



Binar Academy – Data Science
Gold Challenge

DARURAT HATE SPEECH!!!

**Indonesia mendapat
gelar Netizen kasar!**

ZULFIKAR RAMADHAN

https://github.com/zulfikarrmdhn/binar_gold_challenge/



HateSpeech & Abusive!!!

- Dalam pengamatan media sosial, khususnya Twitter, sering kali kita jumpai ujaran kebencian dan bahasa kasar yang dilontarkan oleh netizen.
- Hal tersebut perlu dihindarkan agar tidak adanya konflik baik antar individu maupun kelompok tertentu.



Dataset :

Indonesian Abusive and Hate Speech Twitter Text

Multi-Labeled Hate Speech and Abusive Indonesian Twitter Text by okkyibrohim



<https://www.kaggle.com/datasets/ilhamfp31/indonesian-abusive-and-hate-speech-twitter-text?resource=download>

Tools :



Visual Studio Code
Application Processing



Python
Language Processing



Pandas
Analyze Data



Seaborn
Data Visualization

Matplotlib
Data Visualization



RegEx
Analyze Data



Flask
API Processing



Swagger
Supported by SMARTBEAR

Swagger UI
API Visualization



SQLite

sqlite3
Analyze Data




Canva
Report Visualization

GitHub
Project Saving



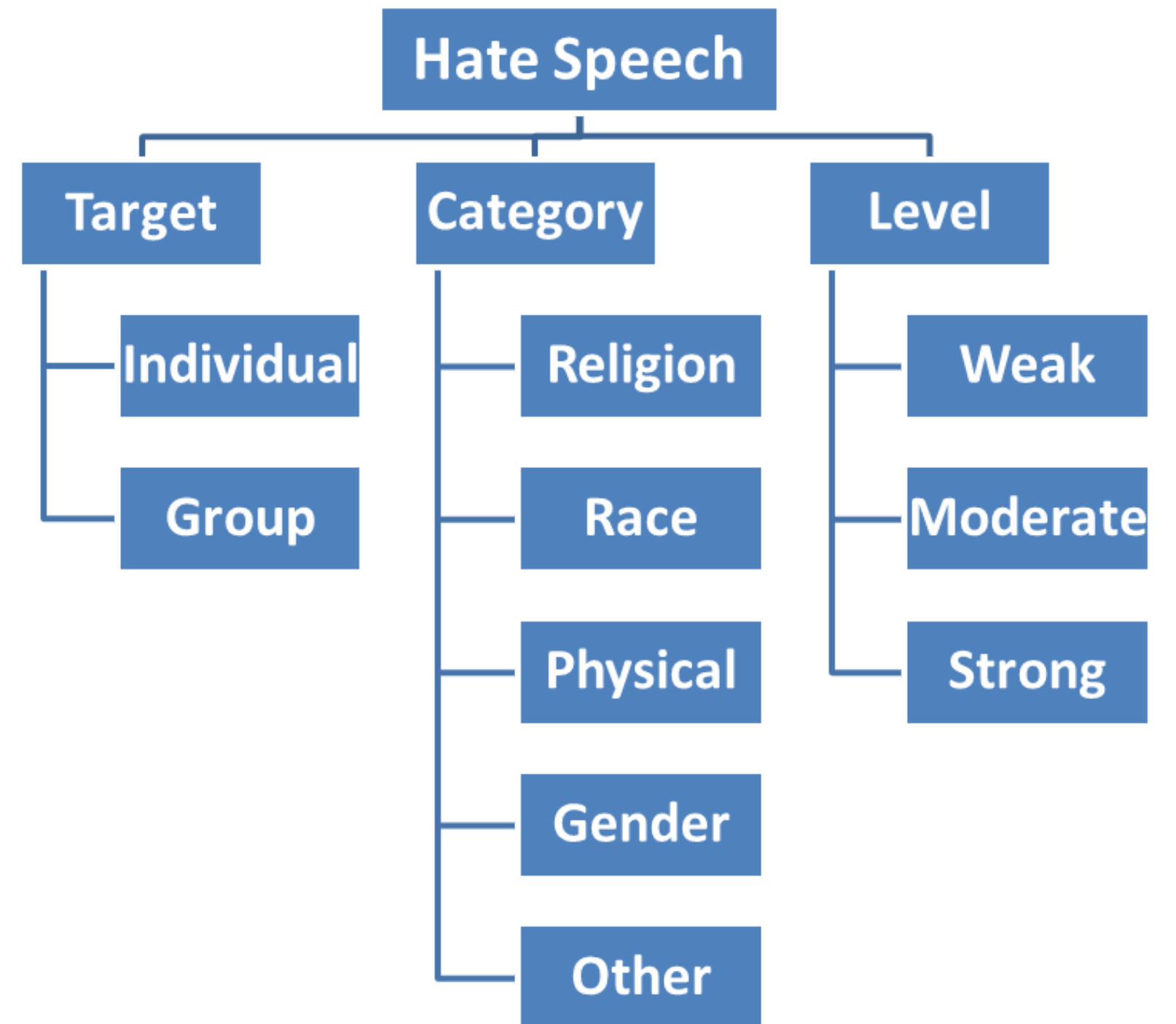
- 
- **Seberapa sering Hate Speech atau Bahasa Abusive muncul?**
 - **Kepada siapa tweet Hate Speech dengan level tertentu tersebut ditujukan?**

- 
- **Kategori apa yang paling sering muncul dalam tweet Hate Speech dengan level tertentu?**
 - **Apakah tweet Hate Speech yang muncul berpotensi menimbulkan konflik?**

Defining Problem

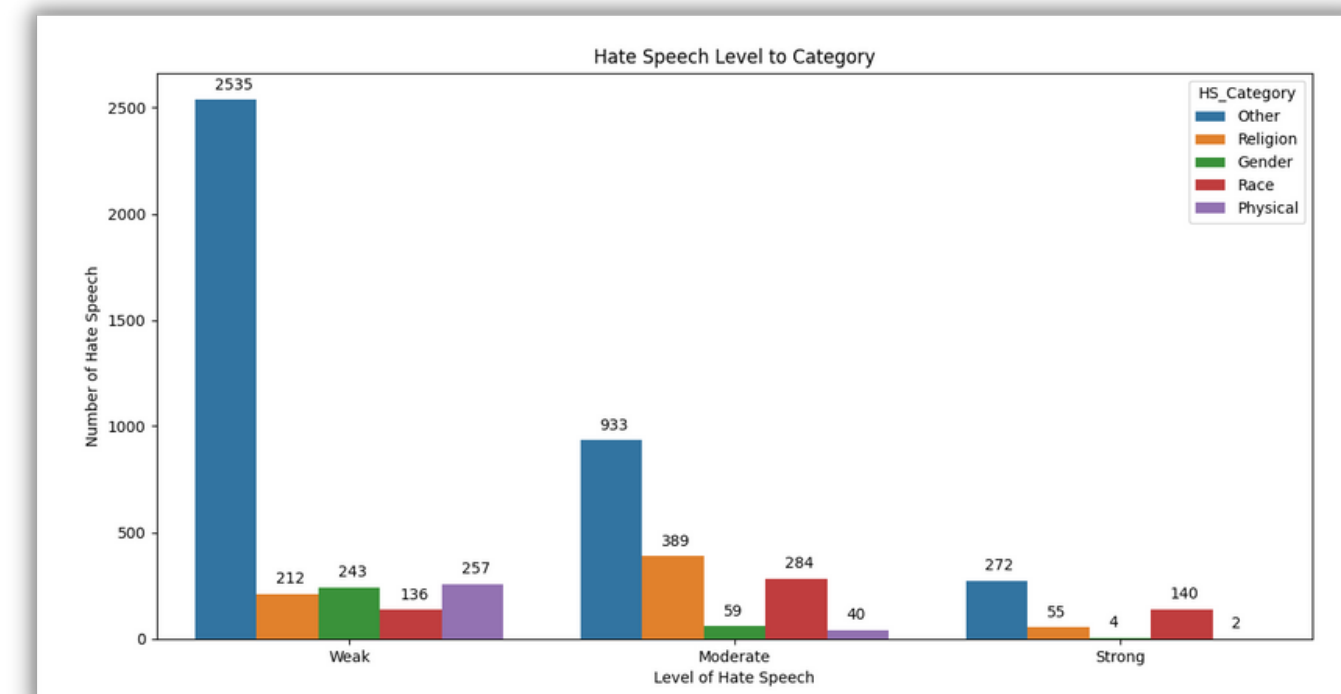
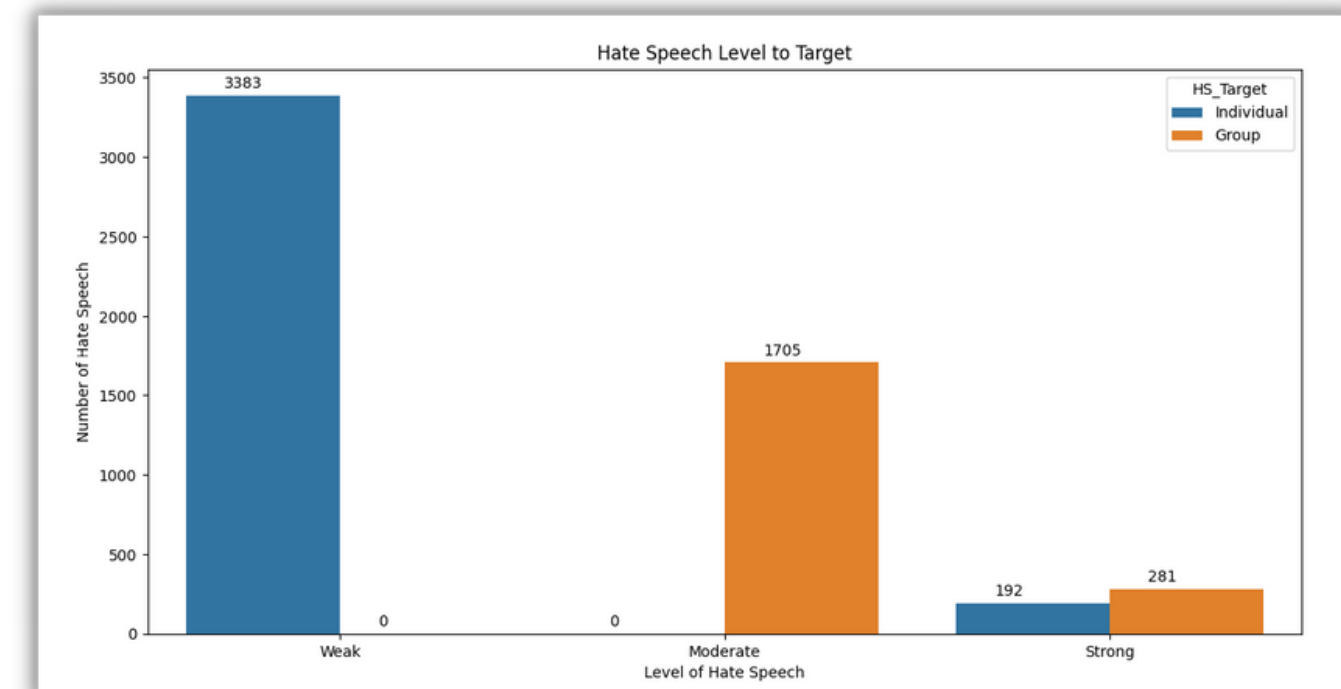
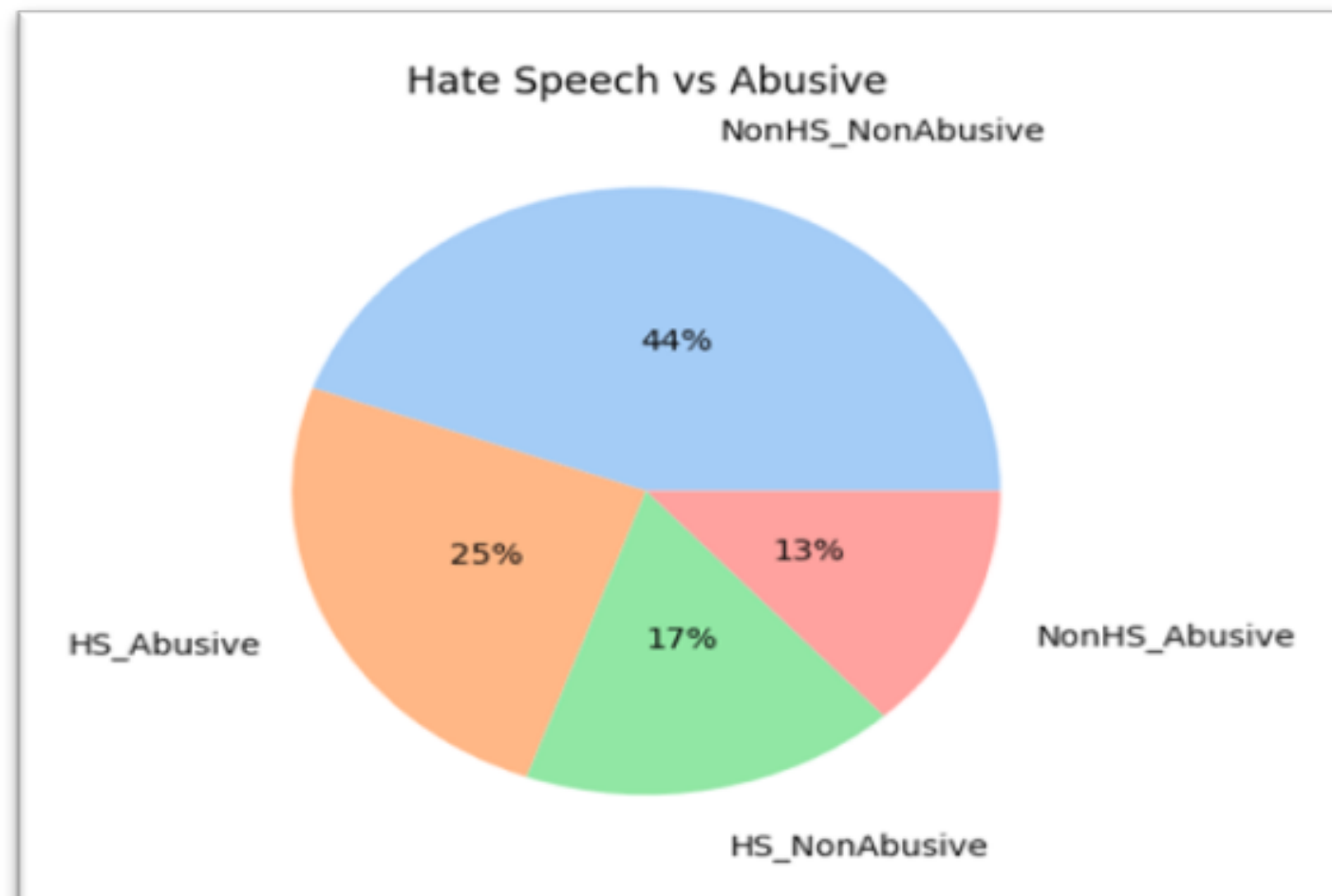
Data Exploration

- **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_Gender** : hate related to other invective/slander;
- **HS_Weak** : weak hate speech;
- **HS_Moderate** : moderate hate speech;
- **HS_Strong** : strong hate speech.



Indonesian HateSpeech Tweet Dashboard

Tweet		13.169	
Hate Speech Tweet	5.561 (42,23%)	Abusive Tweet	5.043 (38,29%)



Conclusions

- Dari 13.169 Tweet, sebanyak 5.561 Tweet (42,23%) merupakan HateSpeech, dan ada 5.043 Tweet (38,29%) yang menggunakan kata-kata kasar/ bahasa abusive.
 - Tidak heran kita seringkali menemukan tweet yang bernuansa negatif.
- Pada level weak (lemah), Tweet HateSpeech seringkali ditujukan kepada individu dan dilontarkan dengan hinaan fisik.
- Pada level moderate (sedang), Tweet HateSpeech seringkali ditujukan kepada suatu kelompok dan dilontarkan dengan hinaan agama.
- Pada level strong (kuat), Tweet HateSpeech seringkali ditujukan kepada suatu kelompok dan dilontarkan dengan hinaan ras atau etnis.
- Tweet HateSpeech dengan level strong yang dapat menimbulkan kekacauan jarang dijumpai. Hanya sekitar 473 Tweet atau 3,6% dari keseluruhan Tweet, namun hal tersebut tetap harus ditindak dan diarahkan sehingga persentasi tersebut berkurang dan tidak menimbulkan konflik yang nyata.

Recomendation

- Setiap orang perlu memperbaiki sikapnya dalam bersosial media.
- Ketahuilah aturan aturan yang berlaku pada UU ITE, atur penggunaan sosial media dengan lebih positif.
- Hindari follow akun-akun yang menimbulkan konflik. Jika perlu, kita bisa report akun tersebut sehingga tidak ada lagi konten ujaran kebencian.
- Kurikulum pendidikan di Indonesia perlu dikaji ulang untuk mendidik anak bangsa supaya memiliki sopan santun dalam penggunaan media sosial.
- Selalu disebarakan terkait etika sosial media yang berfokus pada hak dan kebebasan dalam menciptakan masyarakat yang damai.

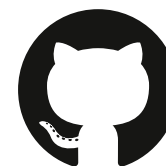
THANK YOU!

Let's Connect!

You can find me on:



<https://www.linkedin.com/in/zulfikarramadhan22/>



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



<https://www.instagram.com/rzulfikarr/>

