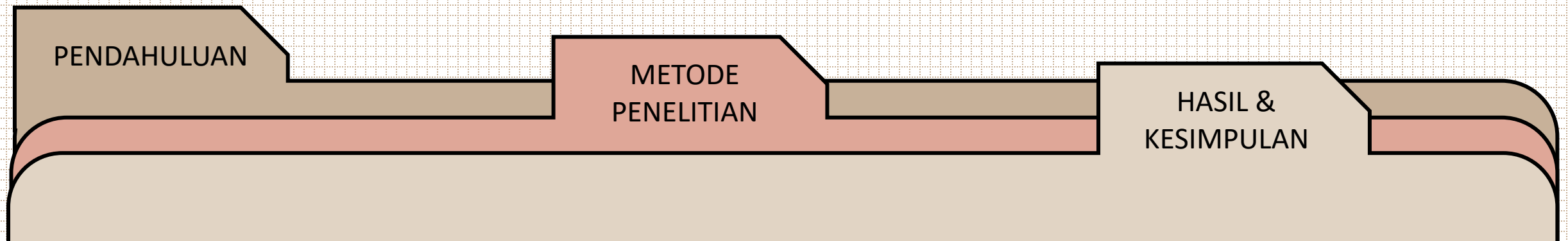
IMPLEMENTASI TEXT MINING TERKAIT HATE SPEECH PADA SOSIAL MEDIA TWITTER

Afdhah Nur Riadhoh



PENDAHULUAN

Media sosial adalah salah satu platform yang banyak digunakan oleh masyarakat untuk berinteraksi, berbagi informasi, dan menyampaikan pendapat. Namun, media sosial juga dapat menjadi sarana untuk menyebarkan ujaran kebencian atau hate speech.

Hate speech, yaitu menyerang atau merendahkan kelompok tertentu, seperti suku, agama, ras, etnis, golongan, atau lainnya. Ujaran kebencian dapat menimbulkan dampak negatif bagi individu maupun masyarakat, seperti diskriminasi, kekerasan, perpecahan, dan lain sebagainya.

METODE PENELITIAN

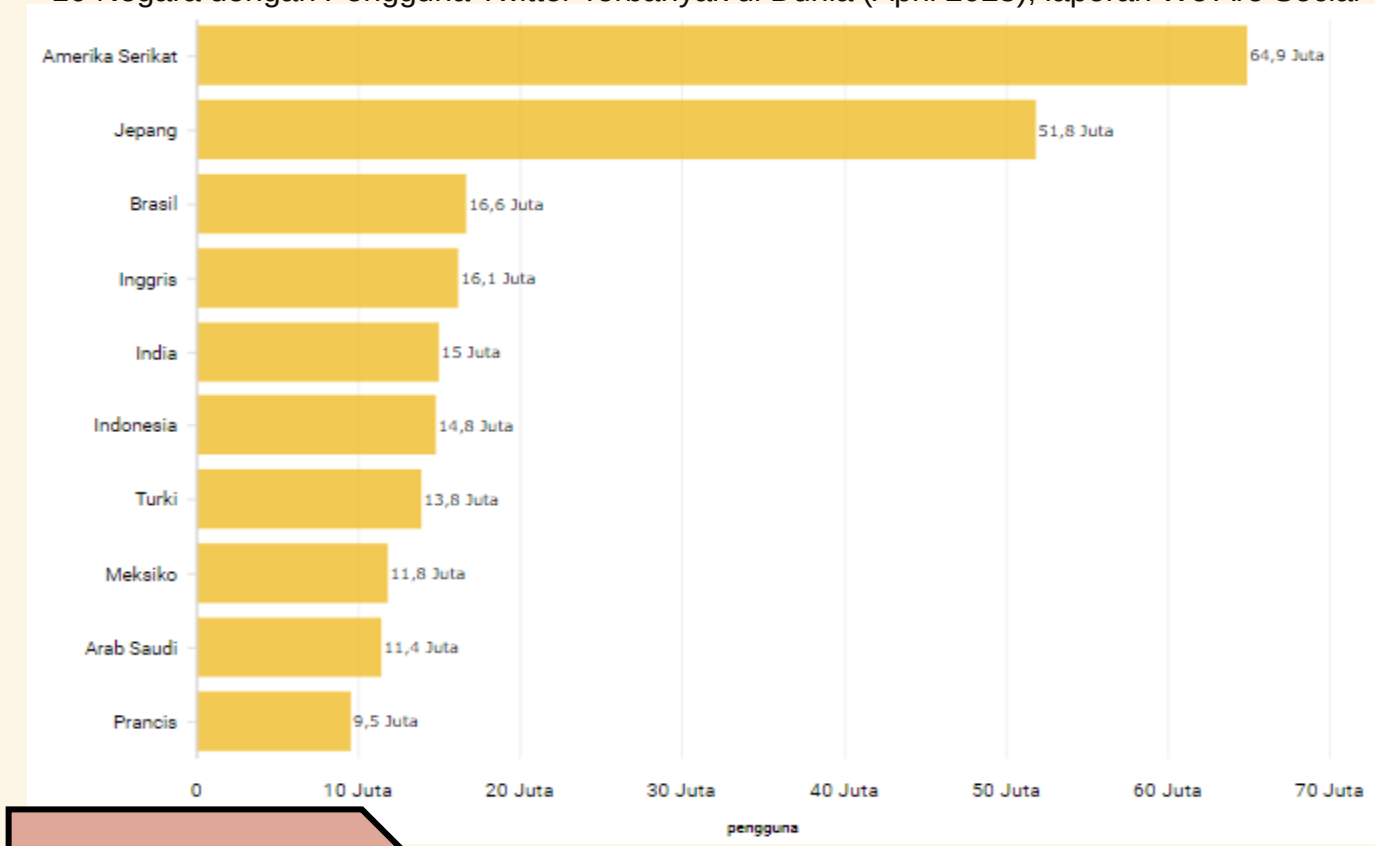
HASIL & KESIMPULAN

PENDAHULUAN

Ada berbagai macam media sosial yang ada saat ini dan masih banyak penggunanya yang aktif, salah satunya adalah Twitter.

Namun, Twitter juga rentan terhadap ujaran kebencian yang dilontarkan oleh para pengguna.

10 Negara dengan Pengguna Twitter Terbanyak di Dunia (April 2023), laporan We Are Social



METODE PENELITIAN

HASIL & KESIMPULAN

PENDAHULUAN

Text Mining

Untuk mengatasi masalah ujaran kebencian di Twitter, salah satu metode yang dapat digunakan adalah *text mining*. *Text mining* adalah metode untuk mengekstrak informasi dan pengetahuan yang bermakna dari sejumlah besar sumber data berupa teks, seperti dokumen Word, PDF, teks kutipan, berita online, komentar dan lain-lain.

Dengan text mining, kita dapat menemukan pola dari data Twitter yang banyak dan beragam, seperti *hate speech*. Diharapkan dapat memberikan manfaat bagi pihak-pihak yang terkait dengan masalah ujaran kebencian di media sosial, seperti pemerintah, lembaga swadaya masyarakat, akademisi, dan masyarakat umum.

METODE PENELITIAN

HASIL & KESIMPULAN

Deskripsi Data

SUMBER DATA

- berita/dokumen (teks) pada sosial media Twitter yang termasuk dalam kategori hate speech
- data teks dari twitter yang digunakan (data sekunder) merupakan data yang disediakan oleh Binar Academy (pemiliki asli : Muhammad Okky Ibrohim dan Indra Budi. 2019. Multi-Labeled Hate Speech and Abusive Indonesian Twitter Text.

Isi Data

- HS (Hate Speech)
- Abusive
- HS_Individual
- HS_Group
- HS_Religion
- HS_Race
- HS_Physical
- HS_Gender
- HS_Other
- HS_Weak
- HS_Moderate
- HS_Strong

Deskripsi Data

Penjelasan Isi Data

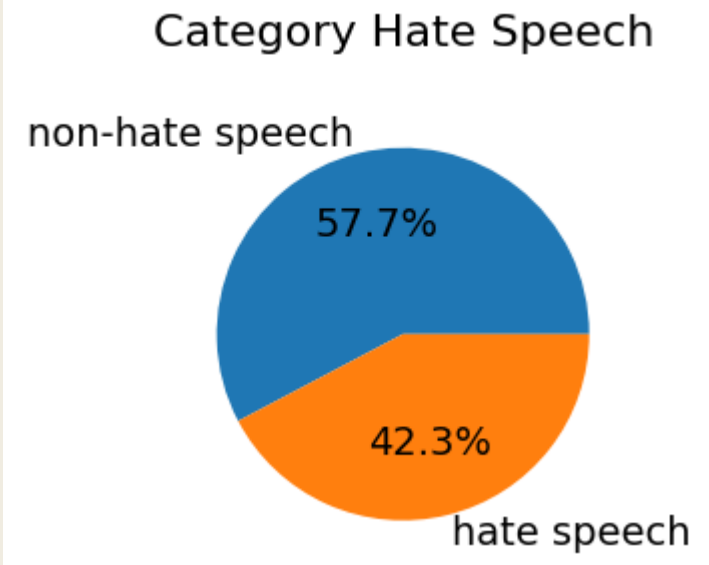
HS : label ujaran kebencian;
Abusive : label bahasa yang kasar;
HS_Individual : ujaran kebencian yang ditujukan kepada individu;
HS_Group : ujaran kebencian yang ditargetkan pada suatu kelompok;
HS_Religion : ujaran kebencian yang berkaitan dengan agama/keyakinan;
HS_Race : ujaran kebencian terkait ras/etnis;
HS_Physical : ujaran kebencian yang berhubungan dengan fisik/cacat;
HS_Gender : ujaran kebencian terkait gender/orientasi seksual;
HS_Other : kebencian yang berhubungan dengan makian/fitnah lainnya;
HS_Weak : ujaran kebencian yang lemah;
HS_Moderat : ujaran kebencian yang moderat;
HS_Strong : ujaran kebencian yang kuat.

Untuk setiap label, 1 berarti tweet termasuk label tersebut, 0 berarti tweet tidak termasuk dalam label tersebut.

Tahapan Analisis Data

1. Cleaning : pembersihan dan mengurangi informasi data yang tidak perlu.
2. Tokenizing : sebuah dokumen teks terdiri dari sekumpulan kalimat, proses tokenization memecah dokumen menjadi beberapa bagian dari kata-kata yang disebut token
3. Filter/stopwords: tahap filtering digunakan algoritma stop-word removal. Penghapusan kata bertujuan untuk menghilangkan kata-kata yang tidak penting. Contoh kata: "yang", "dan", "Anda", "sampai“.
4. Normalization: memperbaiki kata singkatan
5. Stemming : proses penguraian bentuk suatu kata (yang memiliki kata imbuhan) ke dalam bentuk kata dasarnya
6. Analisis untuk melihat topik utama dan kata yang melekat menggunakan visualisasi *wordcloud* dan asosiasi kata, serta visualisasi lainnya untuk melihat gambaran umum dari data.

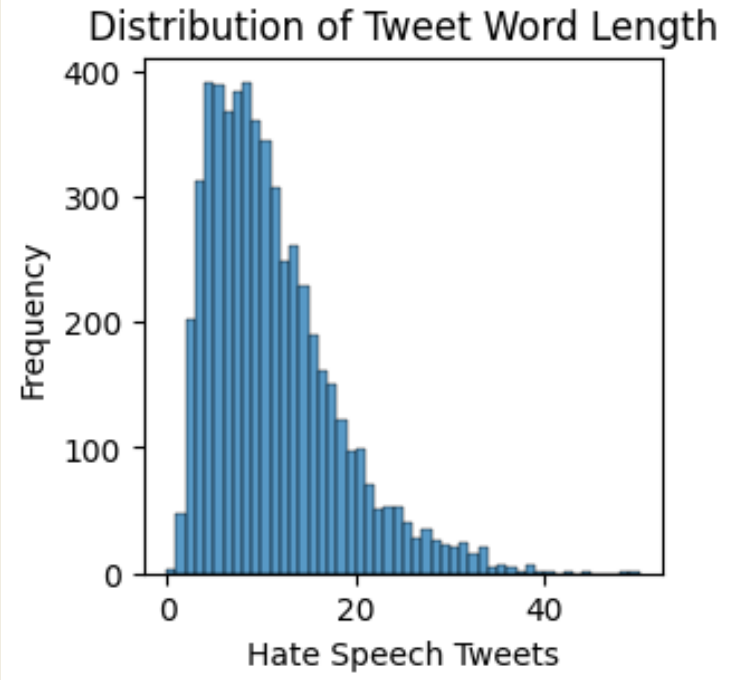
Hasil dan Pembahasan

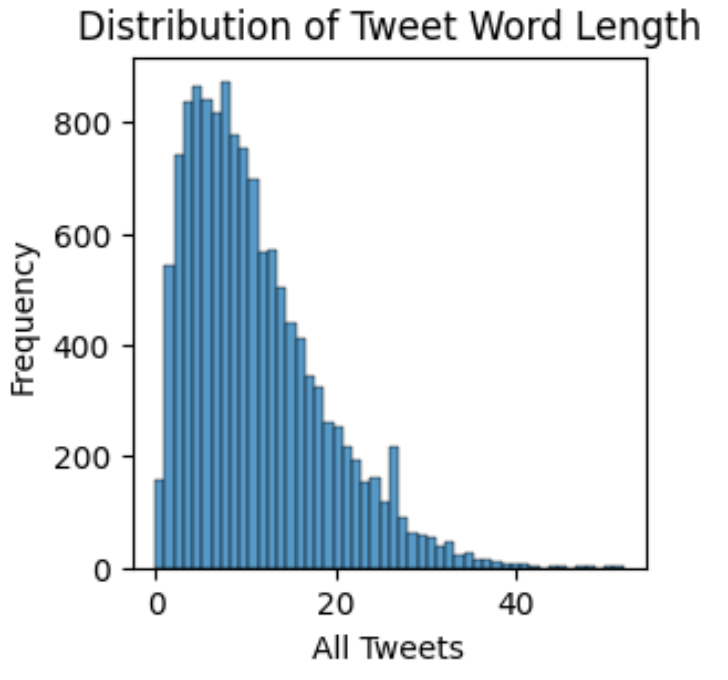


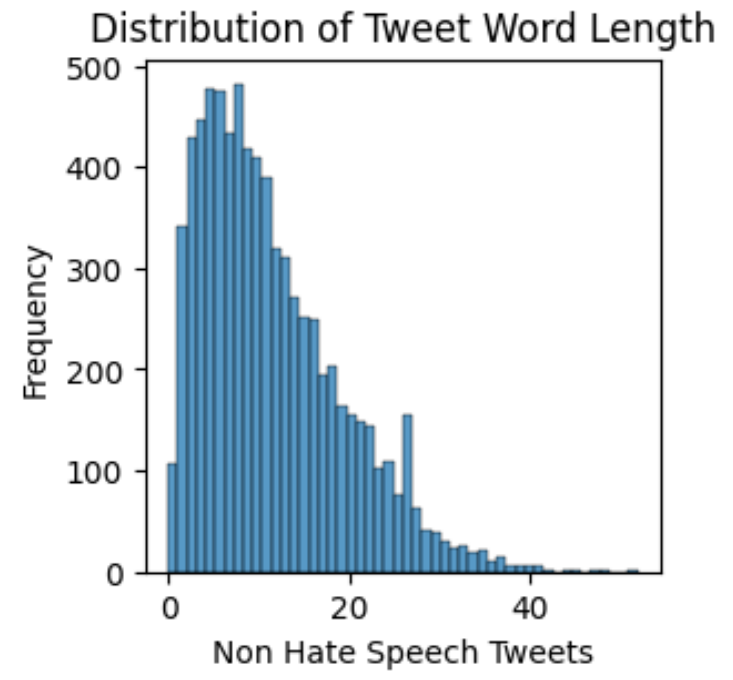

```
kategori_hs
non-hate speech    7576
hate speech        5549
Name: count, dtype: int64
```

- Terbanyak : Kategori non-hate speech (tweet tidak termasuk hate speech sebesar 57,7%)

Hasil dan Pembahasan

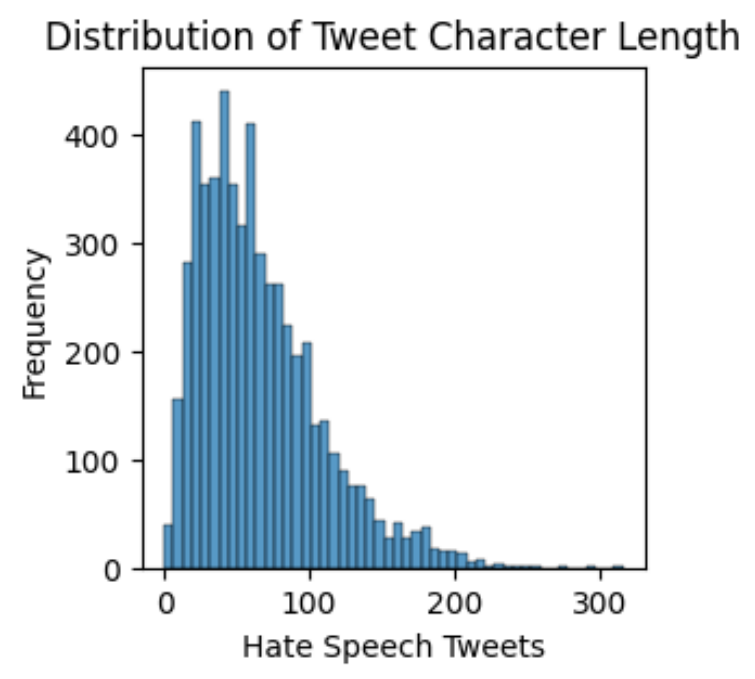


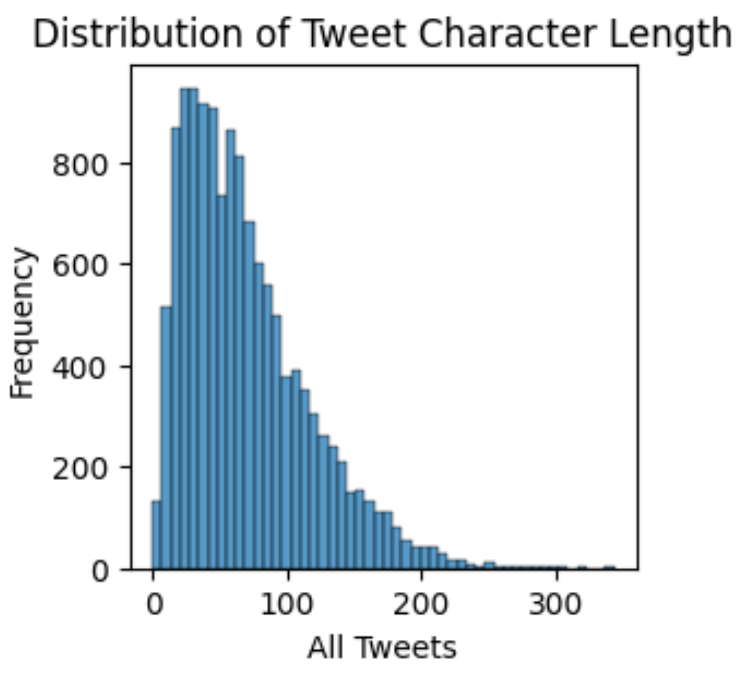


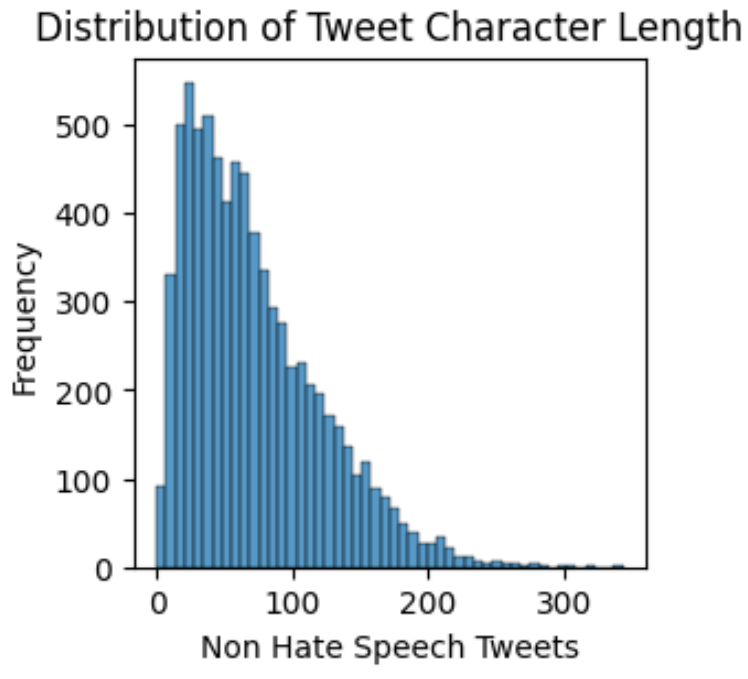


- Panjang kata tweet dari ketiganya memiliki distribusi hampir sama, serta mempunyai kisaran range panjang kata yang sama hingga 50-an kata.

Hasil dan Pembahasan

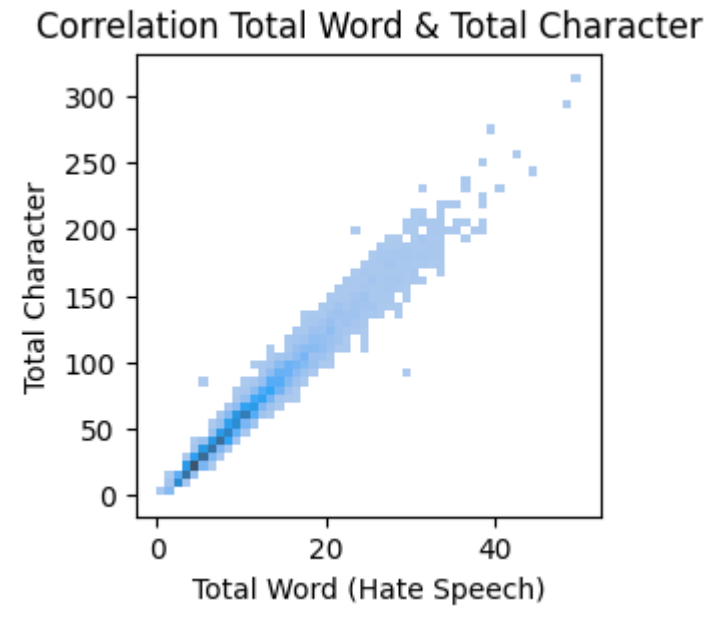


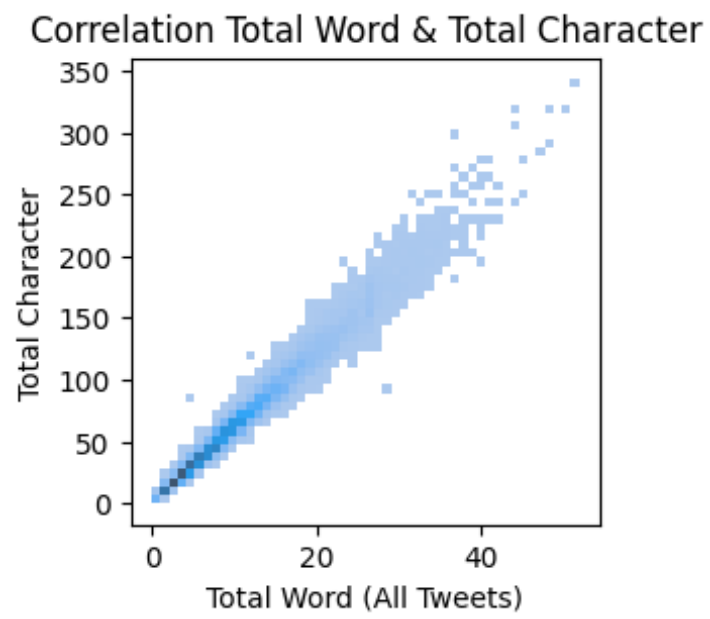


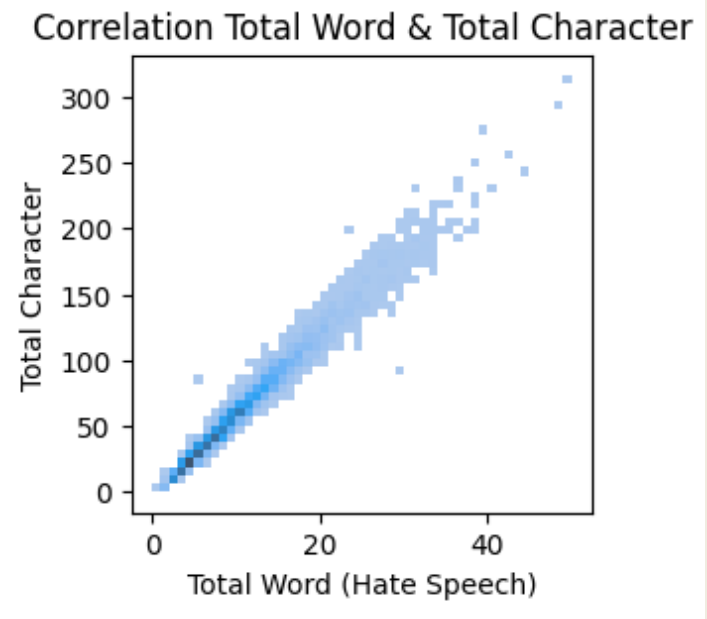


- Distribusi panjang karakter tweet secara keseluruhan maupun yang tweet tidak mengandung hate speech hampir sama. Namun dari ketiganya memiliki kisaran range karakter yang sama hingga 350 karakter.

Hasil dan Pembahasan

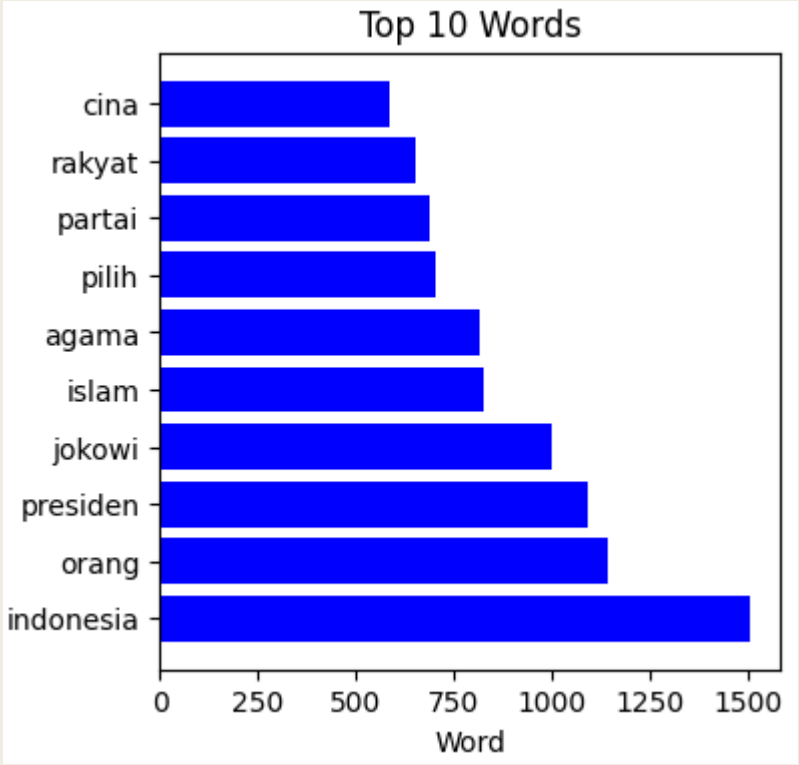






- Korelasi dari ketiganya, memiliki pola yang sama, yaitu semakin panjang kata pada suatu tweet, maka semakin panjang pula karakter tweet tersebut.

Hasil dan Pembahasan

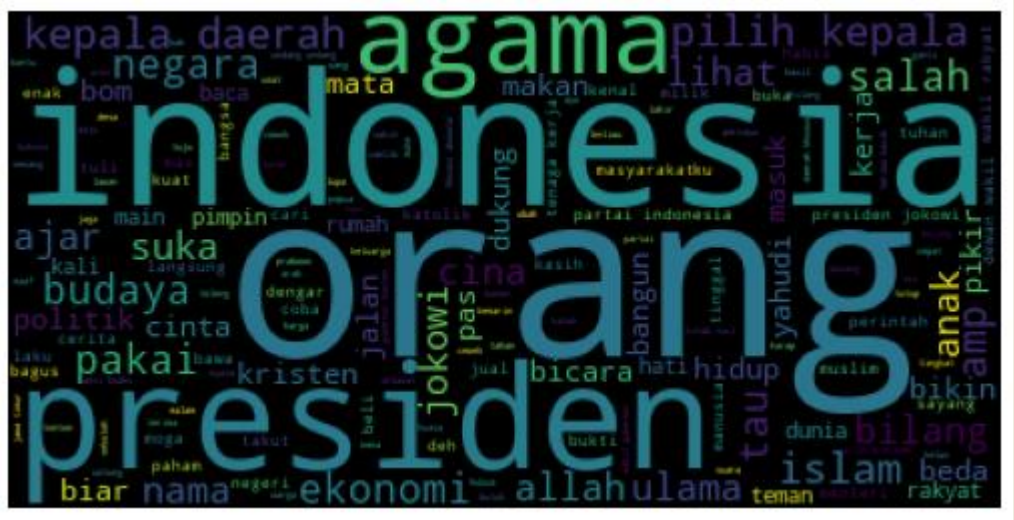


- Kata yang paling banyak muncul pada data tweet secara keseluruhan adalah Indonesia.

Hasil dan Pembahasan

- Kata yang paling banyak muncul pada data tweet yang mengandung hate speech adalah kata ‘jokowi’ dan ‘orang’.

Hasil dan Pembahasan



- Kata yang paling banyak muncul pada data tweet yang tidak mengandung hate speech adalah kata ‘orang’, ‘indonesia’, dan ‘presiden’.

Kesimpulan

1. Data tweet yang tidak mengandung *hate speech* lebih banyak dari tweet yang mengandung *hate speech.*
2. Distribusi panjang karakter tweet secara data keseluruhan maupun tweet yang tidak mengandung hate speech hampir sama. Sedangkan, panjang kata pada tweet *hate speech*, *non-hate speech*, maupun *all tweets* memiliki bentuk distribusi yang mirip.
3. Semakin panjang kata pada suatu tweet, maka semakin panjang pula karakter tweet tersebut.
4. Kata ‘Indonesia’ merupakan kata yang paling banyak dibicarakan pada data twitter secara keseluruhan. Untuk tweet yang mengandung *hate speech*, paling banyak muncul adalah kata ‘jokowi’ dan ‘orang’. Sedangkan kata ‘orang’, ‘indonesia’, dan ‘presiden’ ialah kata yang sering muncul pada tweet yang mengandung *hate speech.*

Saran/Rekomendasi

Pihak terkait dengan masalah ujaran kebencian di media sosial, seperti pemerintah dan lain sebagainya bisa menindaklanjuti mengenai orang yang paling banyak dibicarakan pada tweet yang mengandung *hate speech* agar dapat meningkatkan kesadaran dan tanggung jawab pengguna media sosial dalam berkomunikasi secara etis dan santun.

Thank you