DataGlacier NLP Interns Group

MEMBER DETAILS

- Member 1:
 - Name Gabriel Aluede
 - Email ghabyaluede@gmail.com
 - Country United Kingdom
 - College Teesside University
 - Specialization NLP
- ➤ Member 2:
 - Name Yosha Udaya Shriyan
 - Email yosha.shriyan@gmail.com
 - Country United Kingdom
 - College Royal Holloway University of London
 - Specialization NLP

PROBLEM DESCRIPTION

The term hate speech is understood as any type of verbal, written or behavioural communication that attacks or uses derogatory or discriminatory language against a person or group based on what they are, in other words, based on their religion, ethnicity, nationality, race, colour, ancestry, sex or another identity factor. In this problem, we will take you through a hate speech detection model with Machine Learning and Python.

DATA UNDERSTANDING

Data Attributes

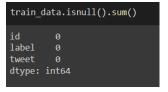
Label: 0 or 1

Text_Format : original tweets with noise

Source: https://www.kaggle.com/datasets/vkrahul/twitter-hate

speech?select=train_E6oV3lV.csv

NA Values: None



Train_Data



Test Data



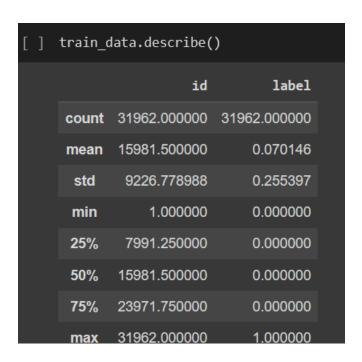
Data Size

```
sum(train_data["label"]==1)

2242

sum(train_data["label"]==0)

29720
```



GITHUB REPO LINK

https://github.com/YoshaRHUL/DataGlacierProject