

API TEXT CLEANSING
PROGRAM

Abusive and Hate Speech Tweet Text in Indonesia

Binar Gold Challenge

Mohammad Raditya Syahmaulana

INTRODUCTION

Survei Digital Civility Index 2020
yang dilakukan Microsoft

Indonesia merupakan salah satu negara yang
menggunakan internet paling kasar

Webinar series BNPB " Modus
Baru Cyber Crime di Tengah
Pandemi COVID-19 "

Dari data polri pada bulan april 2020,
terdapat 3 kasus dengan angka tertinggi
yang dilaporkan yaitu *provocative, hate
conten and hate speech*

13.169 tweet

Telah terkumpulah tweet warga indonesia,
apakah tweet tersebut membawa pesan
negatif atau ujaran kebencian

Defining Problems

1

Seberapa sering muncul
Abusive dan Hate Speech ?

2

Kepada siapa tweet ini
dituju?

3

Seberapa besar level tweet
yang akan dituju?

4

Kategori dan level apa yang
sering dituju dalam tweet ?

Tools



Visual Studio Code

Application Processing



Language Processing



Analyse Data

RegEx

Analyse Data



Analyse Data



Data Visualization



Data Visualization

Canva

Report Visualization



API Processing



API Visualization



Project Saving

Data Exploration

Dataset yang diterima merupakan ribuan tweet yang telah diberi label. Berdasarkan jenis dan data tersebut, data ini merupakan data bagus karena tidak ada data kosong atau null

Tweet	HS	Abusive	HS_Individual	HS_Group	HS_Religion	HS_Race	HS_Physical	HS_Gender	HS_Other	HS_Weak	HS_Moderate	HS_Strong
- disaat semua cowok berusaha melacak perhatian...	1	1	1	0	0	0	0	0	1	1	0	0
41. Kadang aku berfikir, kenapa aku tetap perc...	0	0	0	0	0	0	0	0	0	0	0	0
USER USER AKU ITU AKU\n\nKU TAU MATAMU SIPIT T...	0	0	0	0	0	0	0	0	0	0	0	0
USER USER Kaum cebong kapir udah keliatan dong...	1	1	0	1	1	0	0	0	0	0	1	0
USER Ya bani taplak dkk xf0\x9f\x98\x84\xfd\x...	1	1	0	1	0	0	0	0	1	0	1	0

```
In [73]: df.isnull().sum()
Out[73]: Tweet      0
          HS         0
          Abusive    0
          HS_Individual 0
          HS_Group    0
          HS_Religion  0
          HS_Race      0
          HS_Physical  0
          HS_Gender    0
          HS_Other     0
          HS_Weak     0
          HS_Moderate  0
          HS_Strong    0
          new_tweet    0
          dtype: int64
```

Text Cleansing

1. Mengubah karakter menjadi lowercase

```
def lowercase(s):  
    return s.lower()
```

2. Menghilangkan karakter yang tidak perlu

```
def remove_punct(s):  
    s = re.sub(r'\n', ' ', s)  
    s = re.sub('user', ' ', s)  
    s = re.sub('rt', ' ', s)  
    s = re.sub('((www\.[^\s]+)|(https?://[^\s]+)|(http?://[^\s]+))', ' ', s)  
    s = re.sub(r'^\w\d\s+', ' ', s)  
    s = re.sub(r' +', ' ', s)  
    s = re.sub(r'^\0-9a-zA-Z+', ' ', s)  
  
    return s
```

3. Normalisasikan kata alay menjadi baku dan menghilangkan kata-kata Abusive

```
db = sqlite3.connect('/Users/mohammadraditya/Kuis Gold/gabungan.db', check_same_thread=False)  
q_kamus_alay= 'select * from kamus_alay'  
t_kamus_alay= pd.read_sql_query(q_kamus_alay, db)  
q_abusive = 'select * from abusive'  
t_abusive = pd.read_sql_query(q_abusive, db)  
  
alay_dict_map = dict(zip(t_kamus_alay['anakjakartaasikasik'], t_kamus_alay['anak jakarta asyik asyik']))  
def normalize_alay(s):  
    for word in alay_dict_map:  
        return ' '.join([alay_dict_map[word] if word in alay_dict_map else word for word in s.split(' ')])  
  
p_abusive = t_abusive['ABUSIVE'].str.lower().tolist()  
def normalize_abusive(s):  
    word_list = s.split()  
    return ' '.join([s for s in word_list if s not in p_abusive])
```

Data Visualization

Total Abusive and Hate Speech

3.294

Total Abusive

2.266

Total Hate Speech

1.747

Total Tweet

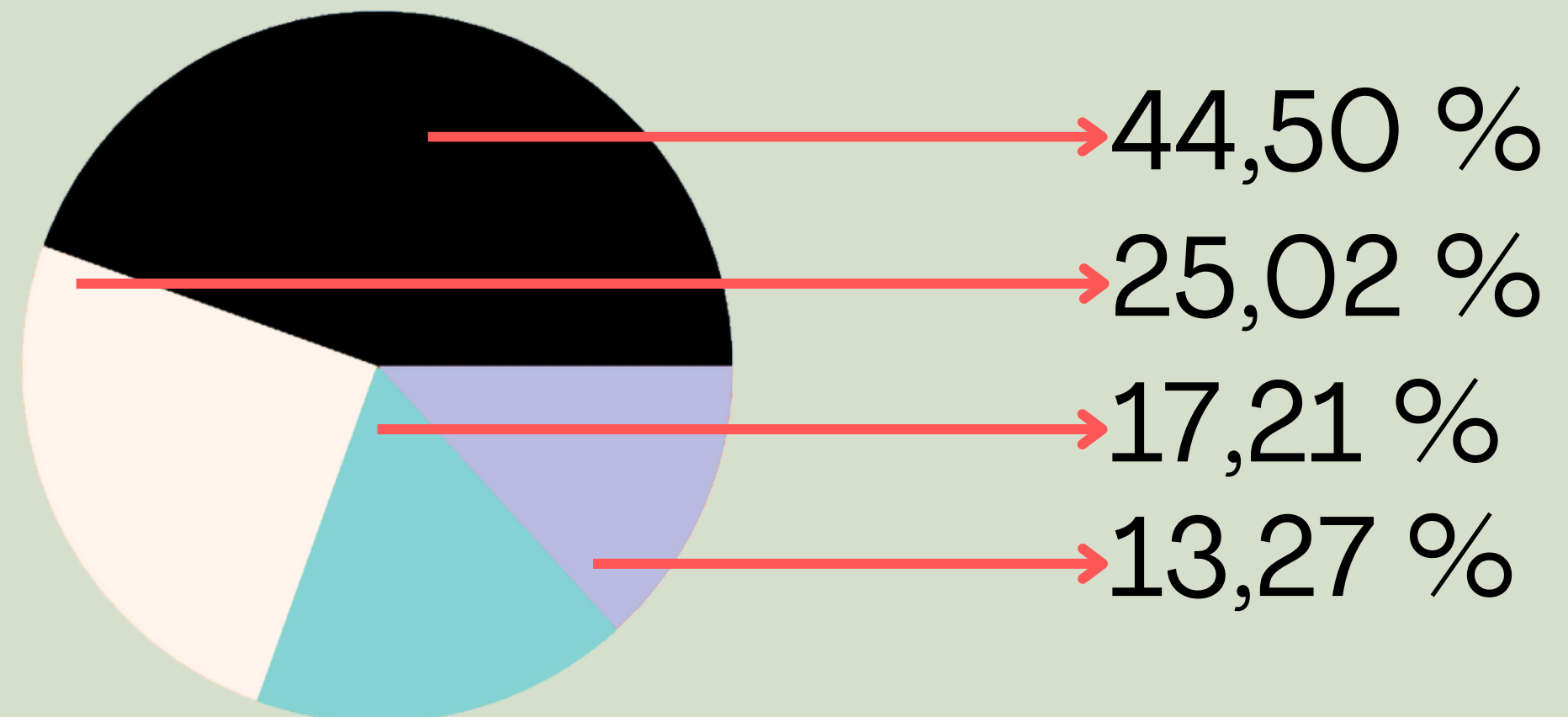
13.169

Non Abusive and Hate Speech

Abusive and Hate Speech

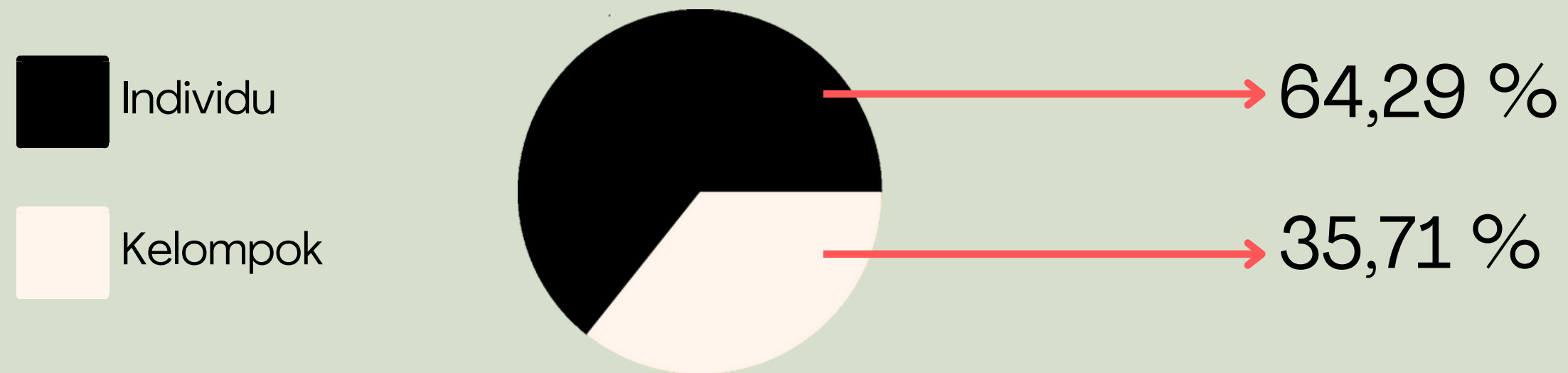
Hate Speech

Abusive

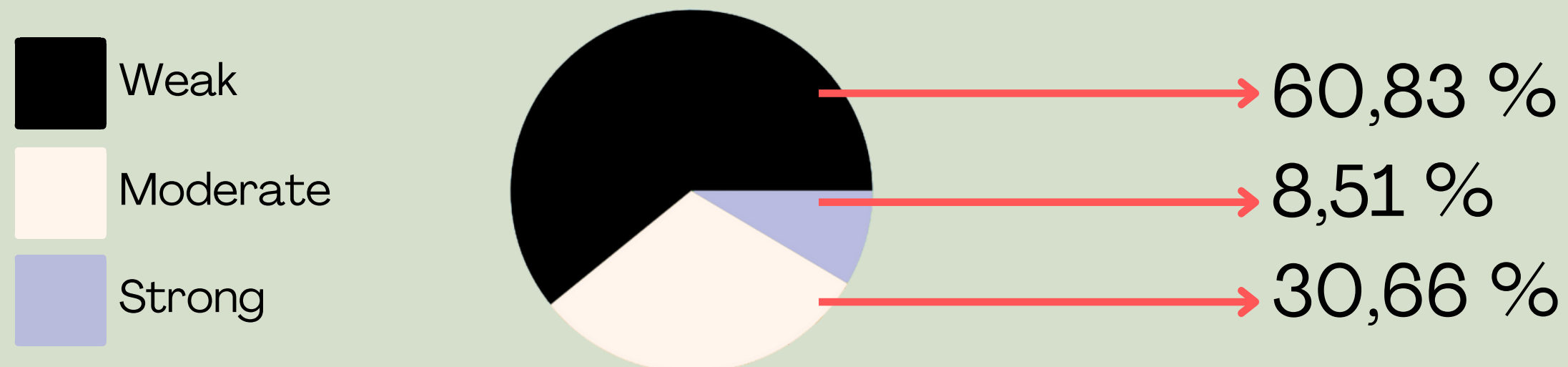


Data Visualization

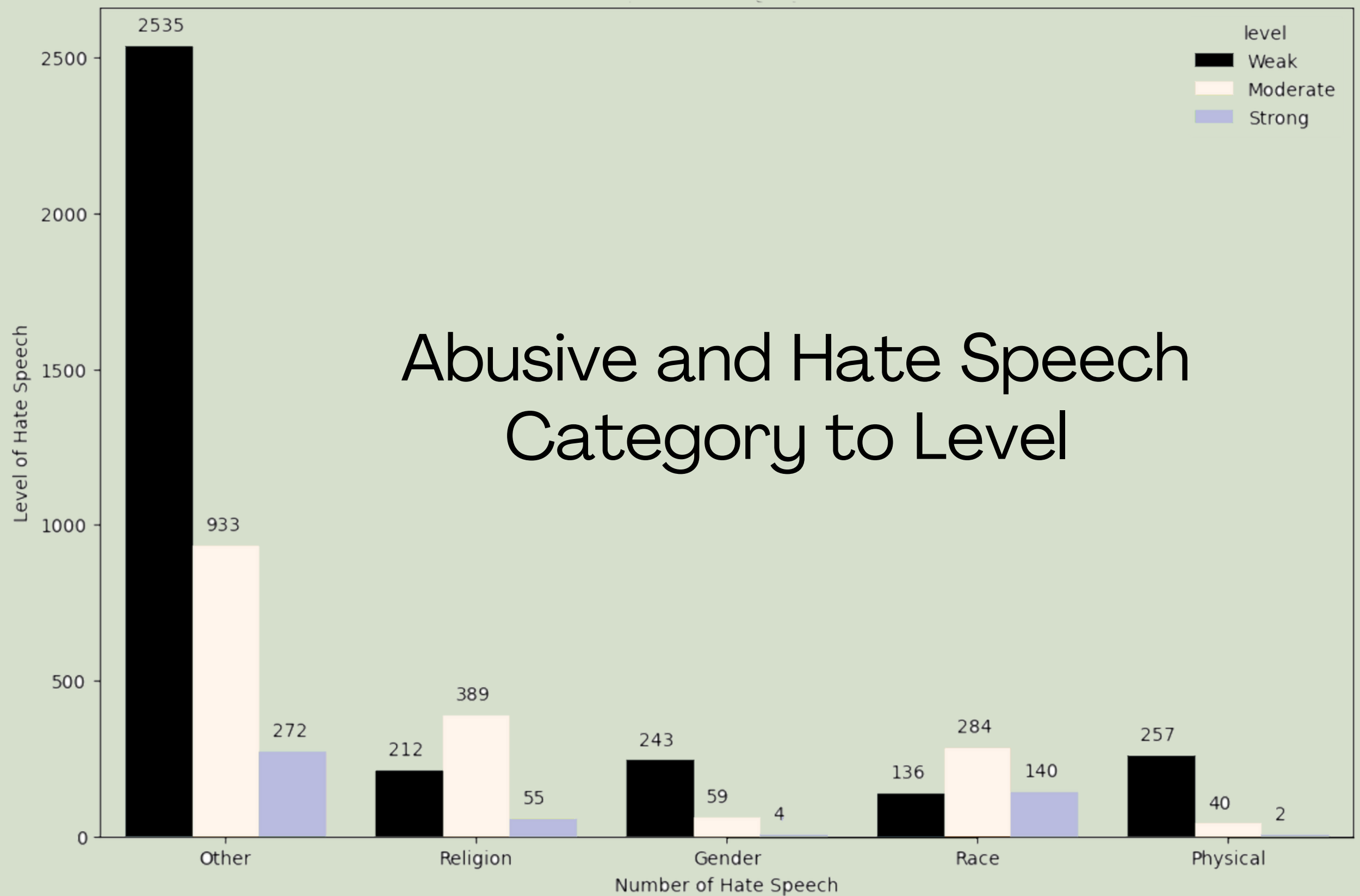
Target Abusive and Hate Speech



Abusive and Hate Speech Level



Data Visualization



Conclusions

- Dari data yang ditemukan terdapat tweet sebanyak 13.169 yang terdiri dari 44,50% tweet tanpa Abusive dan Hate Speech, 25,02% tweet menggunakan Abusive dan Hate Speech, 17,21% tweet menggunakan Hate Speech, dan 13,27 tweet menggunakan Abusive. Dilihat dari jumlahnya tweet terdapat Abusive dan Hate Speech yang sering keluar. Menurut Walters et al. menyebutkan bahwa faktor utama individu melakukannya karena adanya prasangka buruk dan iri hati
- Target yang terkena tweet Abusive dan Hate Speech paling banyak terdapat pada Individu, dan disusul oleh Group. Mengacu pada ilmu psikologi para pelaku melakukan tersebut karena bentuk kepribadian sadism
- Pada level Weak (lemah), sering kali tweet ini dituju kepada individu dan dilontarkan dengan hinaan fisik
- Pada level Moderate (Sedang), sering kali tweet ini dituju untuk suatu kelompok dan dilontarkan dengan hinaan agama
- Pada level Strong (kuat),ditujukan pada suatu kelompok dengan lontaran berbentuk hinaan ras atau etnis

Recomendation

- Setiap orang harus merasa bahwa melakukan tindakan seperti itu tidak baik, harus dapat berpikir bahwa Abusive dan Hate Speech seperti itu tidak menyenangkan apabila ditujukan pada diri sendiri.
- Mengetahui aturan Per UU ITE dan menjalankan dengan baik
- Menenangkan diri dengan tidak mengikuti suatu akun yang menimbulkan konflik dan harus melakukan report pada akun tersebut
- Kurikulum pendidikan di Indonesia perlu dikaji ulang tentang mendidik anak bangsa supaya generasi muda memiliki sopan dan santun dalam interaksi terutama pada penggunaan media sosial

Thank You

Let's Connect

You can find me on:



<https://www.linkedin.com/in/mradityas/>



<https://github.com/mradityas/>