

PASCAL BRILLIANDY K



ANALISIS ABUSIVE & HATE SPEECH X (TWITTER): PERSPEKTIF BAHASA INDONESIA

TABLE OF CONTENT

INTRODUCTION

- Latar Belakang
- Rumusan Masalah
- Tujuan Analisa

RESEARCH METHOD

- Flow Analisis Data
- Deskripsi Data
- Metode Analisis Data

CONCLUSION

- Hasil dan Kesimpulan

INTRODUCTION

LATAR BELAKANG

Di era digital saat ini, X (Twitter) merupakan salah satu platform untuk membagikan pikiran dan opini secara bebas. Namun, dampak negatifnya termasuk ujaran kebencian dan penggunaan kata kasar, menimbulkan kekhawatiran terhadap etika online dan kesejahteraan psikologis individu.

Di Indonesia, keberagaman budaya, agama, dan suku seringkali memicu pertentangan di Twitter. Oleh karena itu, analisis terhadap bahasa kasar dan ujaran kebencian sangat penting untuk mencegah konflik sosial dan menciptakan lingkungan online yang aman dan beradab. Dengan memahami pola perilaku pengguna, strategi efektif dapat dikembangkan untuk mengatasi tantangan ini.

INTRODUCTION

RUMUSAN MASALAH

- Berapa jumlah perbandingan kata yang tidak mengandung HS & Abusive dengan kata yang mengandung HS dan Abusive?
 - Berapa jumlah perbandingan jumlah kata yang mengandung kata yang diklasifikasikan sebagai HS dengan Abusive?
 - Bagaimana perbandingan antara jumlah kasus HS yang ditujukan kepada individu dengan yang ditargetkan pada suatu grup?
 - Apa saja kata-kata kasar yang sering muncul di tweet yang terkait dengan HS di platform Twitter Bahasa Indonesia?"
 - Berapa jumlah panjang karakter dan panjang kata yang tidak mengandung kata alay?
-

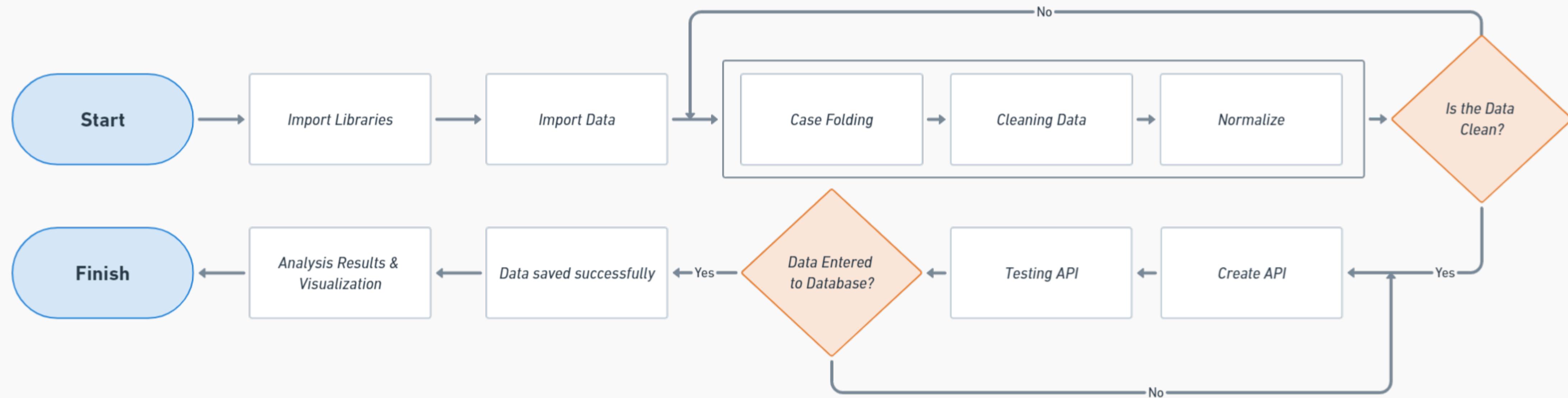
INTRODUCTION

TUJUAN ANALISA

- Mengidentifikasi perbandingan antara tweet yang mengandung HS & abusive dengan yang tidak mengandungnya secara signifikan.
 - Membandingkan proporsi tweet yang mengandung HS dan Abusive dari masing-masing jenis di Twitter
 - Menentukan perbandingan jumlah kasus HS yang ditujukan kepada individu dengan suatu grup untuk mengidentifikasi preferensi atau tren dalam perilaku ujaran kebencian.
 - Mengidentifikasi kata-kata yang paling sering digunakan dalam konteks HS dalam bahasa Indonesia yang digunakan untuk menyebarkan kebencian di platform Twitter.
 - Menganalisis distribusi panjang karakter dan panjang kata dalam tweet Bahasa Indonesia untuk tweet yang tidak mengandung kata alay.
-

RESEARCH METHOD

FLOW ANALISA DATA



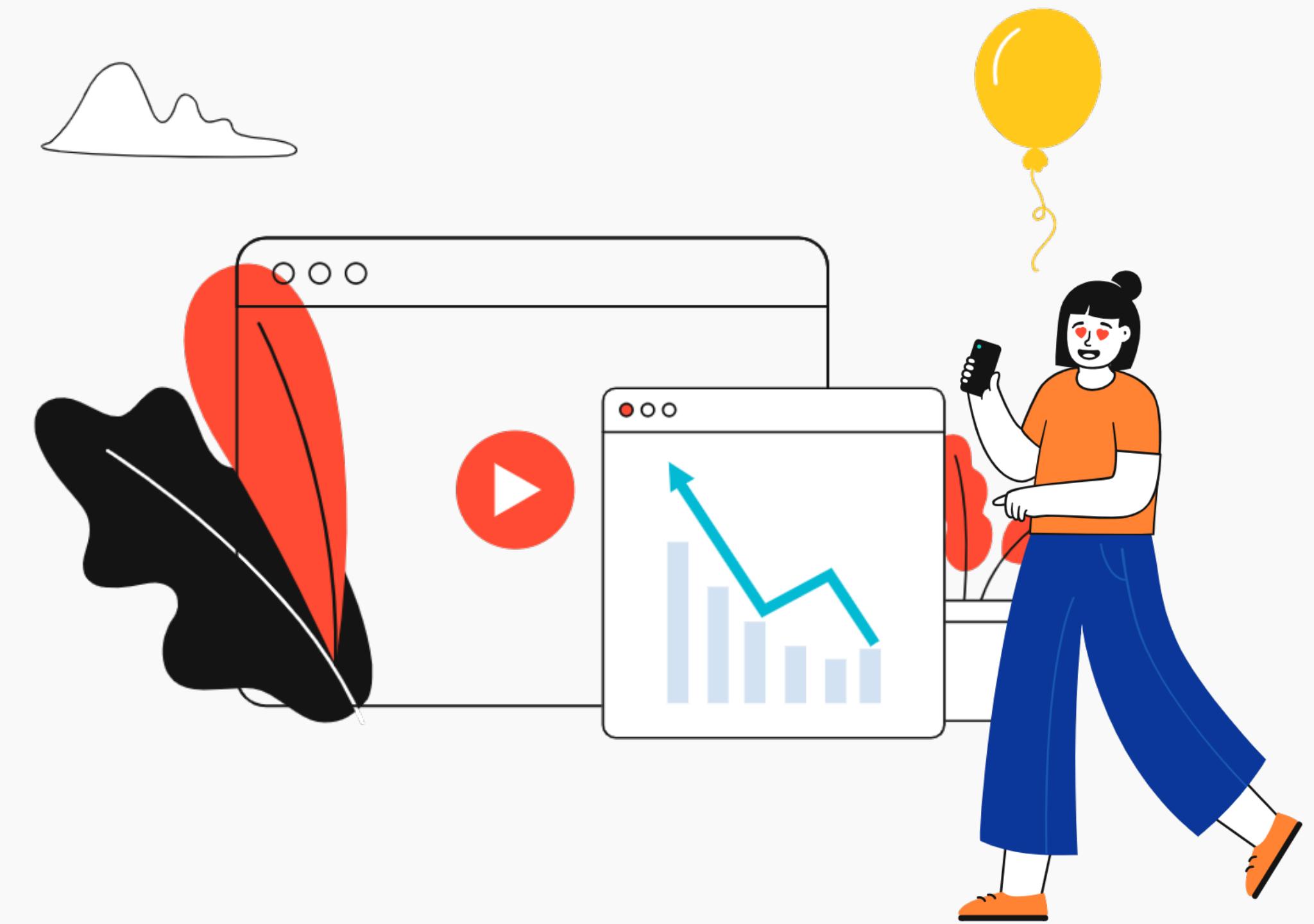
DESKRIPSI DATA

Dataset yang digunakan dalam Challenge ini diambil dari Kaggle dalam bentuk csv dengan menggunakan 3 File csv yaitu data.csv, new_kamusay.csv, abusive.csv. Berikut merupakan kolom yang ada di dataset data.csv, antara lain :

- Tweet : Collection of texts from Twitter
 - HS : hate speech label
 - Abusive : abusive language label
 - HS_Individual : hate speech targeted to an individual
 - HS_Group : hate speech targeted to a group
 - HS_Religion : hate speech related to religion/ creed
 - HS_Race : hate speech related to race/ethnicity
 - HS_Physical : hate speech related to physical/ disability
 - HS_Gender : hate speech related to gender/ sexual orientation
 - HS_Other : hate related to other invective/slander
 - HS_Weak : weak hate speech
 - HS_Moderate : moderate hate speech
 - HS_Strong : strong hate speech
-

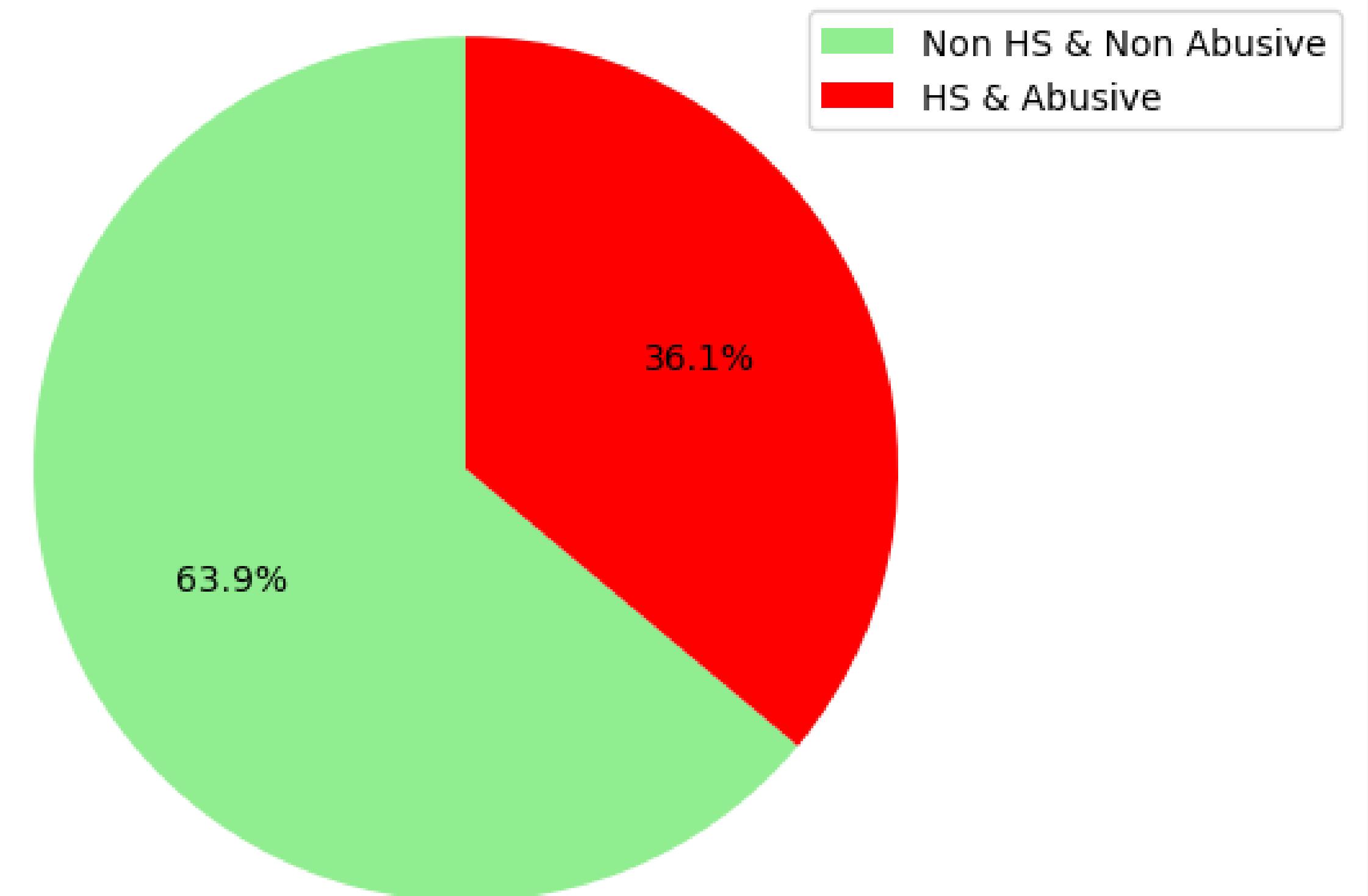
METODE ANALISIS DATA

Metode yang saya gunakan untuk studi kasus ini adalah univariate analytics, yang dimana saya memanfaatkan metode statistika dari measures of central tendency untuk menampilkan gambaran tentang distribusi dari data teks yang diambil dari Twitter



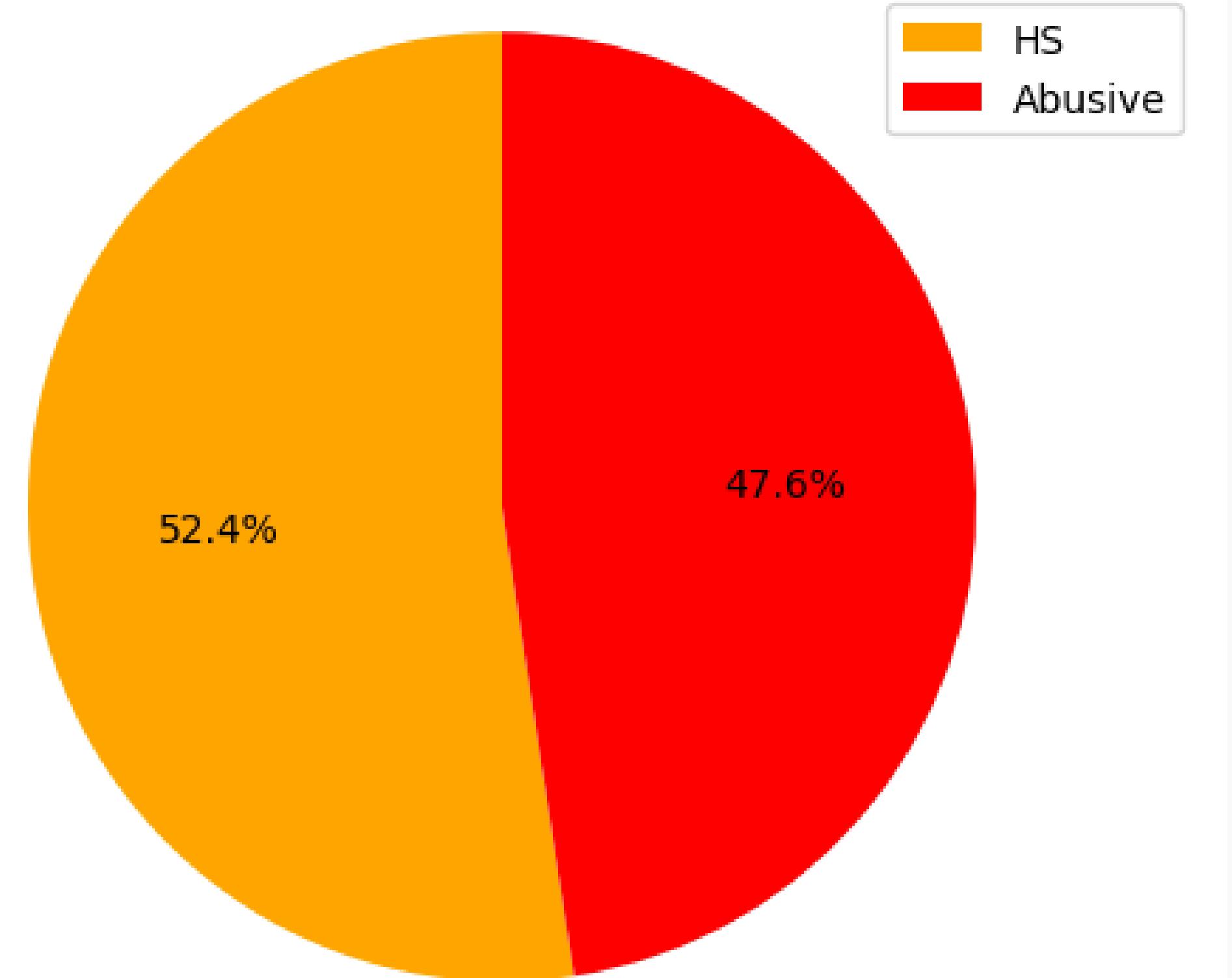
PERBANDINGAN HATE SPEECH & ABUSIVE

Result visualisasi tweet yang mengandung HS & abusive dengan yang tidak mengandung HS & Abusive



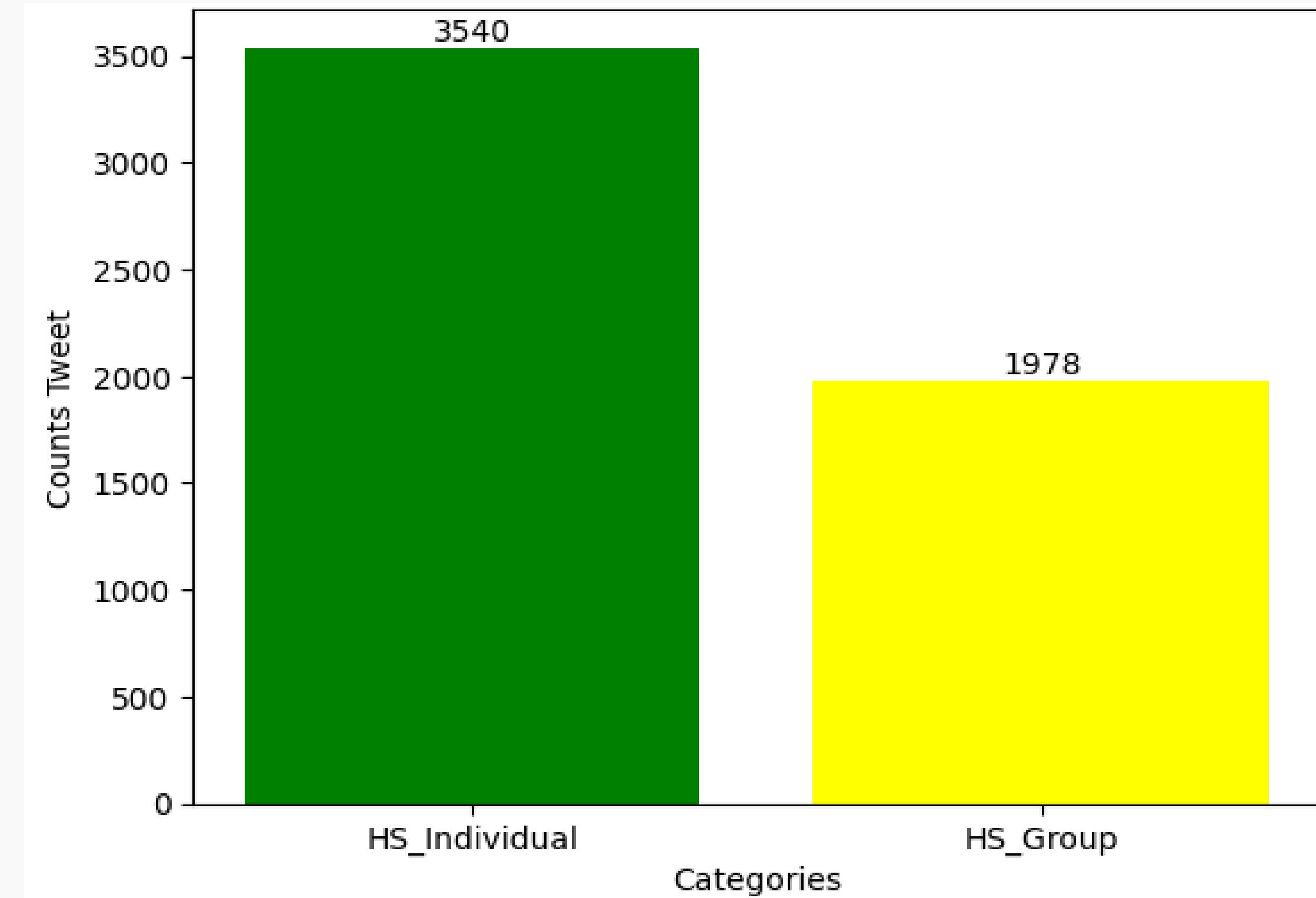
PERBANDINGAN HATE SPEECH & ABUSIVE

Result visualisasi proporsi tweet yang mengandung HS dan Abusive dari masing-masing jenis di Twitter.



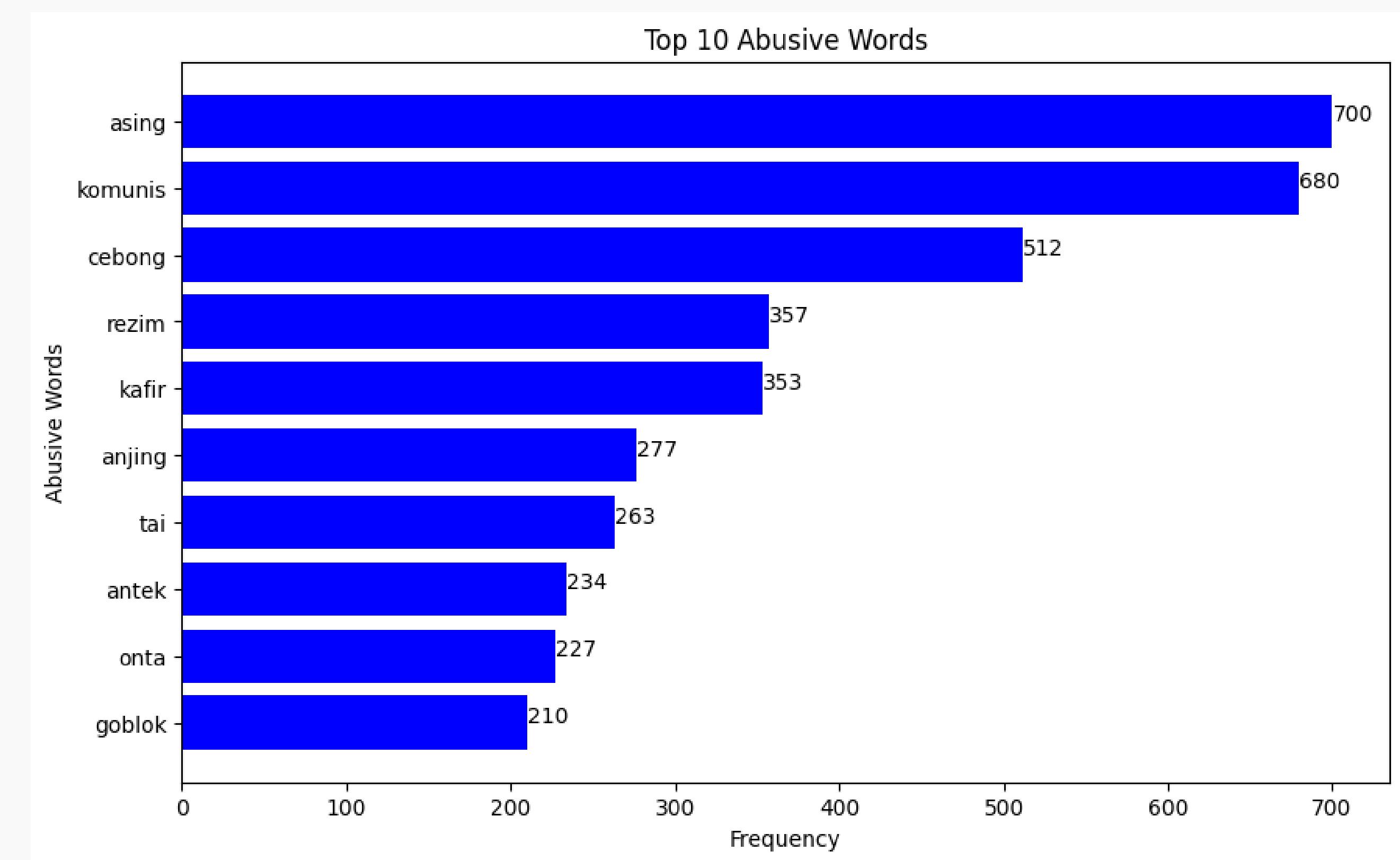
PERBANDINGAN INDIVIDUAL VS GROUP

Result visualisasi perbandingan jumlah kasus HS yang ditujukan kepada individu maupun dalam group.

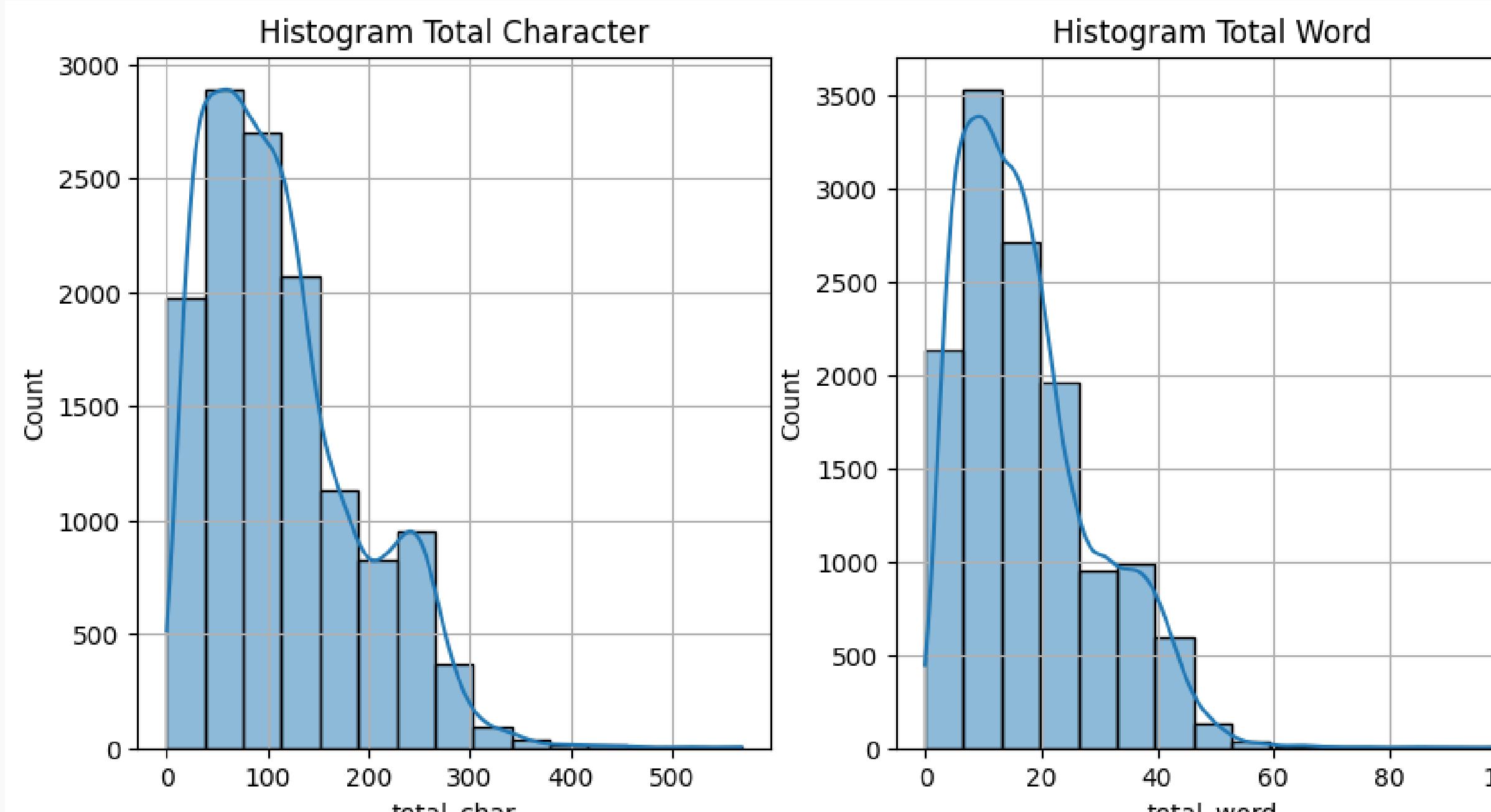


EXPLORATORY DATA ANALYSIS

10 KATA ABUSIVE YANG SERING DIGUNAKAN



EXPLORATORY DATA ANALYSIS



DISTRIBUSI PANJANG CHAR & KATA

Result skewness bernilai > 0 yang berarti menunjukkan bahwa distribusi memiliki skewness positif.

```
print('Skewness total word =',df['Total_word'].skew())
print('Skewness total char =',df['total_char'].skew())
✓ 0.0s
Skewness total word = 0.9469706493179129
Skewness total char = 0.894125421238742
```

CONCLUSION

HASIL DAN KESIMPULAN

- Dari hasil analisa yang didapat bahwa 36.1% cuitan tweet mengandung kata HS & abusive, 63.9% sisa nya yang tidak mengandung kata-kata tersebut.
 - Tweet yang mengandung HS terdapat persentase 52.4%, sedangkan 47.6% sisa nya adalah dari Tweet yang mengandung Abusive.
 - Jumlah kasus ujaran kebencian terhadap individu dan kelompok menunjukkan bahwa masih banyak orang yang melakukan ujaran kebencian terhadap individu daripada kelompok.
 - Kata "Asing" merupakan kata abusive yang sering digunakan dalam Twitter, dengan jumlah frekuensi sebanyak 200.
 - Dari panjang kata dan panjang karakter yang diperoleh, dapat disimpulkan bahwa distribusi memiliki skewness dengan nilai > 0 , yang berarti skewness positif.
-

PASCAL BRILLIANDY K



THANK YOU
EVERYONE