

Analisis Kata dan hubungan klasifikasi tweet pada Data Tweet Hate Speech dan Abusive

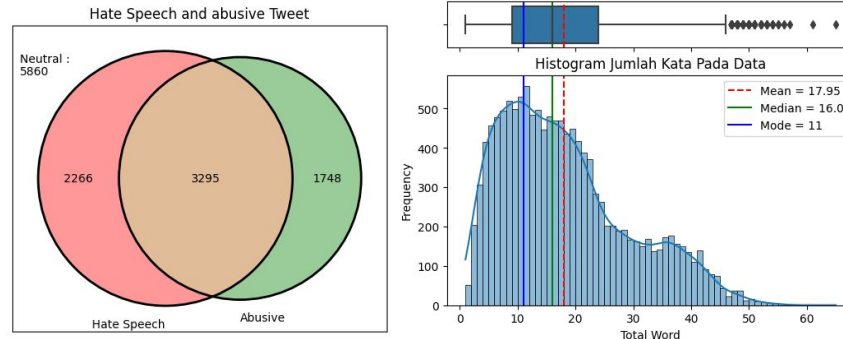
Faiz Agra Kurnia
Institut Teknologi Bandung

Pendahuluan

Indonesia merupakan negara dengan penduduk terbesar keempat di dunia. Pertumbuhan penetrasi internet yang sangat pesat menempatkan Indonesia pada peringkat keempat di dunia dengan pengguna internet terbanyak dan kelima di dunia dengan pengguna media sosial terbanyak.

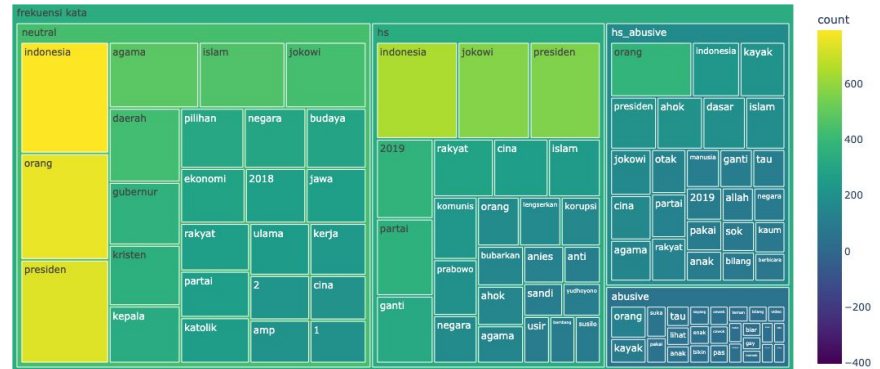
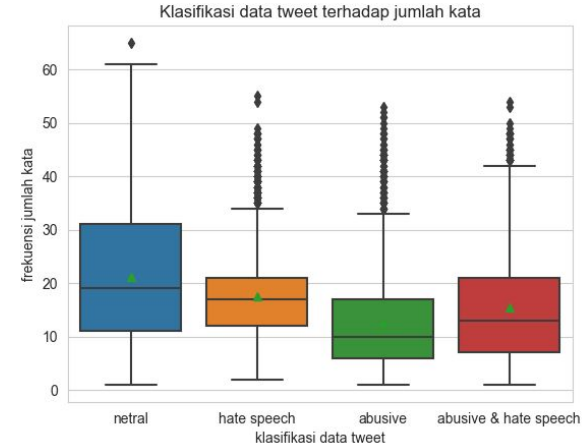
Berangkat dari fakta di atas, ada kecenderungan aktifnya pengguna media sosial di Indonesia untuk berkomentar di dunia maya. Menelaah lebih detail data komentar netizen Indonesia di Twitter menjadi esensial untuk mengetahui karakteristik serta perilaku komentar daring yang dilakukan netizen Indonesia di media sosial.

Oleh karena itu penelitian ini bertujuan untuk menganalisis panjang kata, kata dan juga hubungan klasifikasi data untuk melihat pola terjadi pada komentar tweet yang dilontarkan oleh netizen Indonesia. Harapannya, hasil analisis ini dapat menjadi bahan pertimbangan berbagai pihak kedepannya.



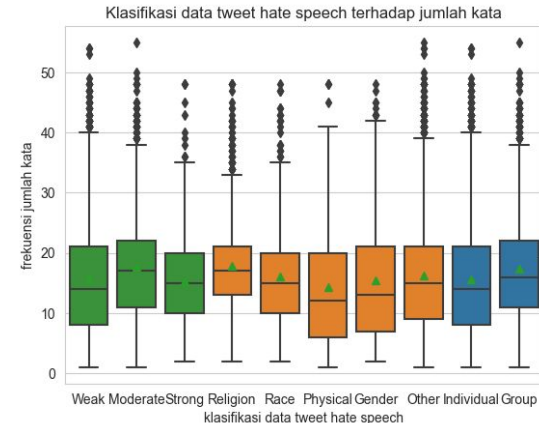
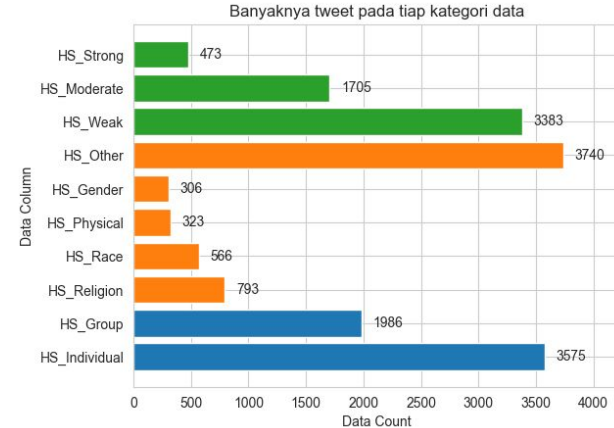
Hasil dan Kesimpulan

- Rata rata jumlah kata pada data netral memiliki jumlah paling tinggi dengan rata rata 21 - 22 kata.
- Outlier jumlah kata pada data netral jauh lebih sedikit jika dibandingkan dengan data dengan kategori lainnya.
- Kata yang paling sering muncul pada data netral:
 - Indonesia, orang, presiden, agama, islam, jokowi, daerah, gubernur, kristen
- Kata yang paling sering muncul pada data hate speech:
 - Indonesia, jokowi, presiden, 2019, partai, ganti, rakyat, cina, islam
- Kata yang paling sering muncul pada data abusive:
 - Orang, kayak, suka, pakai, tau, lihat, anak, sayang, enak, bikin



Hasil dan Kesimpulan

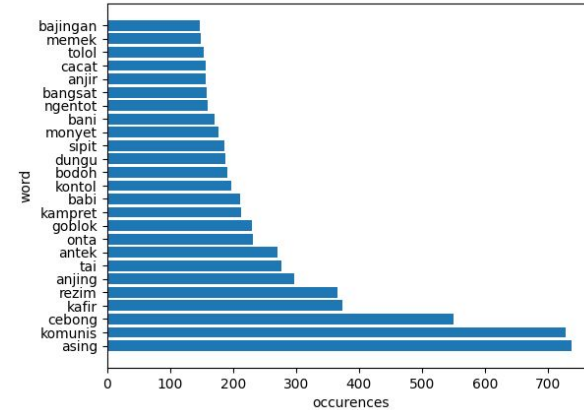
- Kategori dari hate speech dengan frekuensi paling banyak adalah *Other*
- Kategori dari hate speech dengan frekuensi paling sedikit adalah *Gender*
- Level dari hate speech dengan frekuensi paling banyak adalah *weak*
- Target dari hate speech dengan frekuensi paling banyak adalah *individual*
- Rata - rata jumlah kata terbanyak pada kategori hate speech berada pada kategori *religion*
- Rata - rata jumlah kata terbanyak pada target hate speech berada pada target *group*
- Rata - rata jumlah kata terbanyak pada level hate speech berada pada level *moderate*



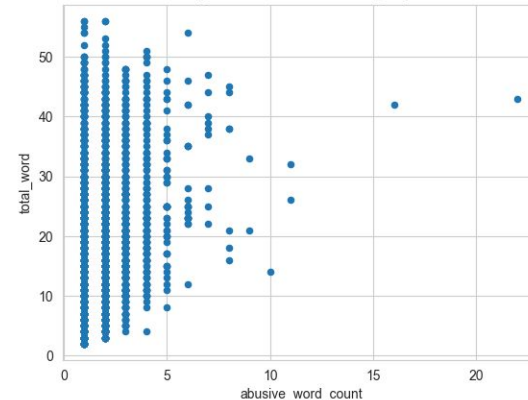
Hasil dan Kesimpulan

- Berdasarkan kamus kata abusive, kata abusive yang paling sering muncul pada seluruh data adalah kata asing yang muncul 738 kali
- Berdasarkan Bivariate Analysis:
 - Dalam Descriptive Statistic korelasi antara jumlah kata abusive dan jumlah kata rendah
 - Dalam visualisasi menunjukkan:
 - Jumlah kata yang tergolong abusive berdasarkan kamus abusive pada setiap tweet abusive tidak memiliki korelasi terhadap jumlah kata pada tiap tweet
 - Jumlah kata yang tergolong abusive berdasarkan kamus abusive paling banyak adalah 22 kata abusive
 - Rata rata kata abusive muncul 1-2 kali pada tiap kalimat

25 Most Common Abusive Word Occurrences in Tweet



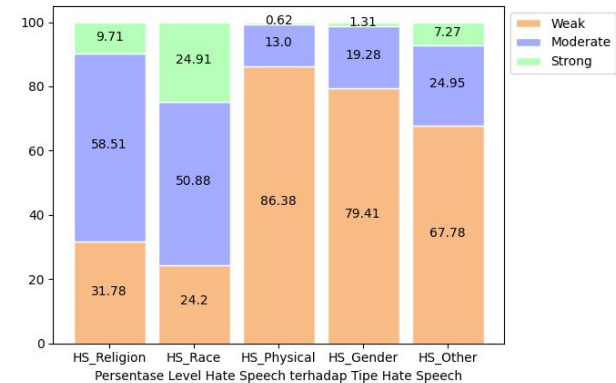
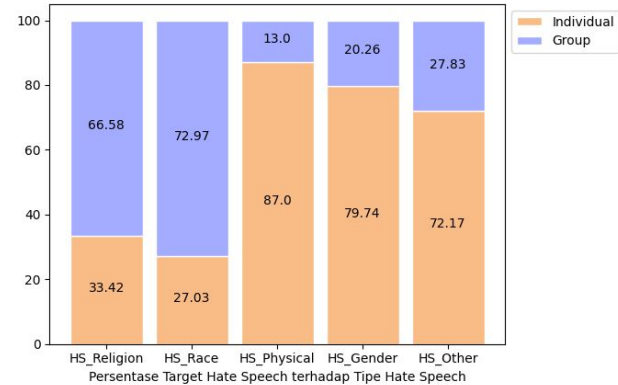
Persebaran jumlah kata abusive dengan jumlah kata



Hasil dan Kesimpulan

Berdasarkan analisis data didapatkan data sebagai berikut :

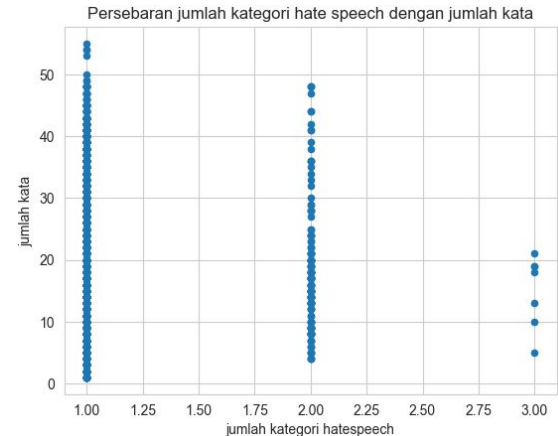
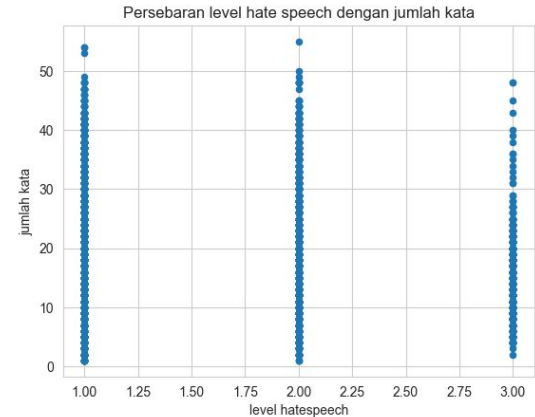
- Berdasarkan Bivariate Analysis:
 - Dalam visualisasi menunjukkan:
 - Tweet dengan klasifikasi *hate speech religion* dan *race* mayoritas ditargetkan ke kelompok orang, lalu untuk *hate speech physical*, *gender* dan *other* mayoritas ditargetkan ke individu.
 - Tweet dengan klasifikasi *hate speech religion* dan *race* mayoritas tergolong dalam *level moderate*, lalu untuk *hate speech physical*, *gender* dan *other* mayoritas mayoritas tergolong dalam *level weak*



Hasil dan Kesimpulan

Berdasarkan analisis data didapatkan data sebagai berikut :

- Berdasarkan Bivariate Analysis:
 - Dalam Descriptive Statistic menunjukkan level hate speech dengan jumlah kata memiliki korelasi yang sangat rendah jumlah kata dengan jumlah kategori muncul
 - Dalam visualisasi menunjukkan:
 - Jumlah kata tidak mencerminkan tingkat level hate speech, rata - rata tingkat level hate speech di kisaran 1-2
 - Jumlah kata tidak mencerminkan jumlah tipe kategori hate speech, jumlah kategori hate speech paling banyak pada 1 tweet adalah 3 kategori, rata - rata setiap tweet hanya memiliki 1 kategori hate speech



Hasil dan Kesimpulan

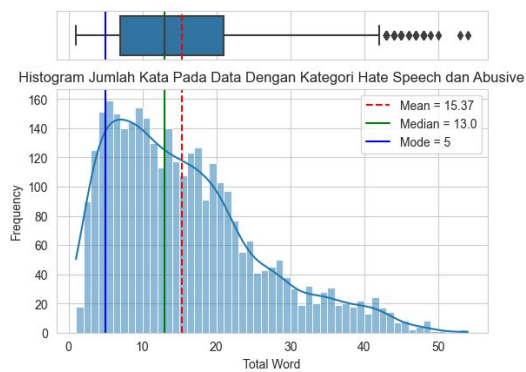
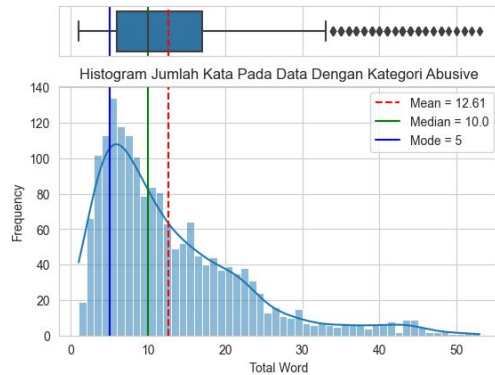
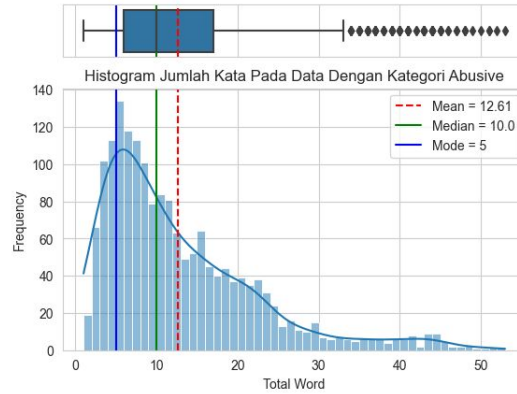
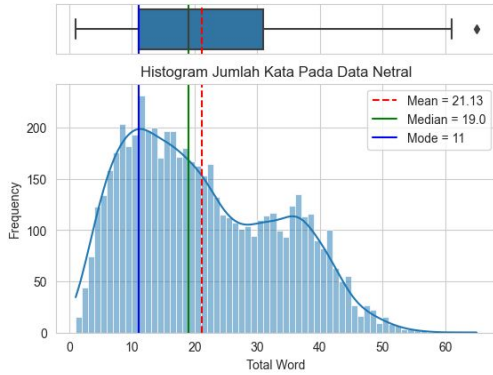
- Dari hasil di atas dapat disimpulkan data didominasi oleh data netral, pada data netral kata yang paling sering muncul adalah “indonesia”, “orang”, “presiden”, “agama”, “islam”, “jokowi”, “daerah”, “gubernur” dan “kristen”.
- Pada data yang tergolong abusive kata yang paling sering dipakai adalah kata “asing”.
- Pada data Hate Speech, klasifikasi hate speech yang paling banyak muncul pada kategori adalah *hate speech other* diikuti dengan *hate speech religion*, pada level adalah weak dan target adalah individual.
- *Hate speech religion* dan *race* mayoritas ditujukan kepada kelompok dengan level *moderate*, sedangkan untuk *hate speech physical*, *gender* dan *other* mayoritas ditujukan kepada individu dengan level *weak*
- Korelasi rendah antara jumlah kata dengan jumlah kata abusive yang dipakai, level hate speech maupun jumlah kategori hate speech yang muncul.

Daftar Pustaka

<https://www.kaggle.com/datasets/ilhamfp31/indonesian-abusive-and-hate-speech-twitter-text>

<https://www.kaggle.com/datasets/oswinrh/indonesian-stoplist?resource=download>

Lampiran



Lampiran

