Group Name: Sedats
Name: Sedat CAN (one-member group)
Email: sedatcancan1@gmail.com
Country: Germany Specialization: Data Science, NLP, Data Analyst

https://github.com/sedat-can/Final-project-

Report

<Twitter Hate Speach Detection>

04/01/2023

Agenda

Executive Summary

Problem Statement

EDA

Model

Final

Executive Summary

The hate speech detection model evaluate and classify whether tweets include hate speech and filter these tweets to offer a better user experience.

Logistic Regression is the most efficent model for hate speech detection.

Problem Statement

, Twitter users are harassed and discriminated against because of their religion, ethnicity, nationality, race, color, ancestry, sex, etc. The hate speech detection model would evaluate and classifies whether tweets include hate speech or not and filters these tweets to offer a better user experience.

EDA

The data:

10197*3 values.

The three columns:

id, tweet and label.

‘tweet’:

the user, text and hashtag, special characters ‘(‘ ‘&’,

EDA

Data Glacier
Your Deep Learning Partner

Clear data

```
df['cleartweet'] = df['tweet'].apply(lambda  tweet: re.sub("(@[A-Za-z0-
9]+)|(#)|(RT[\s]+)|(https?:\/\/\S+)|([^a-zA-Z0-9 -])", "", tweet))
```

Lemmitizer

Tweet before processed

tweet
#studiolife #aislife #requires #passion #dedic...
@user #white #supremacists want everyone to s...
safe ways to heal your #acne!! #altwaystohe...
is the hp and the cursed child book up for res...
3rd #bihday to my amazing, hilarious #nephew...

Processed tweet data

lematizer
studiolife aislife requires passion dedication...
white supremacists want everyone to see the ...
safe ways to heal your acne altwaystoheal h...
is the hp and the cursed child book up for res...
rd bihday to my amazing hilarious nephew el...

Model

Random Forest

```
[[400  55]
 [ 70 372]]
              precision    recall  f1-score   support

           0       0.85      0.88      0.86       455
           1       0.87      0.84      0.86       442

    accuracy                           0.86       897
   macro avg       0.86      0.86      0.86       897
weighted avg       0.86      0.86      0.86       897
```

Model

GradientBoostingClassifier

	precision	recall	f1-score	support
0	0.84	0.84	0.84	455
1	0.84	0.83	0.84	442
accuracy			0.84	897
macro avg	0.84	0.84	0.84	897
weighted avg	0.84	0.84	0.84	897

Model

Logistic Regression

```
[[407  48]
 [ 62 380]]
```

	precision	recall	f1-score	support
0	0.87	0.89	0.88	455
1	0.89	0.86	0.87	442
accuracy			0.88	897
macro avg	0.88	0.88	0.88	897
weighted avg	0.88	0.88	0.88	897

Final

The Best Model:

Logistic Regression

Data Glacier
Your Deep Learning Partner

Sedat Can

Data Glacier

Your Deep Learning Partner

Thank You