dengan ujaran kebencian dan abusif dapat terbilang cukup berbahaya karena dapat menimbulkan konflik antara sesama pengguna Twitter. Dengan mengenali lebih dalam tentang karakteristik *tweet* yang mengandung unsur ujaran kebencian dan abusif, maka tindakan selanjutnya dapat diterapkan untuk mencegah dampak buruk yang dapat meluas.

# Metode Penelitian

Data yang digunakan sebagai dasar dalam penelitian ini adalah dengan menggunakan data yang tercantum pada [paper berikut](#). Data diambil dengan mengambil beberapa data *tweet* berupa teks dari beberapa sumber terbuka. Data juga kemudian diperkaya dengan data teks *tweet* yang diambil dengan menggunakan Twitter Search API.

Untuk memudahkan dan meningkatkan analisis terhadap data, maka dilakukan pembersihan atau *cleansing* terlebih dahulu terhadap teks non-alphanumeric, *stop words*, dan *slang words*. Analisis Deskriptif kemudian diterapkan pada data untuk mengetahui deskripsi dari data. Pola dan karakteristik penyebaran kata dan huruf juga dapat ditemukan dengan analisis deskriptif.

Analisis deskriptif akan dilakukan dengan melibatkan kolom satu variabel (*univariate*) dan dua variabel (*bivariabel*). Statistik deskriptif akan diterapkan untuk mengetahui persebaran kata dan huruf dari data. Persebaran dari data kemudian juga akan diilustrasikan melalui visualisasi

# Hasil dan Kesimpulan

## Berdasarkan analisis univariat

- Terdapat *Outlier* yang tidak berpengaruh signifikan terhadap analisis data
- Total huruf menunjukkan panjang 0 - 382 huruf
- Total kata menunjukkan panjang 0 - 56 kata
- Jika diurutkan, maka urutan *Tweet* dengan jumlah terbanyak adalah *tweet* dengan unsur sebagai berikut :
  1. Netral (44 %)
  2. Ujaran Kebencian dan Abusif (25 %)
  3. Ujaran Kebencian (17%)
  4. Abusif (14%)
- Total kata dan huruf yang paling banyak adalah *tweet* netral dengan rentang 0-382 kata dan 0-56 Huruf

## Berdasarkan analisis bivariat

- Total kata dan huruf memiliki korelasi positif.
- Total karakter dan huruf yang paling banyak terdapat pada *tweet* dengan unsur netral.

