

Analisis Kebiasaan Pengguna Tweeter Berdasarkan Sentiment Kalimat Menggunakan Descriptive Analytics

Pendahuluan

Tweeter adalah salah satu media sosial yang paling populer di dunia, termasuk di Indonesia. masyarakat indonesia tidak asing lagi dengan media sosial tweeter. oleh karena itu tidak jarang juga masyarakat indonesia yang menggunakan media sosial tersebut.

Salah satu kelebihan dari tweeter adalah kita bisa mengekspresikan perasaan kita melalui kata-kata yang bisa kita bagikan ke orang lain dengan mudah. namun, kebebasan berekspresi menimbulkan masalah lain. kebebasan tersebut menimbulkan banyak pengguna tweeter mengekspresikan kata-katanya secara tidak baik dan cenderung abusive.

Oleh karena itu, penelitian ini bertujuan untuk menganalisis behavior pengguna twitter dalam menggunakan tweeter untuk mengekspresikan kata-kata dan perasaan mereka berdasarkan panjang tweet, panjang kata, pilihan kata, tingkat abusive tweet tersebut, dan jenis abusive yang digunakan. Harapannya hasil dari analisis ini, dapat memberikan wawasan terhadap behavior pengguna twitter untuk berbagai pihak kedepannya.

Metode Penelitian

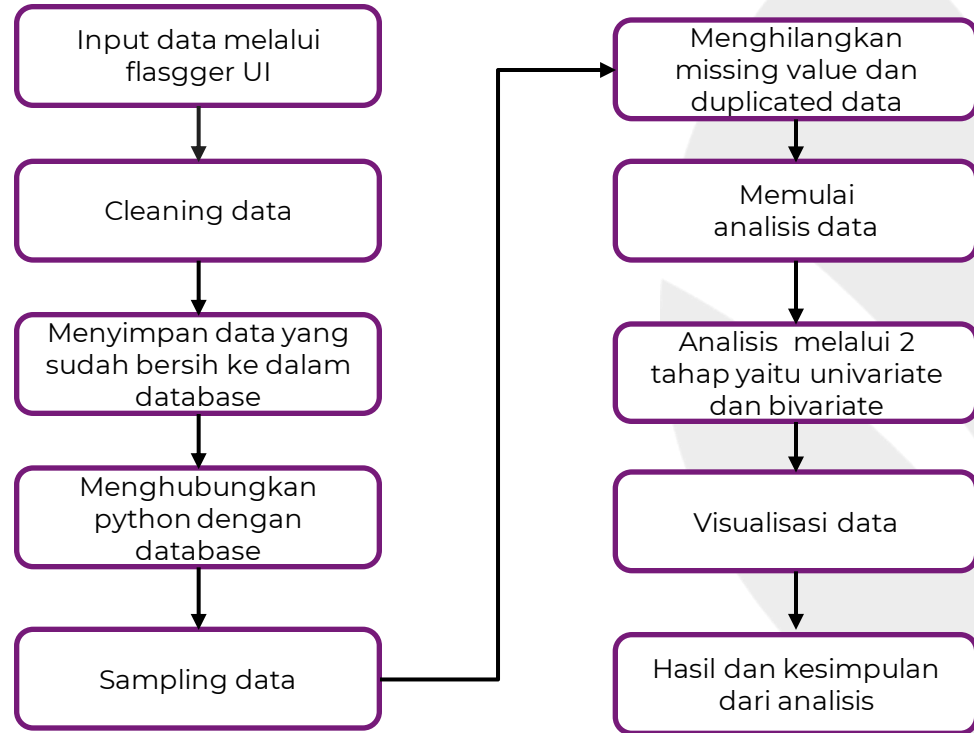
Data pada penelitian ini bersumber dari dataset Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter (Ibrohim M, 2019). Data yang dianalisis adalah data yang memuat tweet pengguna yang telah dianalisis dan diberi label sentiment sesuai dari kalimat pada tweet tersebut.

Metode analisis yang dipakai dalam penelitian ini adalah descriptive analysis dengan tujuan untuk mendeskripsikan pola data. Jenis analisis tersebut sesuai dengan fokus penelitian ini untuk mencari tahu behavior pengguna tweeter dalam menggunakan kebebasannya untuk berekspresi pada tweet dan komen.

Proses analisis dilakukan dengan berdasarkan kolom yang diproses yakni 1 variabel (univariate analysis) dan 2 variabel (bivariate analysis). Dalam setiap prosesnya menerapkan metode descriptive statistic dan visualisasi. Descriptive statistic digunakan untuk mencari tahu persebaran data secara angka sedangkan visualisasi untuk mencari tahu persebaran data secara visual.

Work Flow

1. Analisis sentiment
2. Descriptive statistic
3. Korelasi
4. Visualisasi



Rumusan Masalah

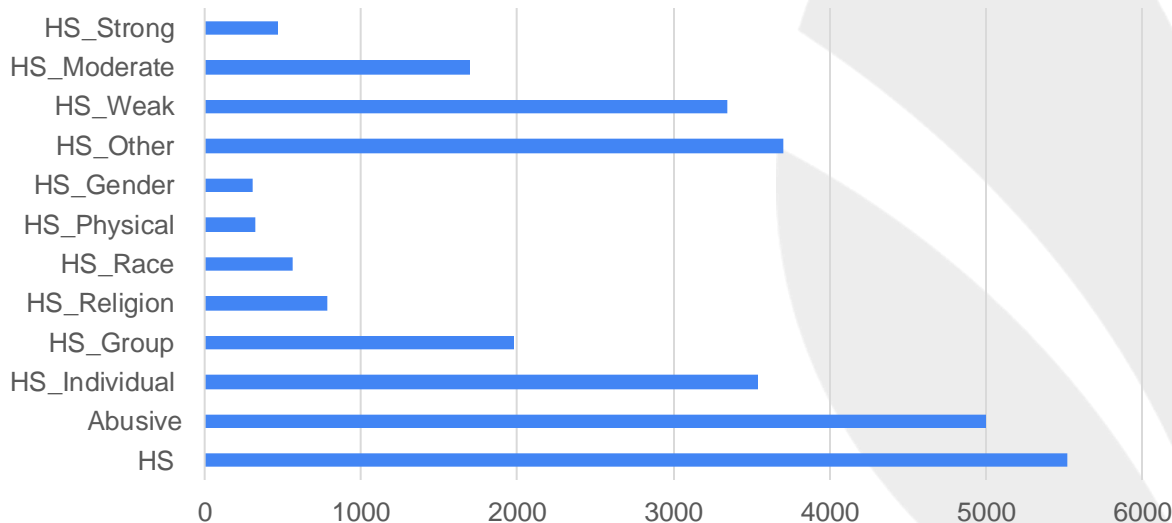
1. Berapa jumlah Hate speech?
2. Apa kategori hate speech yang paling sering dilakukan?
3. Apa tingkatan hate speech yang paling sering dilakukan?
4. Berapa rata-rata karakter dan kata yang biasa digunakan pengguna tweeter untuk mengekspresikan setiap kategori hate speech dan abusive?
5. Apa kata hate speech yang paling sering dikatakan pada tingkatan setiap kategori hate speech?
6. Apa kata hate speech yang paling sering dikatakan pada tingkatan setiap tingkatan?
7. Apa hubungan panjang karakter dan kata?

Hasil Univariate Analisis

Berdasarkan analisis yang sudah kita lakukan dapat hasilnya dapat dijabarkan sebagai berikut::

Jumlah tweet yang telah diberi label sentimen:

- HS 5518
- Abusive 5005
- HS_Individual 3540
- HS_Group 1978
- HS_Religion 789
- HS_Race 563
- HS_Physical 322
- HS_Gender 304
- HS_Other 3706
- HS_Weak 3348
- HS_Moderate 1698
- HS_Strong 472

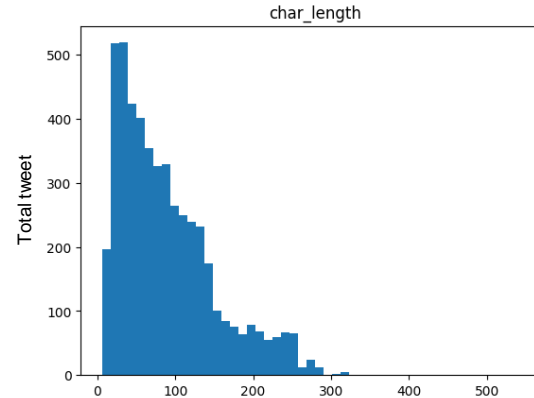
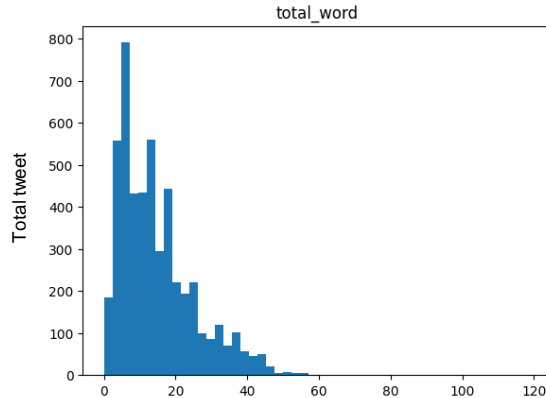


Hasil Univariate Analisis

Pada kategori Abusive ditemukan bahwa:

Rata-rata kata	: 15
Median Kata	: 12
Modus Kata	: 4
Rata-rata Karakter	: 89
Median Karakter	: 74
Modus Karakter	: 68

Berikut adalah persebaran data dari penggunaan kata dan karakter di Twitter pada kategori Abusive :

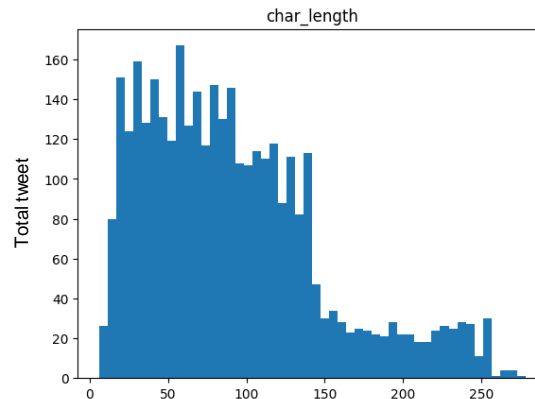
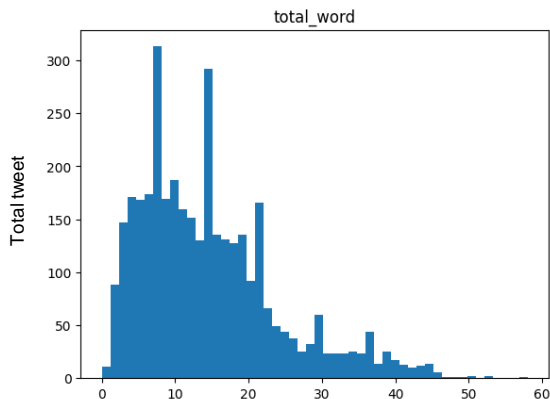


Hasil Univariate Analysis

Pada kategori HS_Individual ditemukan bahwa:

Rata-rata kata	: 14.5
Median Kata	: 13
Modus Kata	: 10
Rata-rata Karakter	: 92
Median Karakter	: 82.5
Modus Karakter	: 49

Berikut adalah persebaran data dari penggunaan kata dan karakter di Twitter pada kategori HS_ Individual :

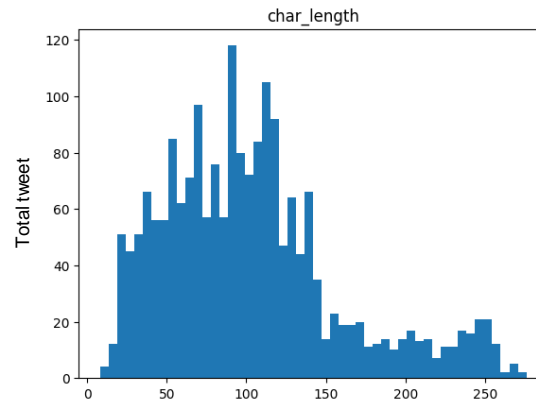
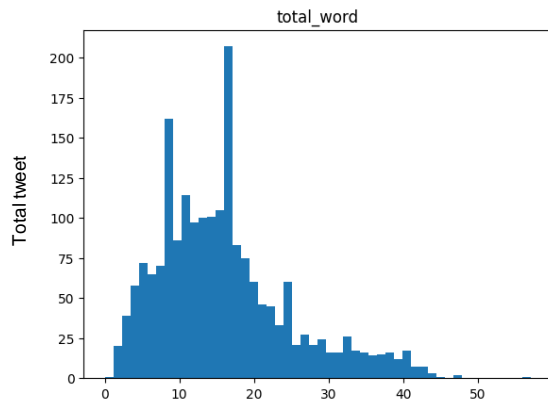


Hasil Univariate Analysis

Pada kategori HS_Group ditemukan bahwa:

Rata-rata kata	: 16
Median Kata	: 15
Modus Kata	: 11
Rata-rata Karakter	: 103
Median Karakter	: 95
Modus Karakter	: 116

Berikut adalah persebaran data dari penggunaan kata dan karakter di Twitter pada kategori HS_Group :

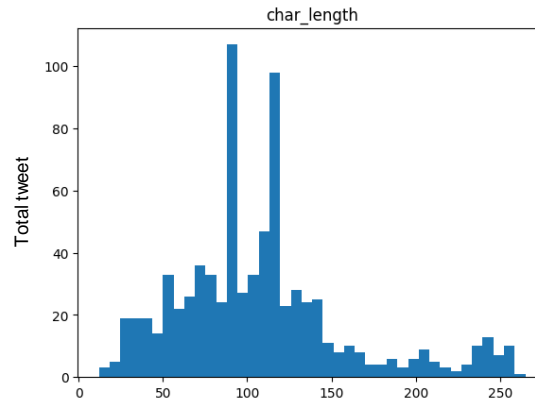
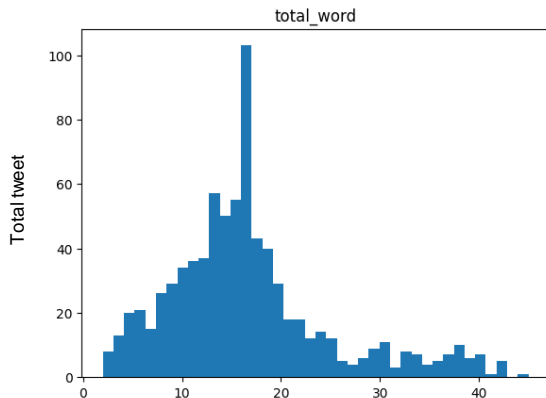


Hasil Univariate Analysis

Pada kategori HS_Religion ditemukan bahwa:

Rata-rata kata	: 17
Median Kata	: 15
Modus Kata	: 17
Rata-rata Karakter	: 108
Median Karakter	: 102
Modus Karakter	: 116

Berikut adalah persebaran data dari penggunaan kata dan karakter di Twitter pada kategori HS_Religion:

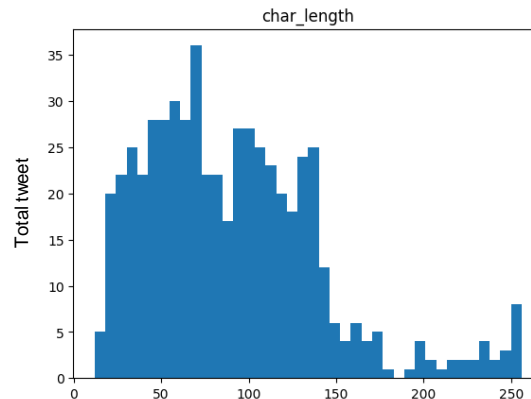
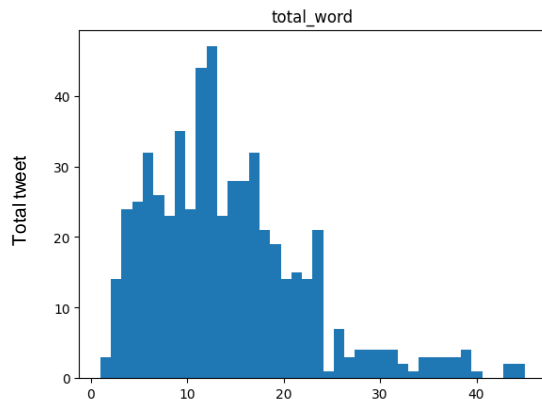


Hasil Univariate Analysis

Pada kategori HS_Race ditemukan bahwa:

Rata-rata kata	: 14
Median Kata	: 13
Modus Kata	: 11
Rata-rata Karakter	: 91
Median Karakter	: 83
Modus Karakter	: 127

Berikut adalah persebaran data dari penggunaan kata dan karakter di Twitter pada kategori HS_Race :

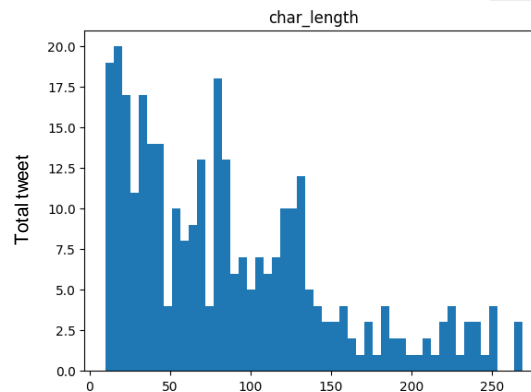
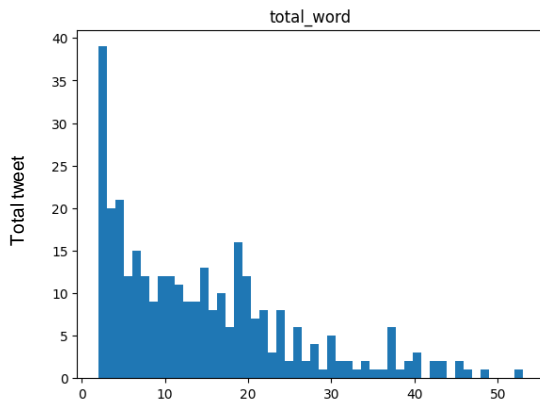


Hasil Univariate Analisis

Pada kategori HS_Physical ditemukan bahwa:

Rata-rata kata	: 14.5
Median Kata	: 12
Modus Kata	: 3
Rata-rata Karakter	: 87
Median Karakter	: 79
Modus Karakter	: 25, 82

Berikut adalah persebaran data dari penggunaan kata dan karakter di Twitter pada kategori HS_Physical:

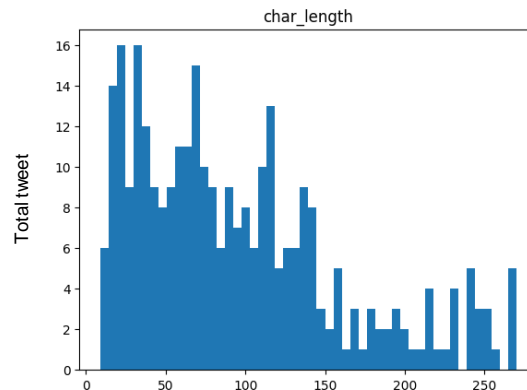
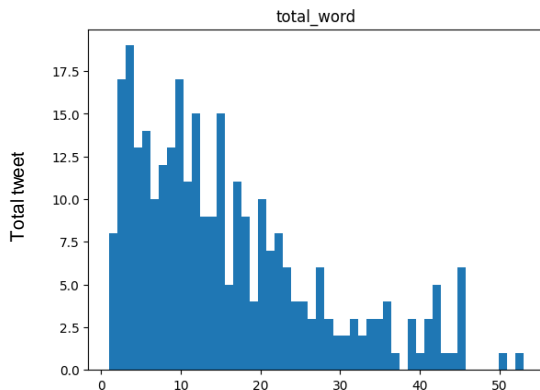


Hasil Univariate Analisis

Pada kategori HS_Gender ditemukan bahwa:

Rata-rata kata	: 16
Median Kata	: 13
Modus Kata	: 4
Rata-rata Karakter	: 95.5
Median Karakter	: 80.5
Modus Karakter	: 19, 23, 26, 31, 67

Berikut adalah persebaran data dari penggunaan kata dan karakter di Twitter pada kategori HS_Gender:

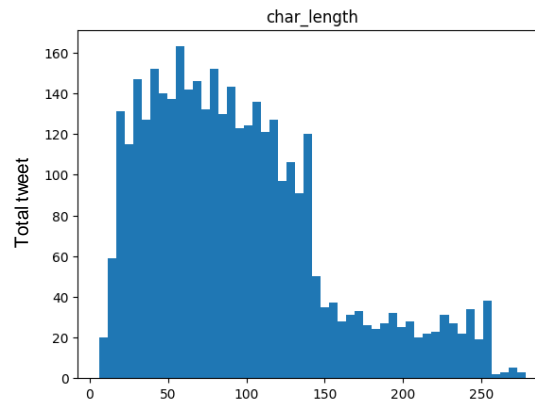
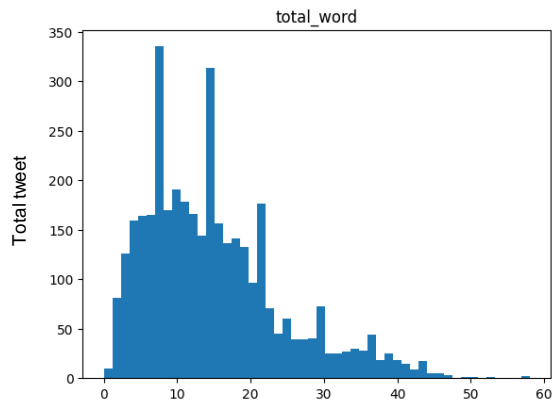


Hasil Univariate Analisis

Pada kategori HS_Other ditemukan bahwa:

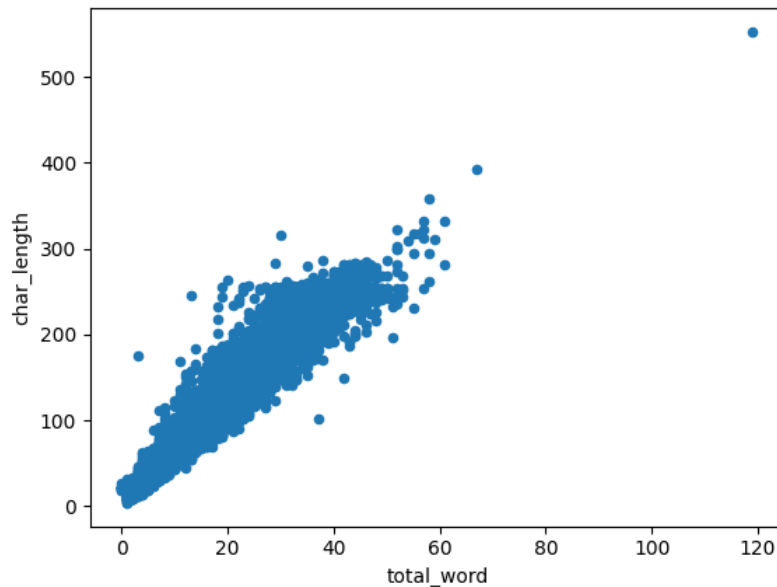
Rata-rata kata	: 15
Median Kata	: 13
Modus Kata	: 10
Rata-rata Karakter	: 95
Median Karakter	: 86
Modus Karakter	: 68

Berikut adalah persebaran data dari penggunaan kata dan karakter di Twitter pada kategori HS_Other:



Hasil Bivariate Analisis

Hasil dari analisis korelasi Panjang kata dan karakter pada tweet, ditemukan bahwa korelasi antara kedua variable tersebut mendapatkan nilai sebesar 0.96. dari temuan tersebut dapat disimpulkan bahwa Panjang kata dan Panjang karakter pada tweet mempunyai korelasi positif dan memiliki korelasi sangat kuat.

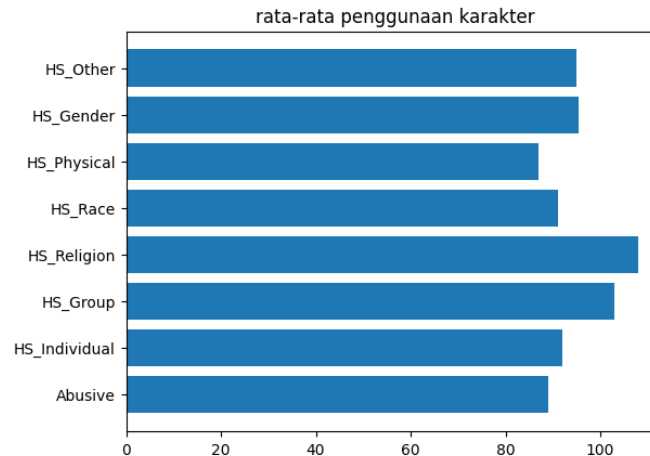
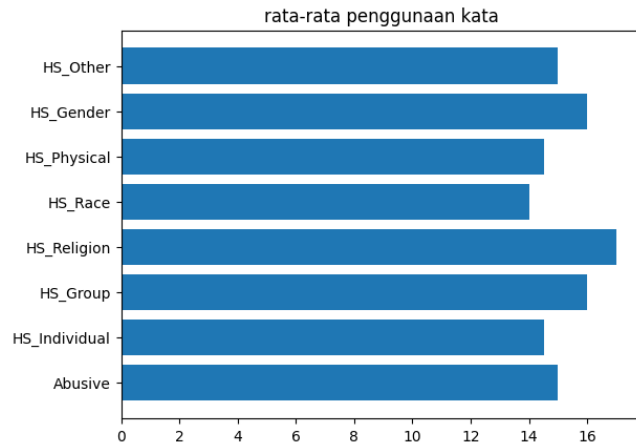


Hasil dan Kesimpulan

1. Minimal karakter pada tweet adalah 4 karakter dan maksimal karakternya adalah 553 karakter.
2. Minimal kata pada tweet adalah 0 kata dan maksimal kata-nya adalah 119 karakter.
3. Tweet dengan sentimen hate speech terdapat 5518 tweet dari 13044 tweet
4. Tweet dengan abusive hate speech terdapat 5005 tweet dari 13044 tweet
5. Tweet hate speech terbanyak ada pada kategori Hate speech other dengan jumlah 3706.
6. Tweet hate speech paling sedikit dan hate speech kepada gender dengan jumlah 304.
7. Kata yang paling sering digunakan dalam mengekspresikan hate speech dan kata abusive pada masing-masing kategori adalah:
 - Abusive: Cebong, Dasar, Babi, Bajingan, Tolol, Kampret
 - Individual: Jokowi, Cebong, Ahok, Presiden, Rakyat, Prabowo
 - Group: Indonesia, Cebong, Cina, Islam, Komunis, PKI, Kafir, Rezim
 - Religion: Islam, Kafir, Agama, Muslim, Anti Islam, Ahok, Indonesia, Penista Agama
 - Race: Cina, PKI, Indonesia, Komunis, Usir, Ganyang Pki, kafir
 - Physical: Idiot, Bolot, Picek, Budek, Cacat, Sarap, Gila
 - Gender: Banci, Bencong, Dasar, Kontrol, Cabul, Jablay, Maho
 - Other: Cebong, Jokowi, Rakyat, Presiden, Prabowo, Ahok
8. kata yang paling sering dipakai dalam mengekspresikan hate speech pada masing-masing tingkatan adalah:
 - Weak: Jokowi, Cebong, Presiden, Prabowo, Rakyat, Ahok
 - Moderate: Indonesia, Cebong, Islam, Cina, Komunis, Pki, Agama
 - Strong: Lengserkan Jokowi, Indonesia, Jokowi, Ganyang PKI, PKI, Bubarkan DPR, Cina,

Hasil dan Kesimpulan

9. Rata-rata penggunaan kata terbanyak ada pada kategori hate speech religion yang berarti pengguna twitter mengekspresikan hate speechnya yang berkaitan dengan agama dengan lebih banyak kata dari pada dalam mengekspresikan hate speech lain.
10. Sesuai dengan hasil analisis korelasi sebelumnya, dimana penggunaan karakter berbanding lurus dengan penggunaan kata. Jadi rata-rata penggunaan karakter terbanyak juga pada kategori hate speech religion.



TERIMAKASIH