

Descriptive Analytics terkait data Indonesian Abusive & Hate Speech Twitter

Aziz Awaludin

231000005

Binar Academy Data Science Bootcamp Wave 19

Latar Belakang

Seiring dengan berkembangnya zaman, penggunaan media sosial di kalangan masyarakat menjadi sangat masif, namun penggunaan media sosial dengan cara yang salah dapat menimbulkan permasalahan baru, salah satunya yaitu maraknya ujaran kebencian yang beredar di media sosial.

Hate Speech (Ujaran Kebencian)

Ujaran Kebencian atau Ucapan Kebencian didefinisikan sebagai ujaran, tulisan, tindakan, atau pertunjukan yang ditujukan untuk menghasut kekerasan atau prasangka terhadap seseorang atas dasar karakteristik kelompok tertentu yang dianggap ia wakili, seperti kelompok ras, etnis, gender, orientasi seksual, agama, dan lain-lain.

Dalam arti hukum, *hate speech* adalah perkataan, perilaku, tulisan, ataupun pertunjukan yang dilarang karena dapat memicu terjadinya tindakan kekerasan dan sikap prasangka entah dari pihak pelaku pernyataan tersebut ataupun korban dari tindakan tersebut. (sumber : [Pengertian Ujaran Kebencian \(Hate Speech\) dan Contohnya | Blog Komunikasi Jurnalistik Media Humas \(komunikasipraktis.com\)](#))

Tujuan Penelitian

- Melakukan analisis deskriptif sederhana terhadap data Indonesian Abusive & Hate Speech Twitter Text
- Mengetahui presentase dan jumlah sentiment positif dan negative dari total data tweet pada Indonesian Abusive & Hate Speech Twitter Text
- Mengetahui presentase tweet Hate Speech dan Abusive pada data Indonesian Abusive & Hate Speech Twitter Text
- Mengetahui siapa target dari tweet yang mengandung Hate Speech pada data Indonesian Abusive & Hate Speech Twitter Text
- Mengetahui topik dari tweet yang mengandung Hate Speech pada data Indonesian Abusive & Hate Speech Twitter Text

Metode Penelitian

Metode analisis yang digunakan pada penelitian ini, yaitu Descriptive Analytics, yaitu dengan mendeskripsikan fakta yang terdapat di dalam data Indonesian Abusive & Hate Speech Twitter Text dengan menggunakan visualisasi sederhana pada jupyter notebook.

Exploratory Data Analysis

df													
	Tweet	HS	Abusive	HS_Individual	HS_Group	HS_Religion	HS_Race	HS_Physical	HS_Gender	HS_Other	HS_Weak	HS_Moderate	HS_Strong
0	- disaat semua cowok berusaha melacak perhatia...	1	1	1	0	0	0	0	0	1	1	0	0
1	RT USER: USER siapa yang telat ngasih tau elu?...	0	1	0	0	0	0	0	0	0	0	0	0
2	41. Kadang aku berfikir, kenapa aku tetap perc...	0	0	0	0	0	0	0	0	0	0	0	0
3	USER USER AKU ITU AKU\n\nKU TAU MATAMU SIPIT T...	0	0	0	0	0	0	0	0	0	0	0	0
4	USER USER Kaum cebong kapir udah keliatan dong...	1	1	0	1	1	0	0	0	0	0	1	0
...
13164	USER jangan asal ngomong ndasmu. congor lu yg ...	1	1	1	0	0	0	1	0	0	1	0	0
13165	USER Kasur mana enak kunyuk'	0	1	0	0	0	0	0	0	0	0	0	0
13166	USER Hati hati bisu :(.g\n\nlagi bosan huft \...	0	0	0	0	0	0	0	0	0	0	0	0
13167	USER USER USER USER Bom yang real mudah terdet...	0	0	0	0	0	0	0	0	0	0	0	0
13168	USER Mana situ ngasih('': itu cuma foto ya kuti...	1	1	1	0	0	0	0	0	1	1	0	0

13169 rows × 13 columns

Berikut adalah data yang digunakan pada penelitian ini

Exploratory Data Analysis

Dilakukan pengecekan terhadap data yang digunakan dari mulai info terkait dataframe, info mengenai kolom dengan value null dan info mengenai duplicated value, dan menghilangkan duplicated value dengan `df.drop_duplicates()`

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 13169 entries, 0 to 13168  
Data columns (total 13 columns):  
#   Column          Non-Null Count  Dtype    
---  -  
0   Tweet           13169 non-null  object   
1   HS              13169 non-null  int64    
2   Abusive         13169 non-null  int64    
3   HS_Individual   13169 non-null  int64    
4   HS_Group        13169 non-null  int64    
5   HS_Religion     13169 non-null  int64    
6   HS_Race         13169 non-null  int64    
7   HS_Physical     13169 non-null  int64    
8   HS_Gender       13169 non-null  int64    
9   HS_Other        13169 non-null  int64    
10  HS_Weak         13169 non-null  int64    
11  HS_Moderate     13169 non-null  int64    
12  HS_Strong       13169 non-null  int64    
dtypes: int64(12), object(1)  
memory usage: 1.3+ MB
```

```
df.isnull().sum()
```

```
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  
dtype: int64
```

```
df.duplicated().sum()
```

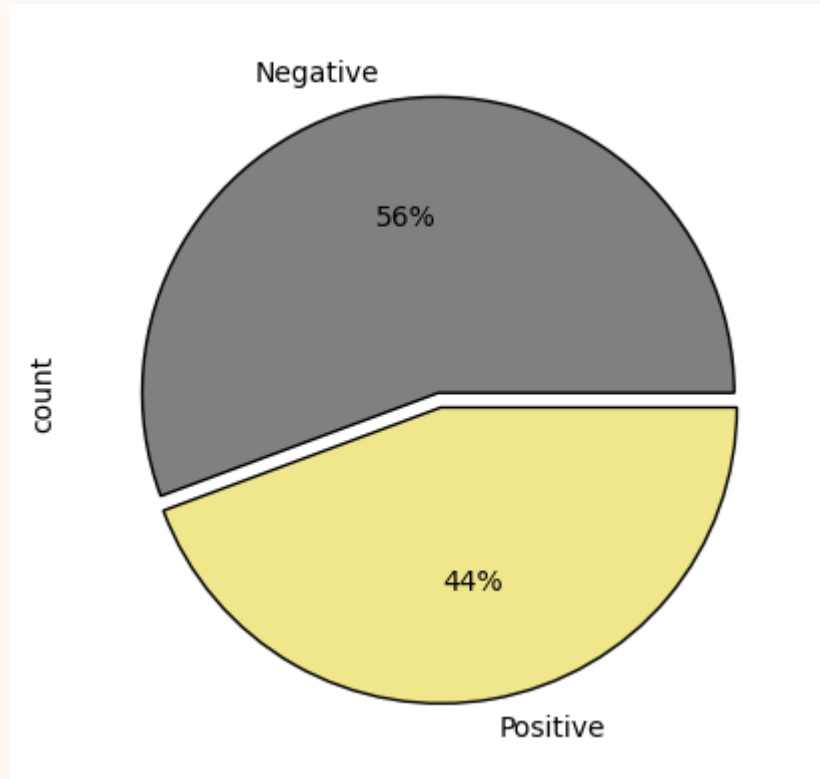
```
125
```

```
df = df.drop_duplicates()
```

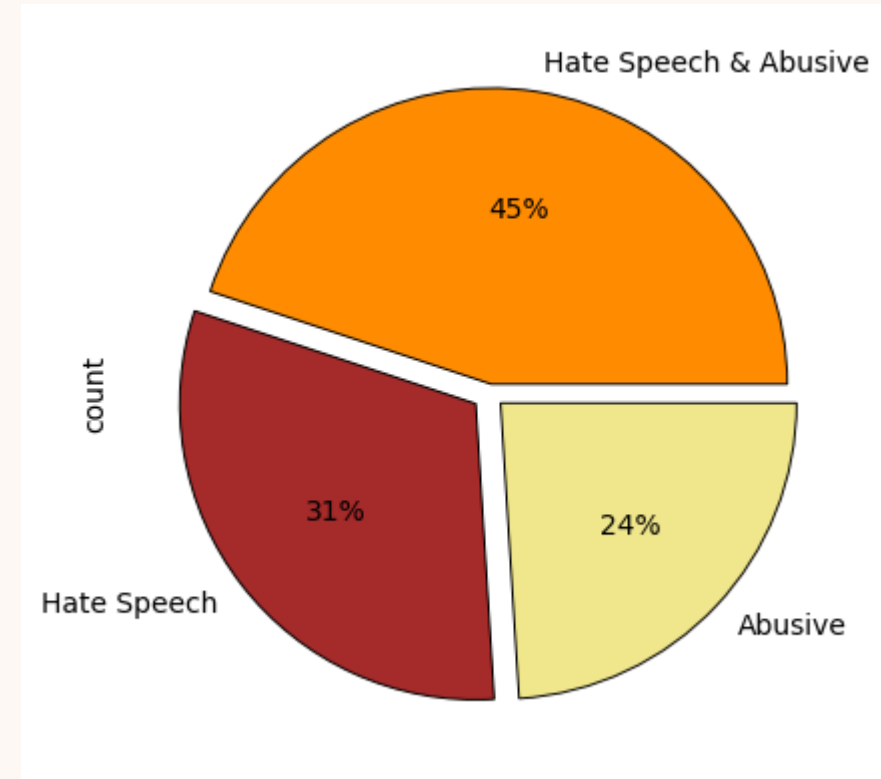
```
df.duplicated().sum()
```

```
0
```

Exploratory Data Analysis

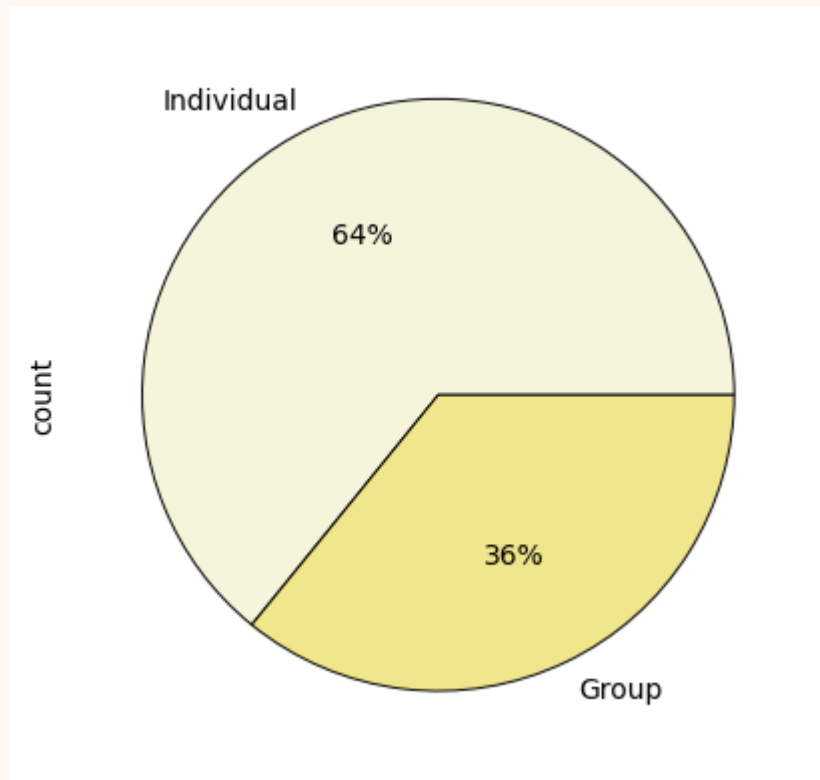


Presentase sentiment positif dan negative dari tweet pada data terkait

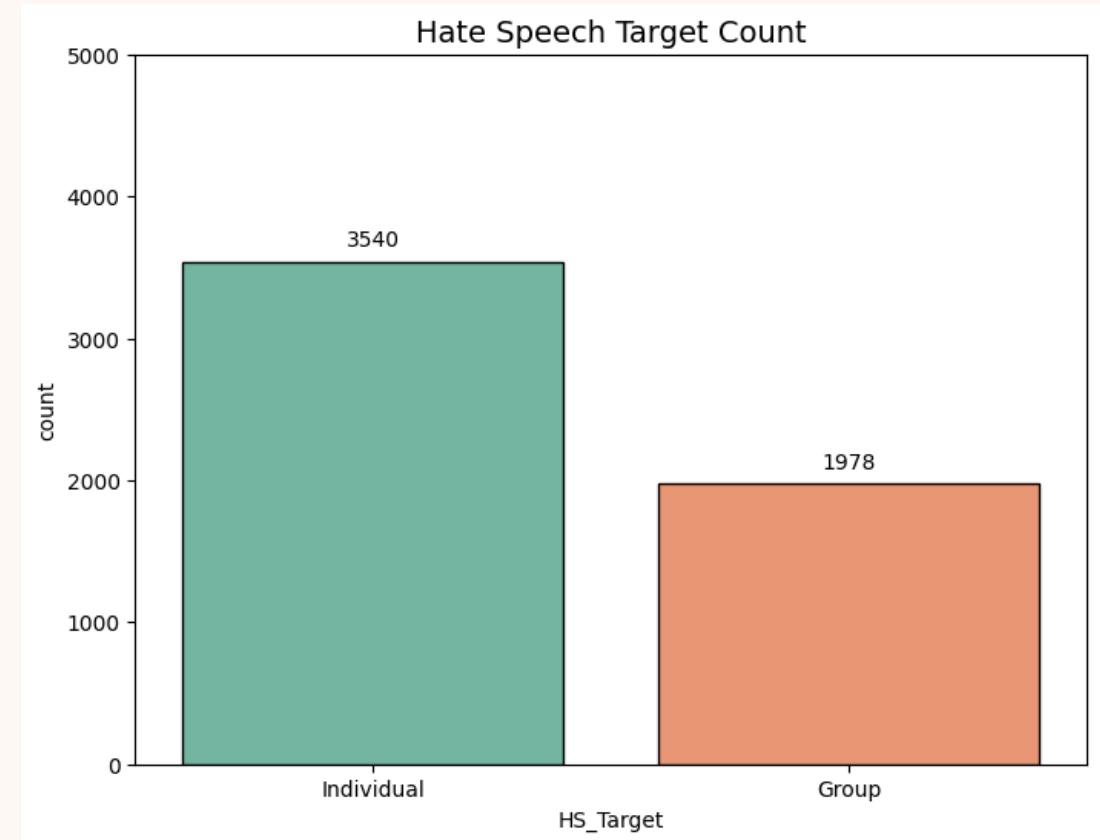


Presentase tweet yang mengandung Hate Speech, Abusive, dan keduanya

Exploratory Data Analysis

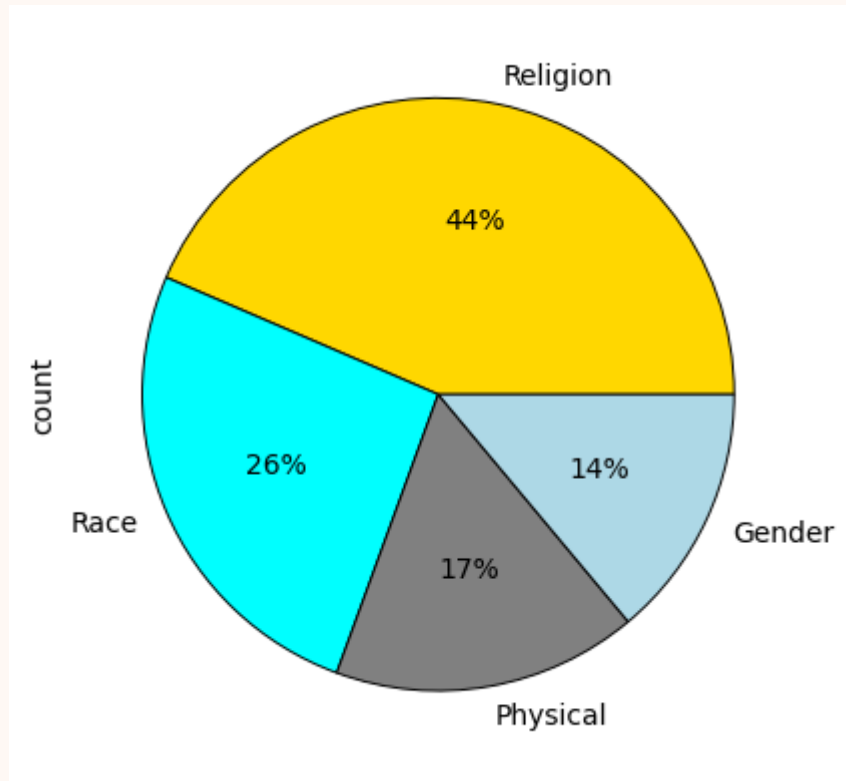


Presentase tweet target Hate Speech pada data terkait

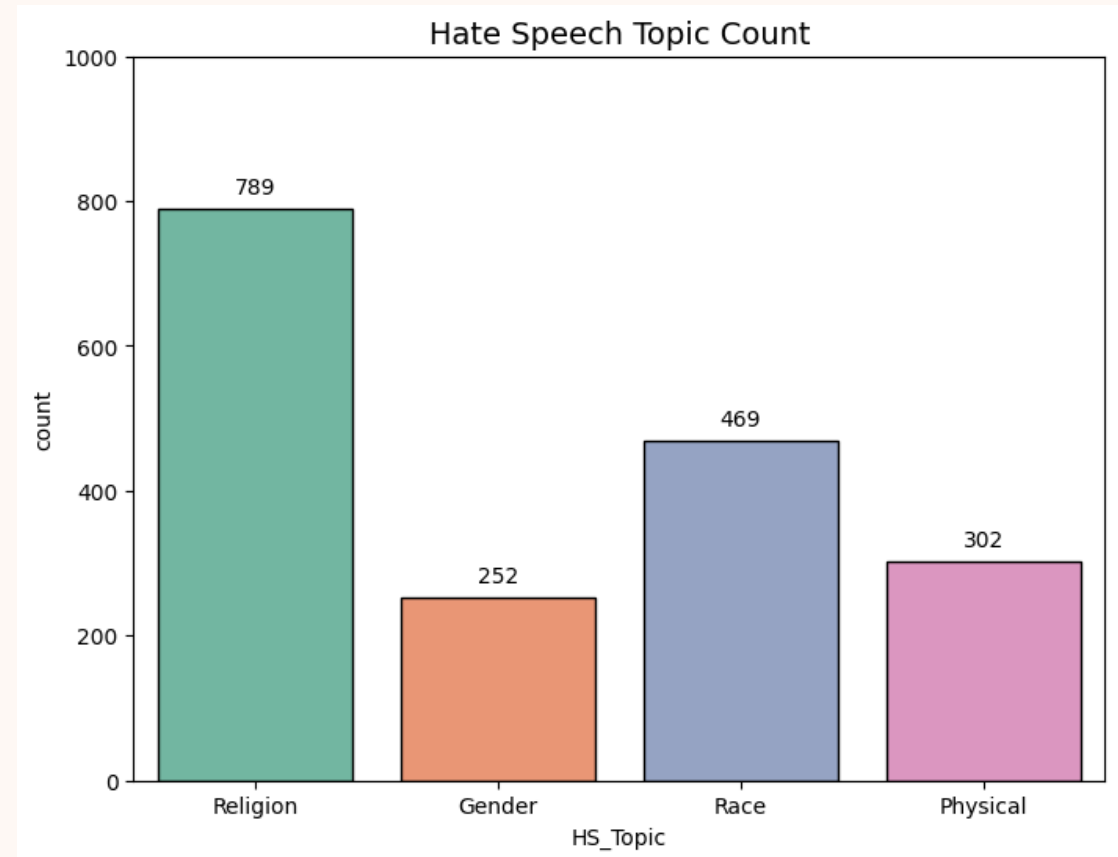


Jumlah tweet target Hate Speech pada data terkait

Exploratory Data Analysis



Presentase tweet topic Hate Speech pada data terkait



Jumlah tweet topic Hate Speech pada data terkait

Exploratory Data Analysis

	Tweet	Sentiment	Type	HS_Target	HS_Topic
0	di saat semua cowok berusaha melacak perhatia...	Negative	Hate Speech & Abusive	Individual	Other
1	siapa yang telat memberi tau kamu edan sarap ...	Negative	Abusive	Other	Other
2	41 kadang aku berpikir kenapa aku tetap percay...	Positive	Neutral	Other	Other
3	aku itu aku dan ku tau matamu sipit tapi dili...	Positive	Neutral	Other	Other
4	kaum cebong kafir sudah kelihatan dongoknya d...	Negative	Hate Speech & Abusive	Group	Religion
5	ya bani taplak dan kawan kawan xf0 x9f x98 x8...	Negative	Hate Speech & Abusive	Group	Other
6	deklarasi pilihan kepala daerah 2018 aman dan ...	Positive	Neutral	Other	Other
7	gue baru saja selesai re watch aldnoah zero pa...	Negative	Abusive	Other	Other
8	nah admin belanja satu lagi po terbaik nak mak...	Positive	Neutral	Other	Other
9	enak lagi kalau sambil ngewe	Negative	Abusive	Other	Other
10	setidaknya gue punya jari tengah buat kamu seb...	Negative	Hate Speech & Abusive	Individual	Other
11	banci kaleng malu tidak bisa jawab pe anyaan ...	Negative	Hate Speech & Abusive	Individual	Gender
12	kalau belajar ekonomi mestinya jago memprivati...	Positive	Neutral	Other	Other
13	aktor huru hara 98 prabowo si ingin lengserkan...	Negative	Hate Speech	Individual	Other
14	bu guru enakan jadi jablay atau guru sekolah ...	Negative	Hate Speech & Abusive	Individual	Gender

Berikut adalah data yang melalui tahapan cleaning pada penelitian ini

Kesimpulan

Berdasarkan analisis deskriptif yang dilakukan pada penelitian ini, diketahui bahwa:

- Terdapat 13169 baris dan 13 kolom pada dataframe
- Terdapat 125 data duplikat dan tidak terdapat kolom dengan value null
- 56% dari total tweet, yaitu sebanyak 7261 tweet memiliki sentimen negatif
- 24% dari total tweet dengan sentimen negatif mengandung perkataan abusive (1743 tweet), 31% dari total tweet dengan sentimen negatif mengandung perkataan hate speech (2256 tweet), dan 45% dari total tweet dengan sentimen negatif mengandung keduanya (3262 tweet)
- Tweet yang mengandung hate speech yang ditujukan pada individu lebih besar dibandingkan hate speech yang ditujukan pada kelompok, yaitu 64% berbanding 36%
- Topik hate speech yang paling banyak yaitu bahasan mengenai agama, disusul dengan ras, fisik, dan gender pada urutan terakhir