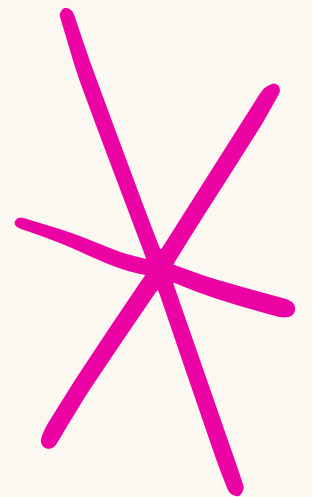
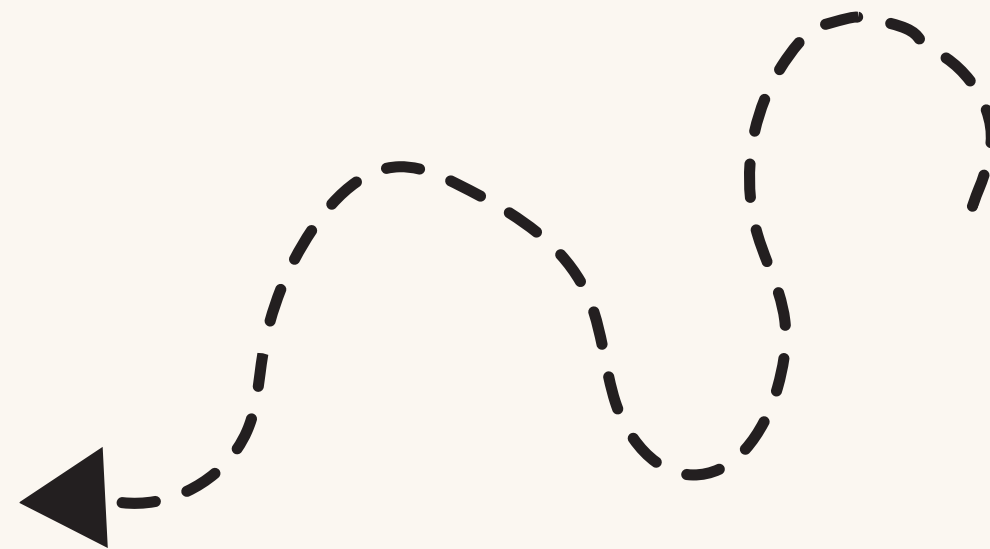




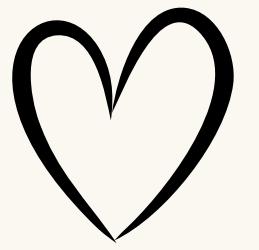
Report

ANALYSIS



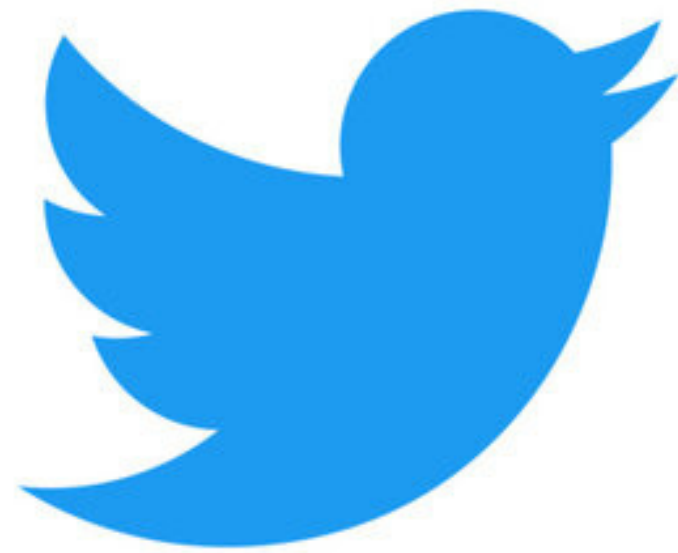


# Pendahuluan



Indonesia adalah salah satu negara dengan jumlah pengguna sosial media yang sangat tinggi yaitu sejumlah 202,6 juta pengguna per tahun 2021 (HootSuite dan We Are Social).

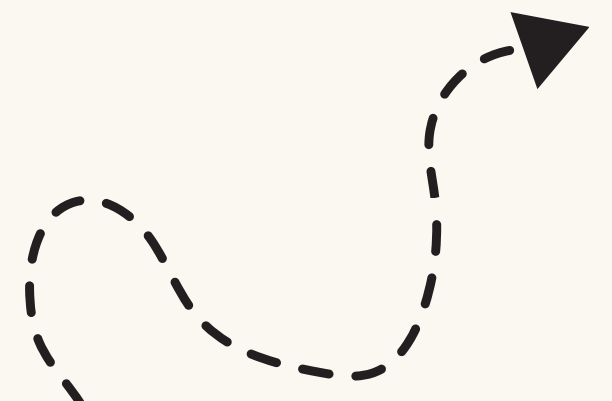
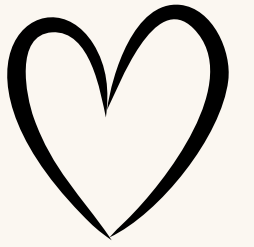
Berdasarkan angka yang fantastis tersebut, sosial media tentunya berpengaruh besar terhadap kehidupan bermasyarakat seperti opini publik, tren, hingga pandangan politik sangat dipengaruhi oleh sosial media.



Twitter

Twitter adalah salah satu platform sosial media yang paling banyak digunakan oleh warganet Indonesia. Per tahun 2021 jumlah pengguna twitter Indonesia mencapai 15,7 juta pengguna (Statista)

Kebebasan untuk memposting dan opini yang ada di dalamnya membuat adalah salah satu alasan mengapa Twitter banyak digemari oleh usernya.



# TWITTER

As an interactive social media



- **BRAND IDENTITY**

Dengan adanya kebebasan berekspresi membuat Twitter rentan akan penggunaan Hate Speech

- **AKIBAT**

- Environment sosial media yang kurang nyaman ketika membahas topik tertentu
- Sosial media rentan menjadi alat tunggangan politik dengan buzzer

- **BENTUK HATE SPEECH**

Bentuk ujaran kebencian yang sering user jumpai beragam seperti SARA, omongan kotor, dan ujaran kebencian lainnya.

- **TUJUAN PENELITIAN**

- Mengetahui topik yang sering dibahas di Twitter dengan menggunakan Hate Speech.

# METODE PENELITIAN



## METODE PENELITIAN

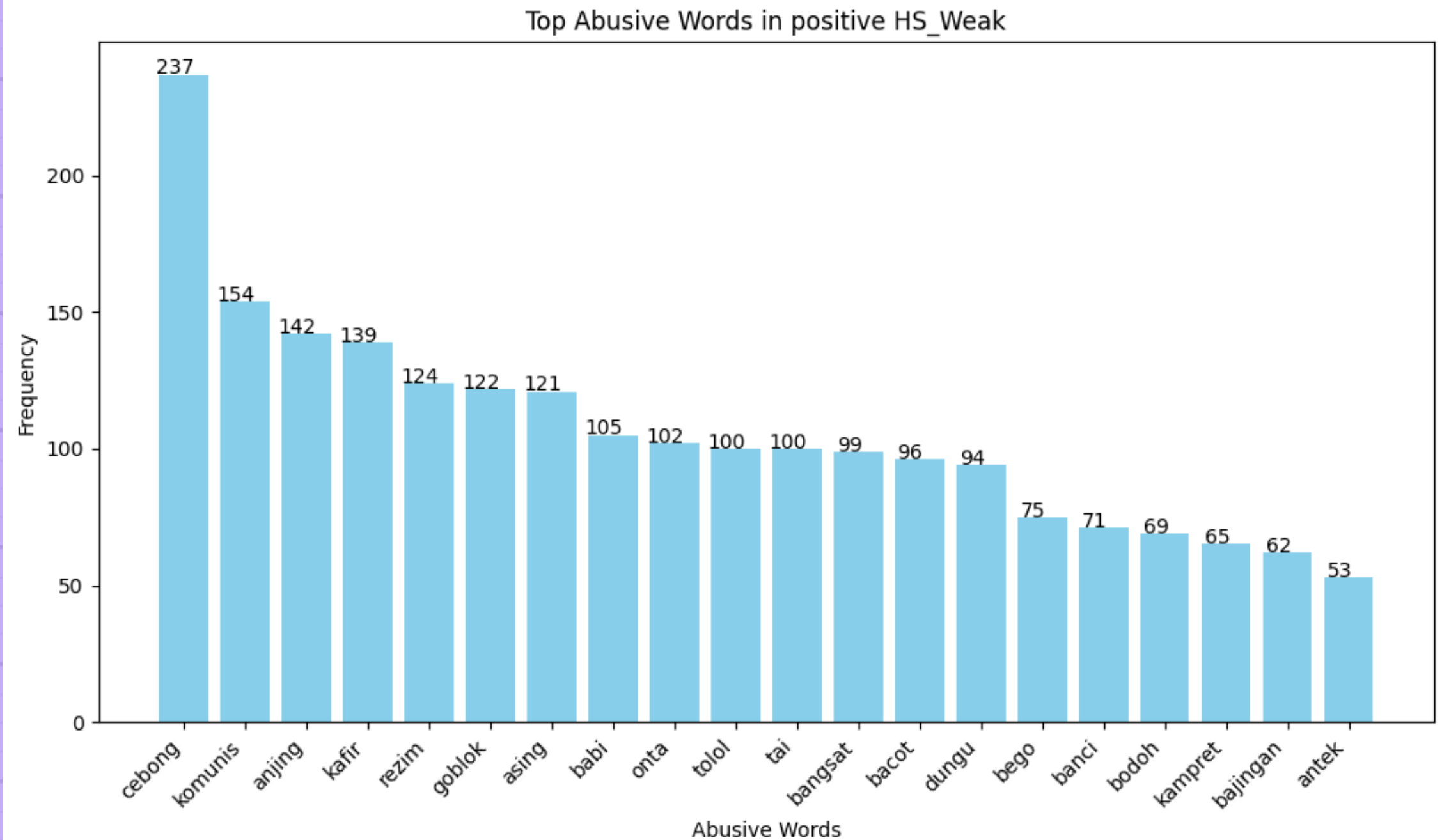
- Sumber data dari Indonesia Hatespeech Dataset bersumber dari Kaggle
- Menganalisis berdasarkan kategori HS\_Weak, HS\_Moderate, dan HS\_Strong
- Melakukan univariate analysis dengan berfokus pada tiap-tiap variabel (HS\_Weak, HS\_Moderate, dan HS\_Strong) serta melihat karakteristik kata yang muncul

Descriptive  
Analytics  
dengan  
menggunakan  
LDA

Descriptive  
Statistic dengan  
menghitung  
frekuensi  
permunculan kata

# HASIL ANALISIS

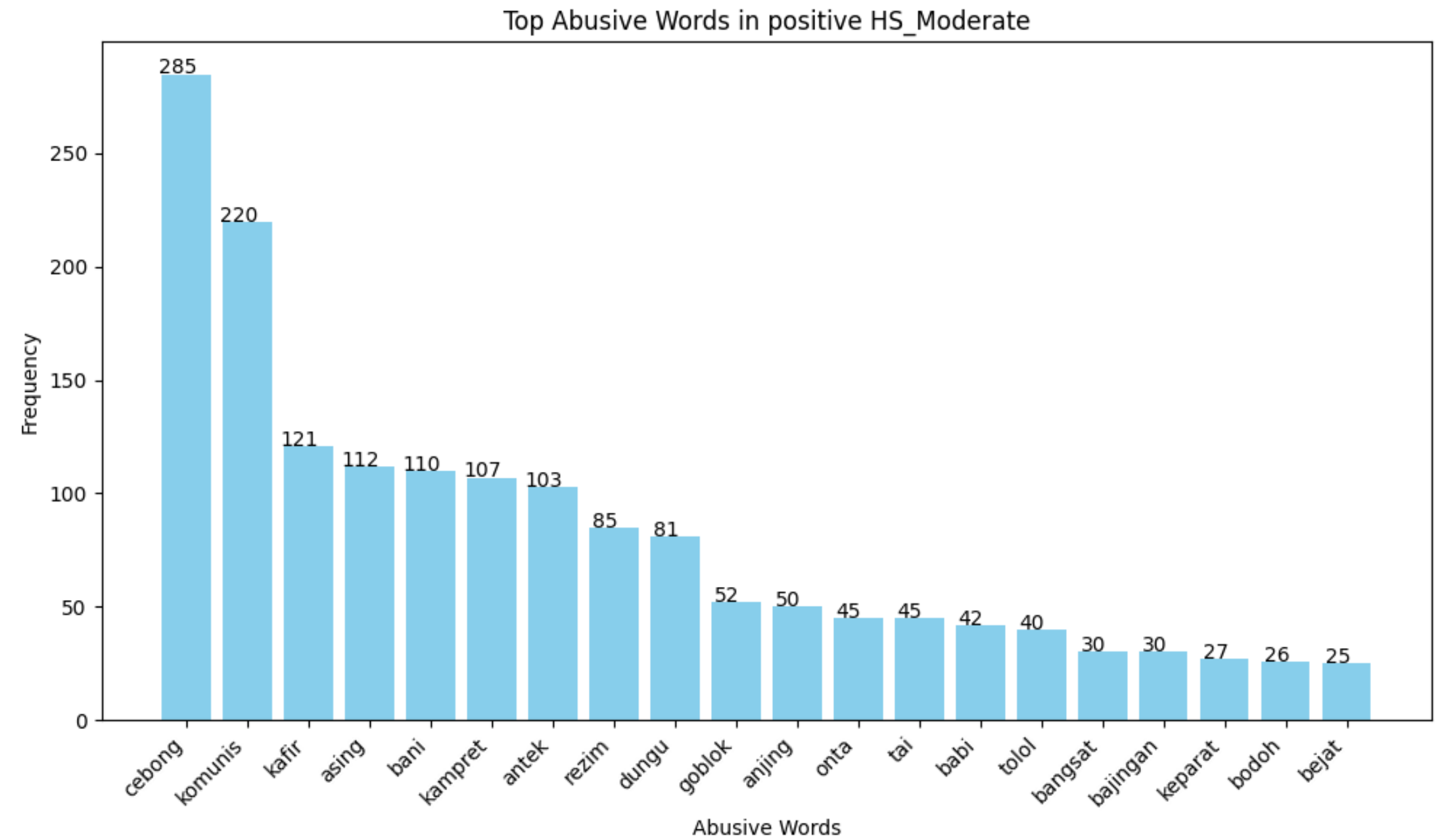
Frekuensi Permunculan kata-kata  
HS Dalam Setiap Kategori





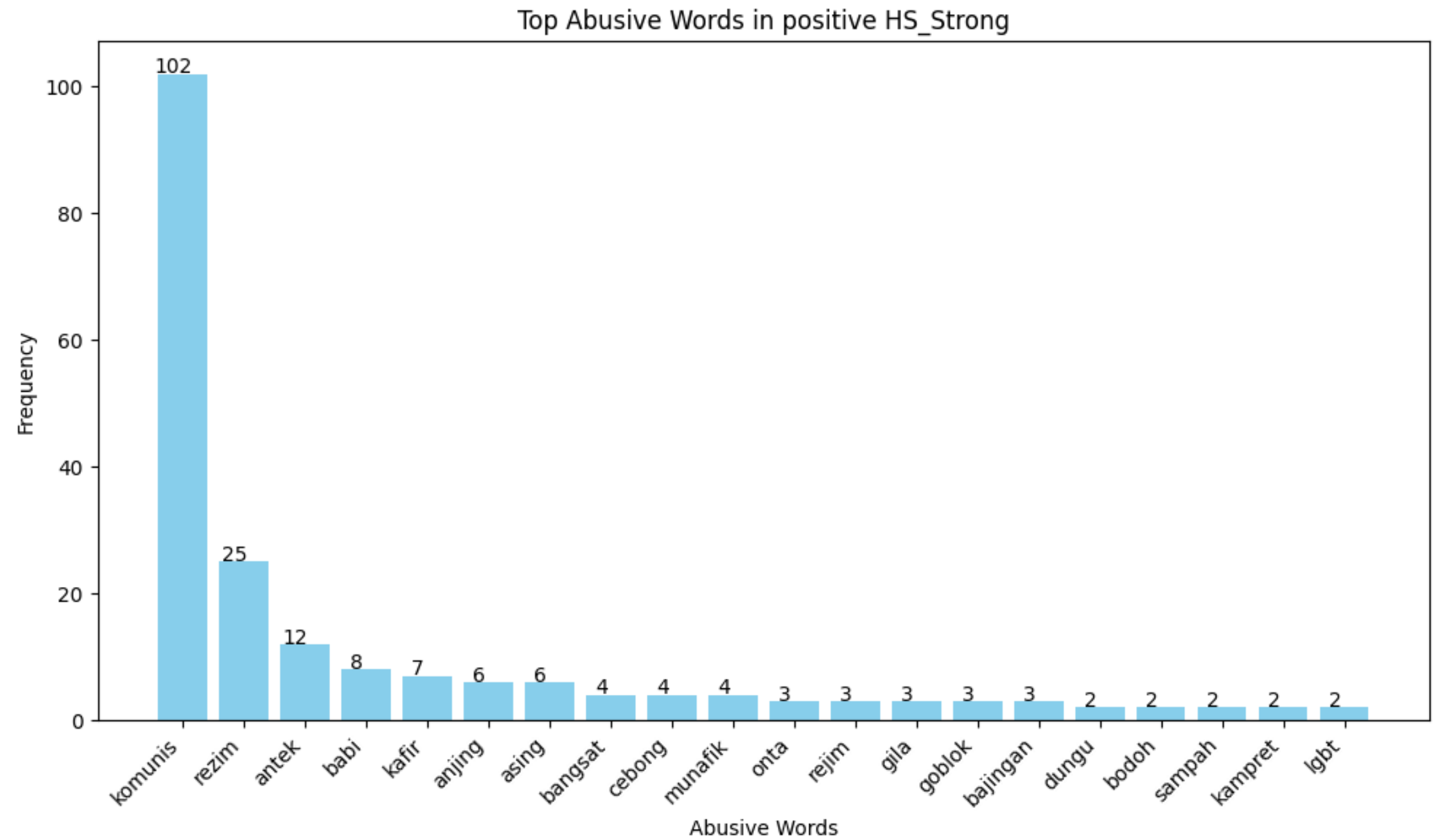
# HASIL ANALISIS

Frekuensi Permunculan kata-kata  
HS Dalam Setiap Kategori




# HASIL ANALISIS

Frekuensi Permunculan kata-kata  
HS Dalam Setiap Kategori







# TOPIC MODELLING WITH LDA



Latent Dirichlet Allocation adalah cara umum untuk melakukan topic modelling dengan mengambil topik dari sekumpulan data teks yang digabungkan (Corpus)



## TOPIC MODELLING (HS\_WEAK)

- 1 Topic 0 membahas mengenai hal yang terkait dengan presiden dan kepala pemerintahan.
- 2 Topic 1 membahas mengenai ideologi komunisme
- 3 Topic 2 membahas tentang political standing dalam dukungan kepada kepala pemerintahan

Topic 0

indonesia  
jokowi  
presiden  
negara  
gubernur  
prabowo  
sandi

Topic 1

orang  
komunis  
indonesia  
dasar  
jokowi  
yudhoyono  
ai  
pa  
anjing

Topic 2

presiden  
ganti  
2019  
kafir  
cebong  
ahok  
gue  
sih  
orang  
ya

## TOPIC MODELLING (HS\_MODERATE)

- 1 Topic 0 membahas mengenai ujaran akan political standing pilpres 2019 yang dikaitkan dengan agama
- 2 Topic 1 membahas mengenai ujaran tentang SARA yang dihubungkan dengan agama
- 3 Topic 2 membahas tentang political standing serta ujaran mengenai agama.



## TOPIC MODELLING (HS\_STRONG)

- 1 Topic 0 membahas mengenai ujaran akan sistem pemerintahan Jokowi
- 2 Topic 1 masih membahas mengenai ujaran tentang sistem pemerintahan seperti DPR
- 3 Topic 2 membahas tentang pemerintahan yang diiringi dengan SARA

Topic 0


lengserkan  
keadilan hancurkan  
undang  
tolak  
hati  
jokowi  
islam  
sejahtera  
paham

Topic 1

dewan korupsi komunis  
bubarkan  
komisi perwakilan  
indonesia  
ai rakyat pa

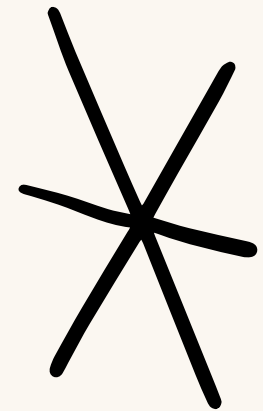
Topic 2

buddha indonesia  
cin  
usir  
negara  
republik  
jokowi  
myanmar  
lengserkan  
kesatuan



**Note!**

Metode LDA mungkin butuh pendefinisian yang lebih dalam dengan cara memunculkan banyak sample topic sehingga penarikan kesimpulan dapat menjadi lebih akurat.





# KESIMPULAN

## HASIL ANALISIS

Berdasarkan hasil analisis dapat dilihat bahwa penggunaan hate speech di Twitter tidak jauh dari konteks politik, agama dan SARA.

## AKIBAT

- Berpotensi menjadi alat politik untuk menggiring opini masyarakat
- Rentan akan penyebaran HOAX untuk kepentingan tertentu
- Menimbulkan perpecahan
- Menimbulkan user environment yang tidak nyaman

## APA YANG BISA DILAKUKAN?

- Memberikan persyaratan pembuatan akun dengan nomor ID sehingga memunculkan akuntabilitas pengguna
- Mengembangkan fitur "added context" dengan baik untuk memerangi hoax
- Melakukan topic modeling setiap 6 bulan sekali agar dapat melihat perilaku user dengan lebih baik





# THANK YOU!

**SOURCE: MUHAMMAD OKKY IBROHIM AND INDRA BUDI. 2019.  
MULTI-LABEL HATE SPEECH AND ABUSIVE LANGUAGE  
DETECTION IN INDONESIAN TWITTER**