Hate speech detection of streaming tweets

Prachi Naik
MS2022019

Problem Statement

- Detecting tweets containing offensive/hateful words
- Streaming tweets

Approach

1. Twitter developer API (Connection by two methods)
2. Preprocessing (Removing stop words, special characters, etc.)
3. PySpark
4. TextBlob

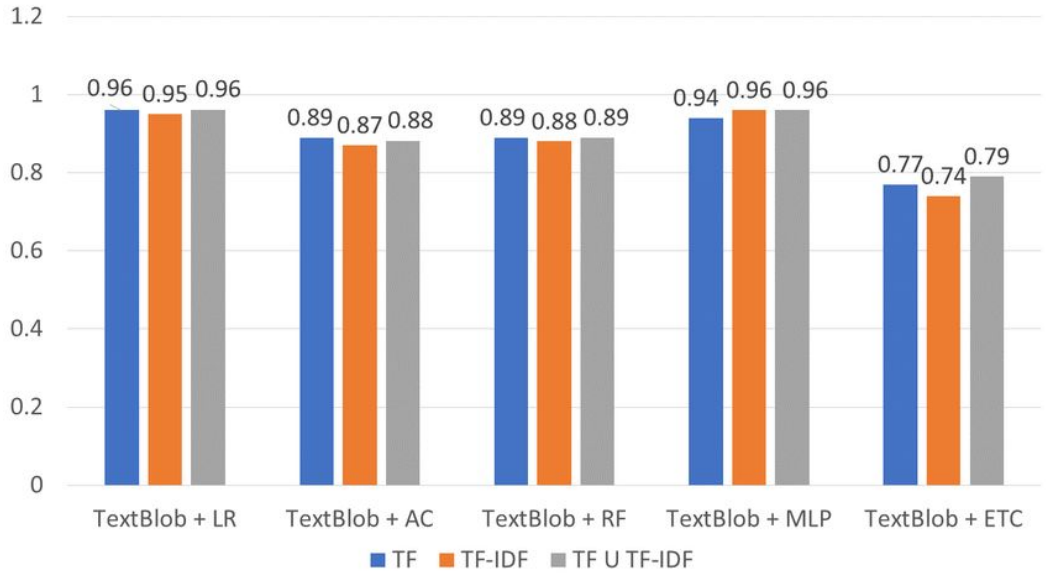
Implementation

- Implemented in PySpark (using Resilient Distributed Datasets)
- Preprocessing the tweets
- Give interval time
- Classify them into Negative, Positive and Neutral tweets (Negative tweets contain hateful words)

Evaluation

- Processes around 200 tweets per minute
- Some other sentiment analysis tools include Vader, Flair.
- Vader and Textblob work same in case of negative sentiments.
- Different factors contribute: capitalization, punctuation, emojis

TEXTBLOB	Precision	Recall	F1 score
Negative	0.67	0.48	0.56
Positive	0.25	0.59	0.35
Neutral	0.26	0.15	0.19



Reference:
https://www.researchgate.net/publication/352393003_Determining_the_Efficiency_of_Drugs_Under_Special_Conditions_From_Users'_Reviews_on_Healthcare_Web_Forums

Challenges faced

```
print("status code", response.status_code)
print("text", response.text)

if "data" in response:
    print("in if")
    ids = list(map(lambda rule: rule["id"], rules["data"]))
    payload = {"delete": {"ids": ids}}

response = requests.post(
        "https://api.twitter.com/2/tweets/search/stream/rules",
        auth=bearer_oauth,
        json=payload
    )
```

```
{"title": "Unauthorized", "type": "about:blank", "status": 401, "detail": "Unauthorized"}
status code 401
text {
  "title": "Unauthorized",
  "type": "about:blank",
  "status": 401,
  "detail": "Unauthorized"
}
```

401	Unauthorized	V1.1 V2	There was a problem authenticating your request. This could be due to missing or incorrect authentication credentials. This may also be returned in other undefined circumstances.

Elevated access

Conclusion and Future Scope

Thus, we have created use case of hate speech detection of streaming tweets.

Future Scope: The NLP models show negative polarity when tweets contain some hateful words. But sometimes the sentiment of tweet is positive and it does not fall under offensive category. We can work on this problem.

Project category

Excellent Category