Sarcasm Detection

Saba Suhail

Introduction

- Sarcasm refers to use of words that mean opposite of what one wants to say

- To insult/irritate/being funny.The limits of sarcasm are not so well defined

- Sarcasm is subjective. Non-native speakers/readers may not get it.

- **Goal:** To predict whether a headline is sarcastic or not

- **Metric:** Accuracy Score

- **Overall Approach:** Clean->Feature Generation and Selection-> Modeling

Dataset

Kaggle-

https://www.kaggle.com/datasets/rmisra/news-headlines-dataset-for-sarcasm-detection

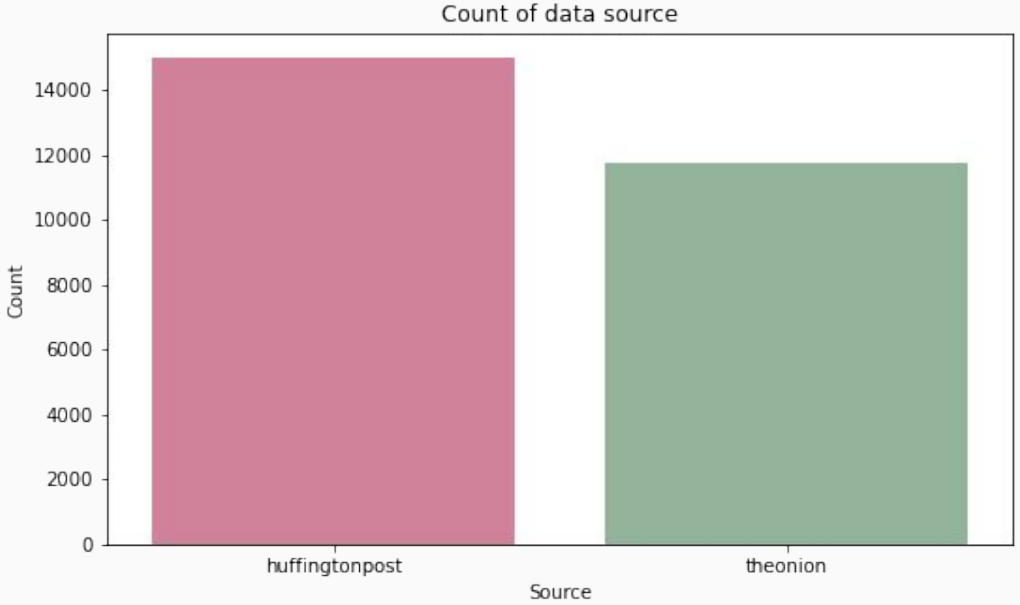
Each record consists of three attributes:

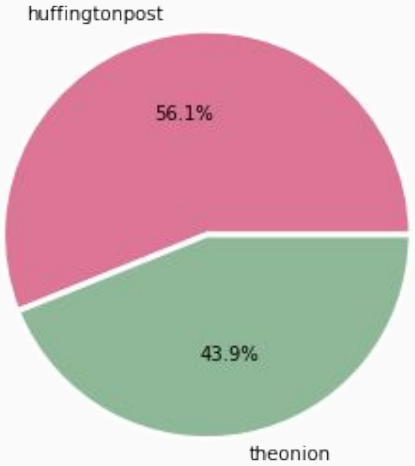
- `**is_sarcastic**`: 1 if the record is sarcastic otherwise 0

- `**headline**`: the headline of the news article

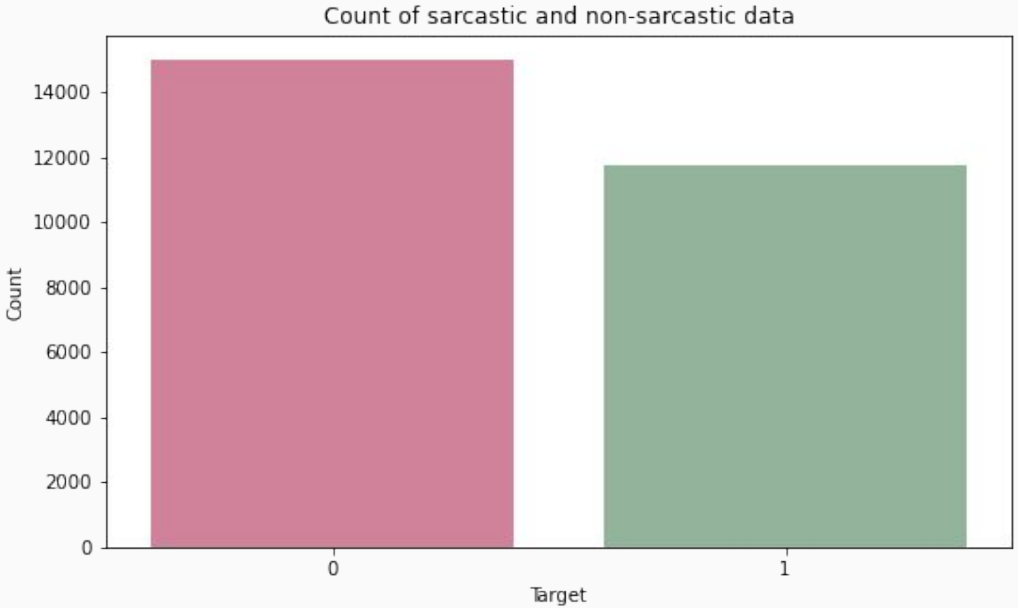
- `**article_link**`: link to the original news article. Useful in collecting supplementary data

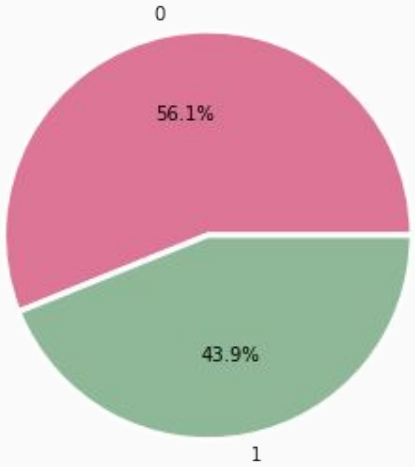
Source


Count of data source
14000
12000
10000
8000
6000
4000
2000
0
Count
huffingtonpost
theonion
Source

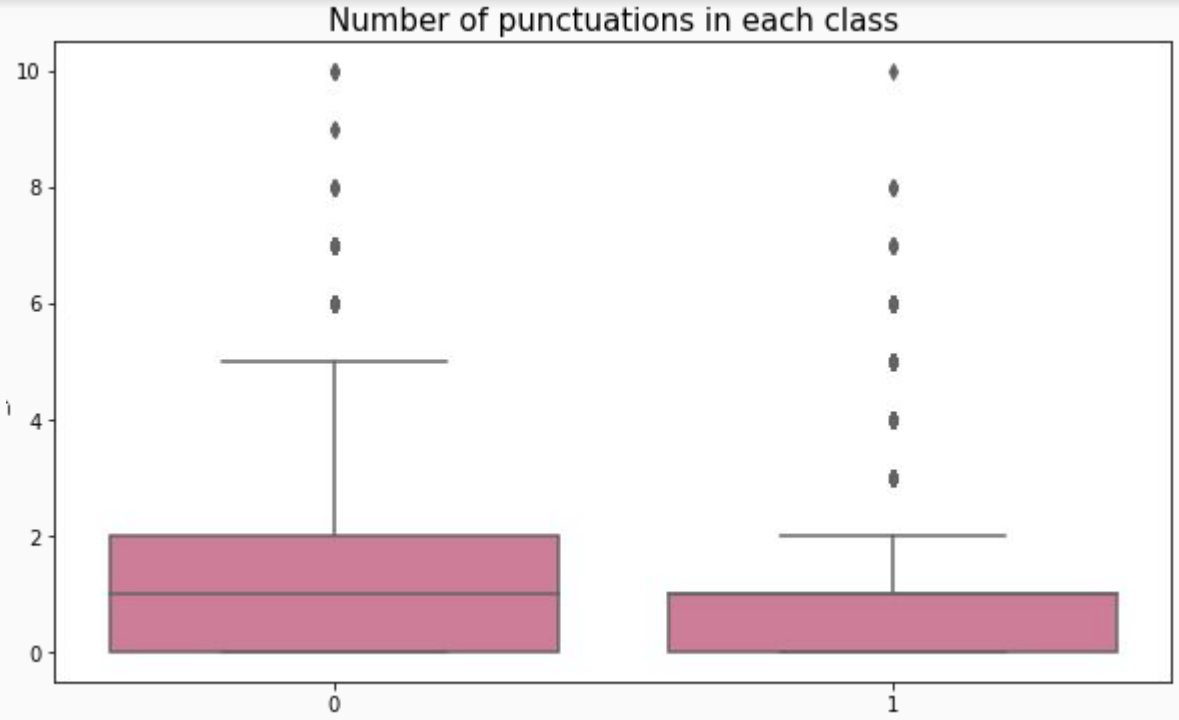

huffingtonpost
56.1%
43.9%
theonion

Target Categories

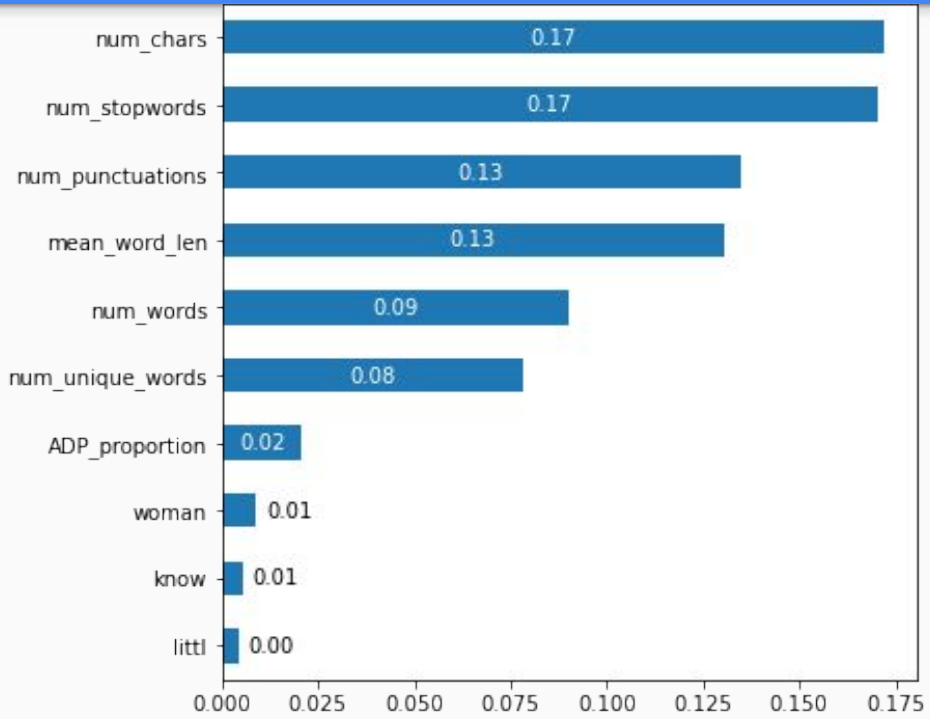




Punctuations in each class

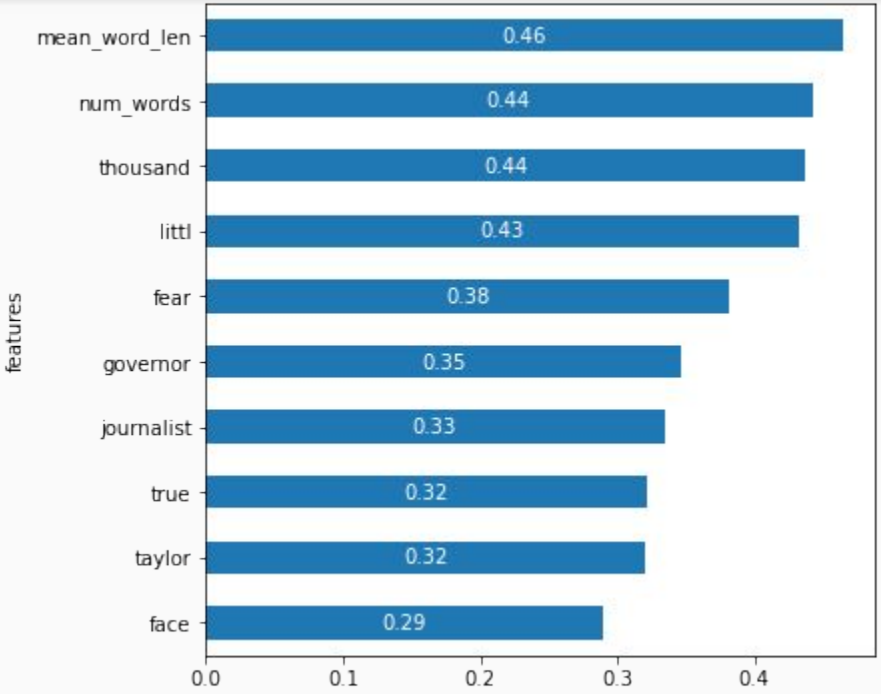


Random Forest Classifier



Accuracy: 65%

Logistic regression



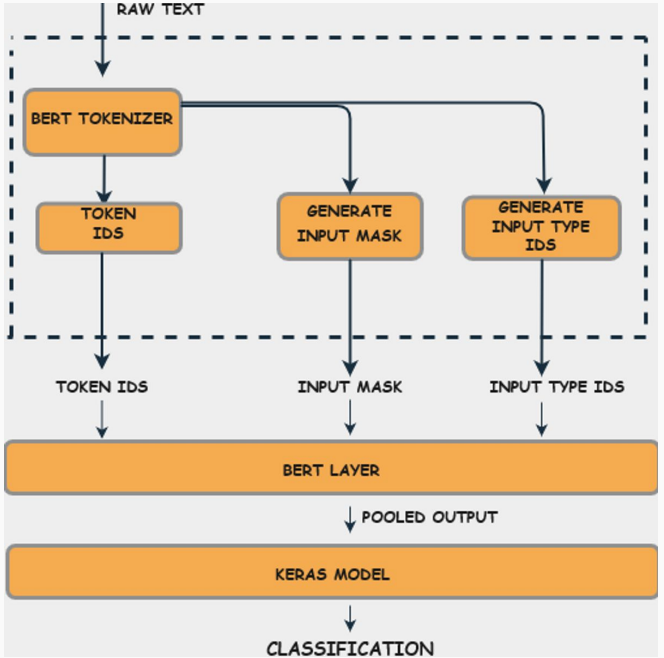
Accuracy: 63%

BERT

- Reads the entire sequence of words at once

- This characteristic allows the model to learn the context of a word based on all of its surroundings (left and right of the word).

- Accuracy:



Conclusion

Conclusion : Context matters for sarcasm and hence BERT seems apt for text classification and performed best among the models in this project.

Limitations: only applicable for English language

Next Steps: Explore more contextual neural nets and techniques

Impact: Chatbots can be more intelligent, Adding emotions into voice detection systems