

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.

Dataset

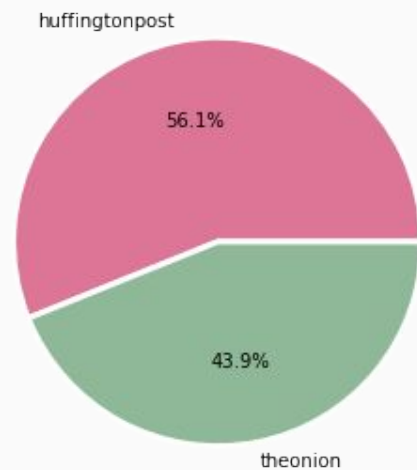
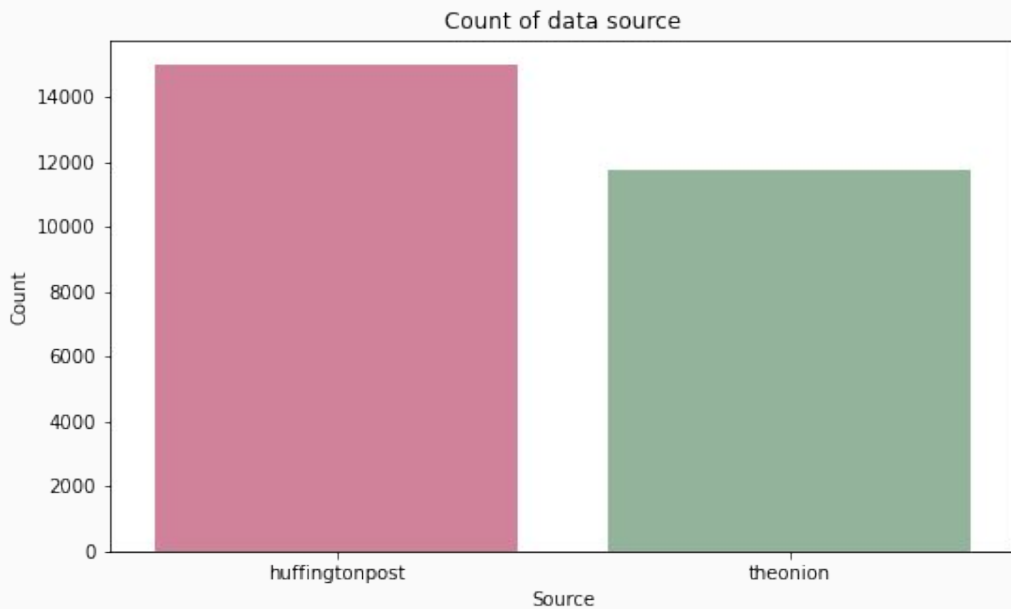
Kaggle-

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

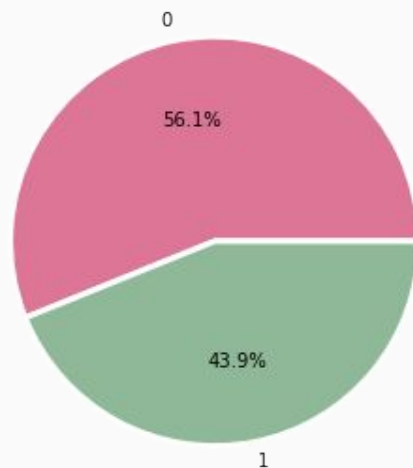
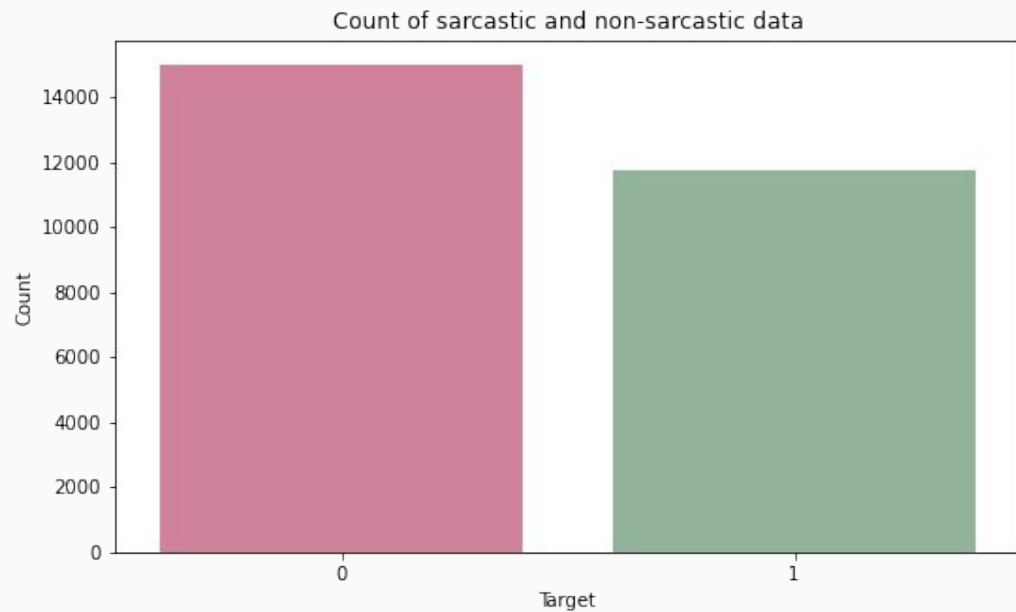
Each record consists of three attributes:

- **`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

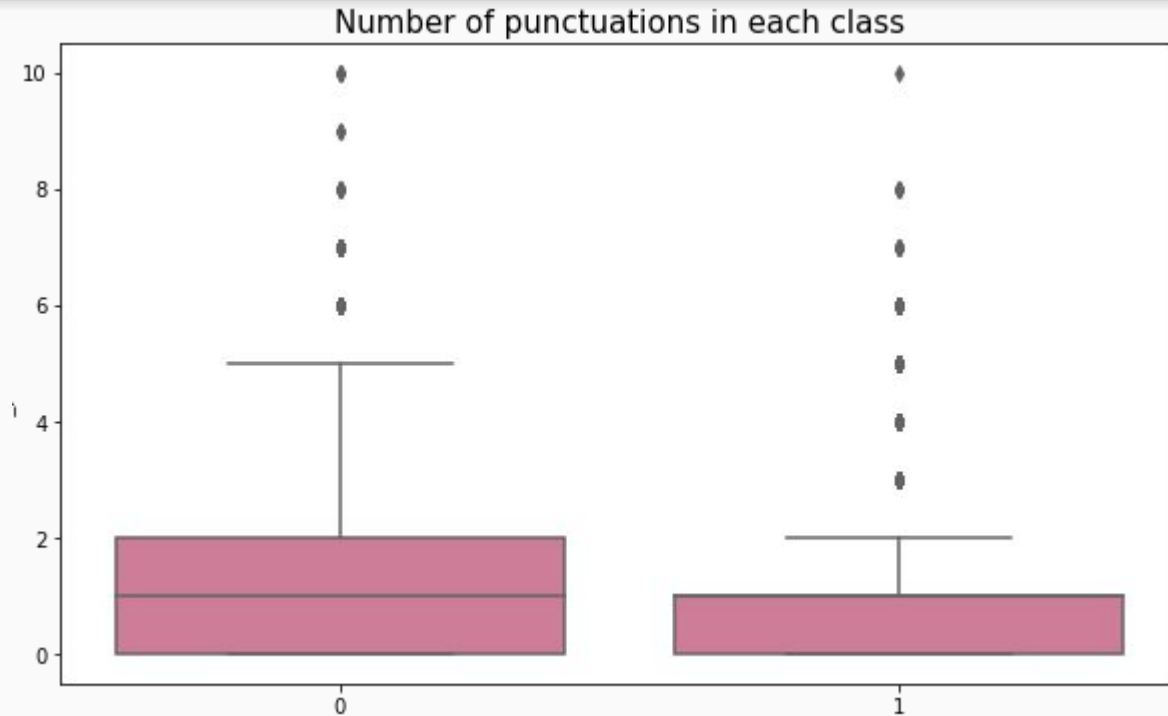
Source



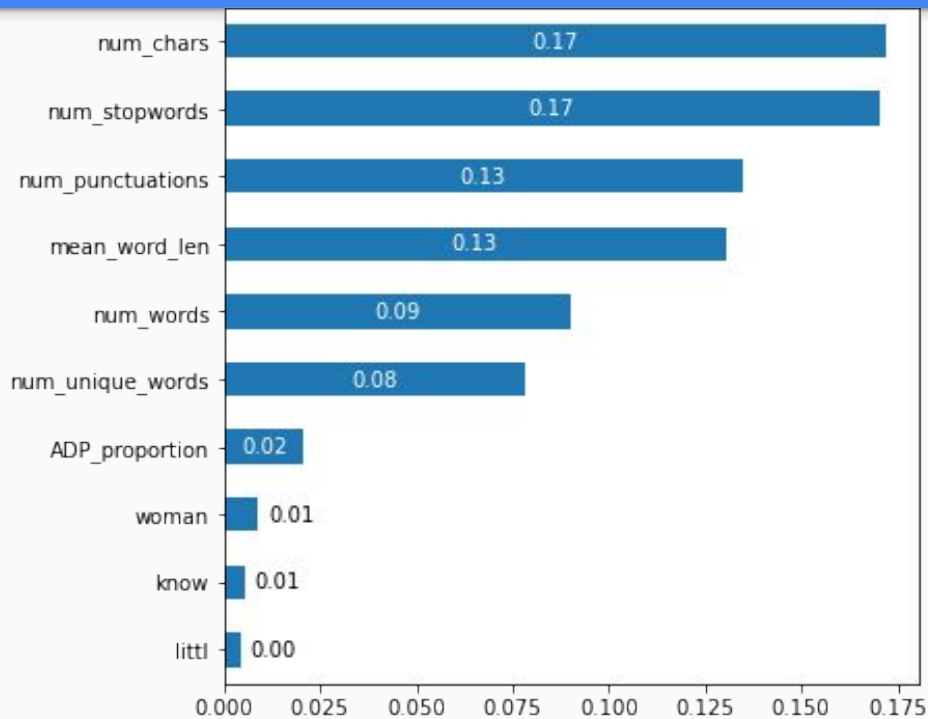
Target Categories



Punctuations in each class

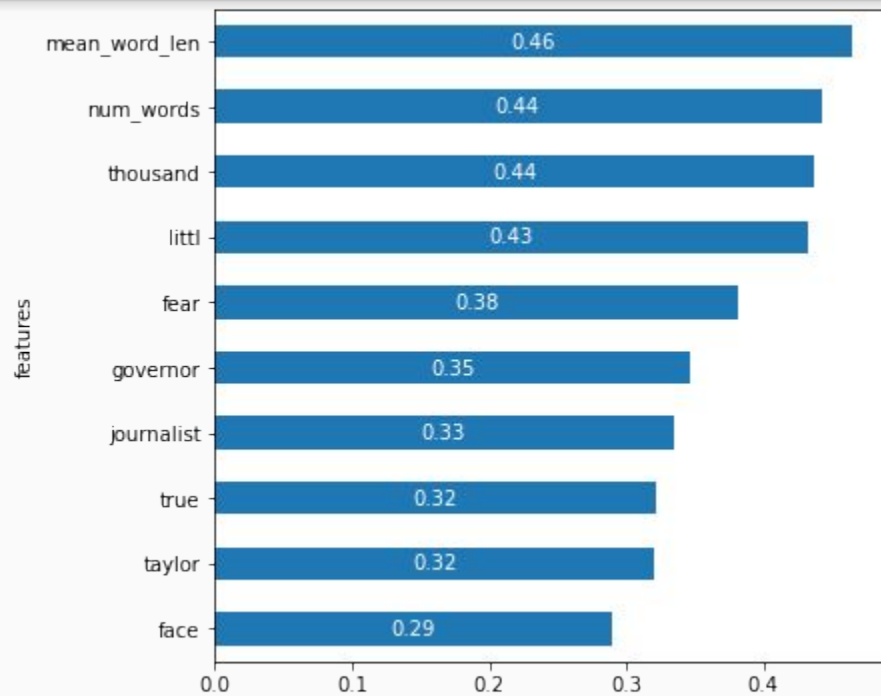


Random Forest Classifier



Accuracy: 65%

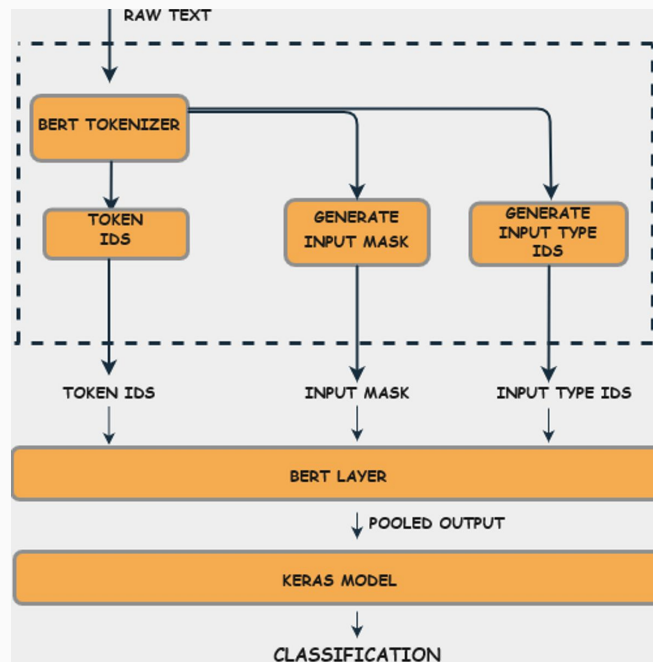
Logistic regression



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

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