

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 text is sarcastic or not

Dataset

Kaggle-

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

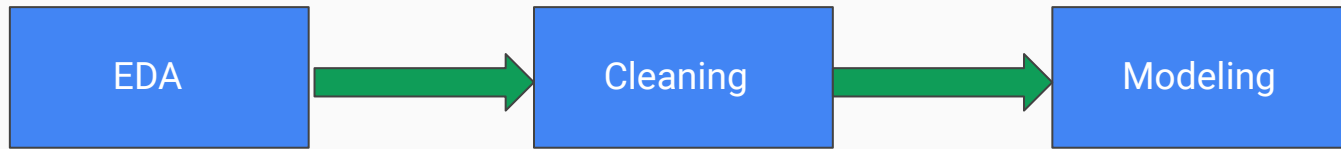
Number of Entries: 26709

Success Metric: Accuracy Score, Precision

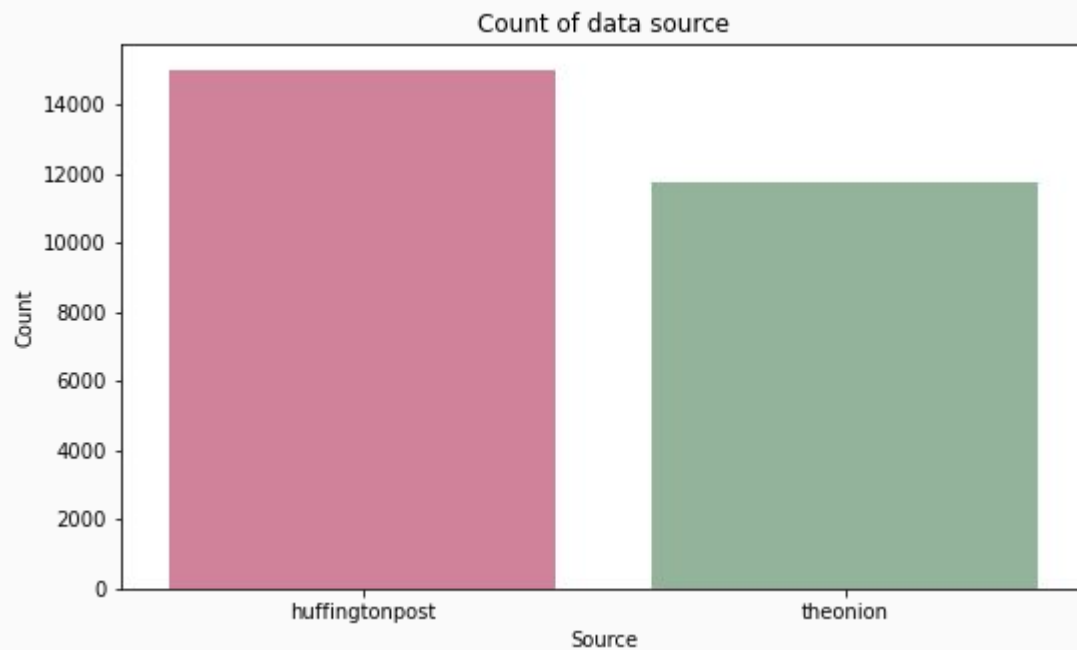
Each record consists of three attributes:

- **`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

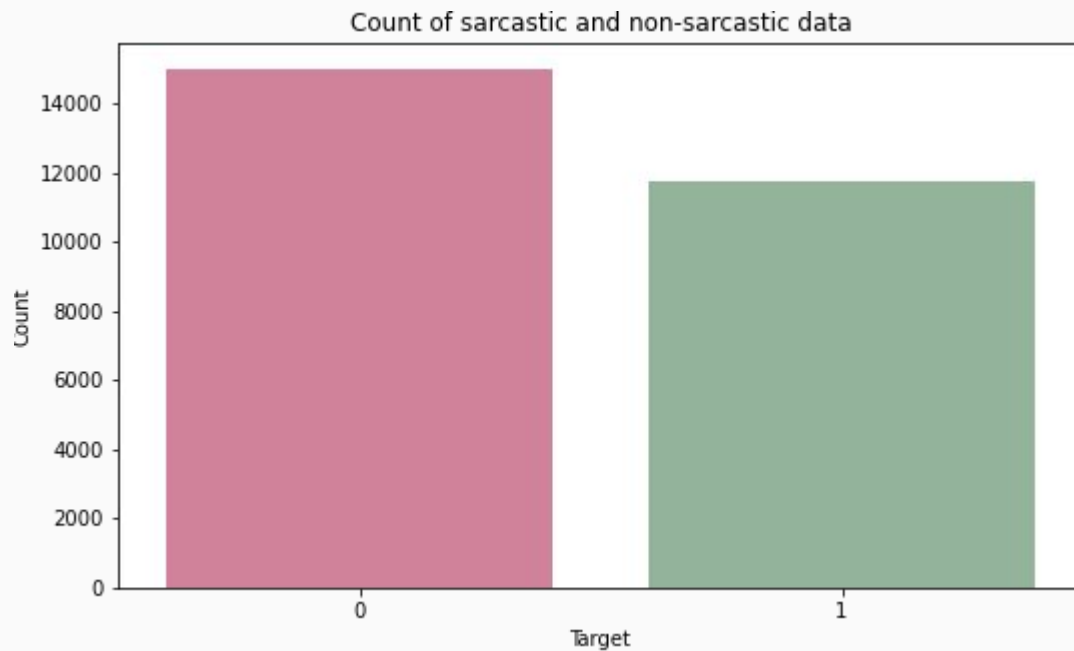
Overview of Project



Source



Target Categories

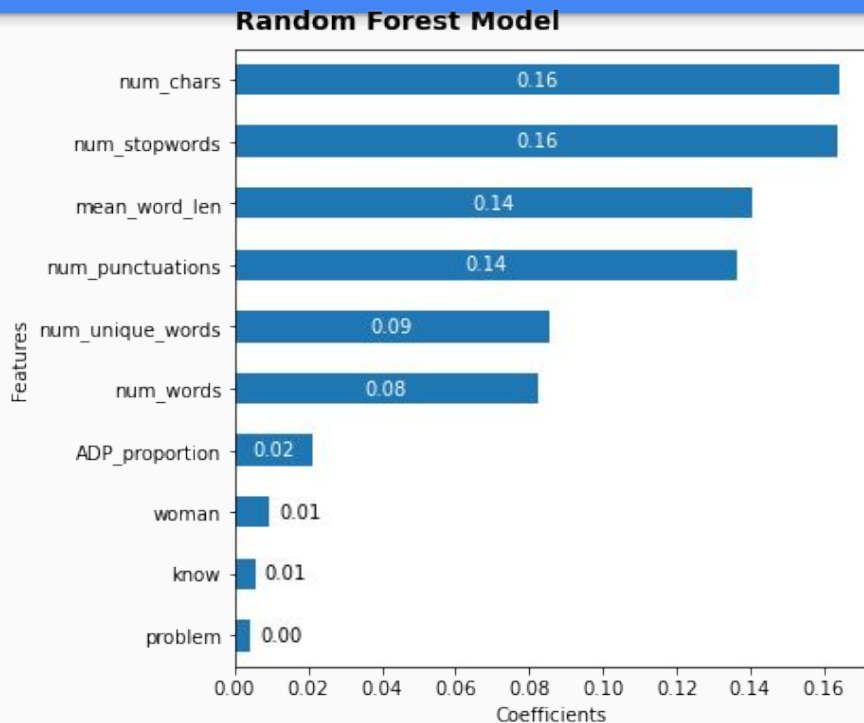


Baseline:56%

Punctuations in each class

- Number of Punctuations in non-sarcastic headlines is noticeably more than that in sarcastic headlines which means it is the wordplay which creates sarcasm.
- Contrary to informal written language probably which uses more punctuation to create sarcasm.

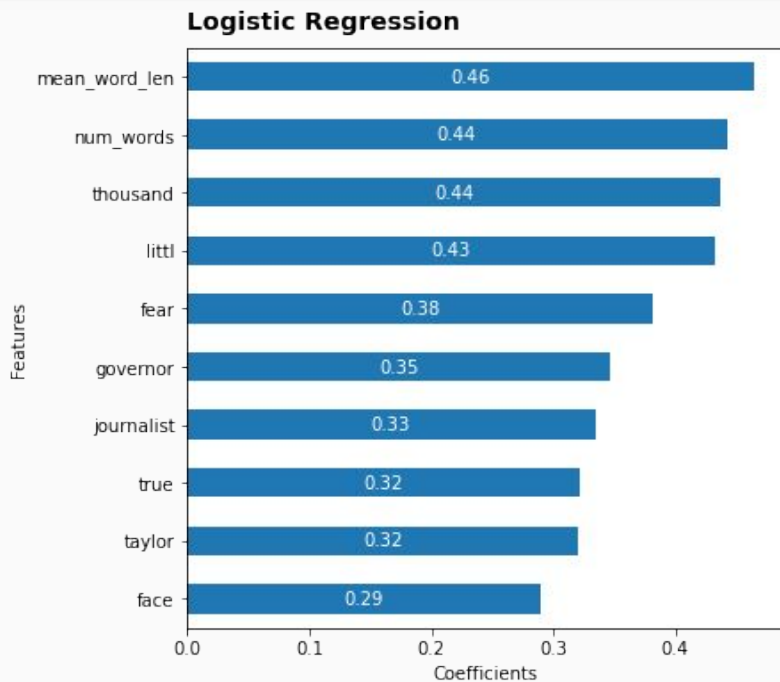
Random Forest Classifier



Accuracy: 65%

Precision: 0->66% 1->64%

Logistic Regression

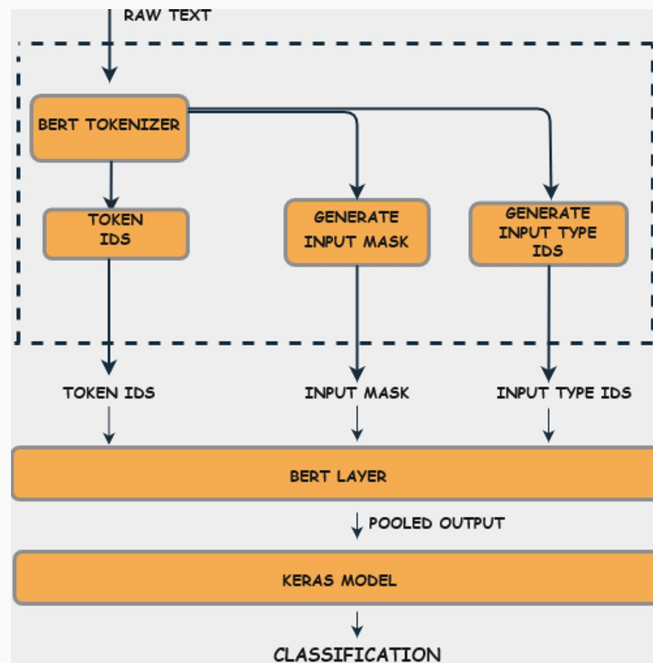


Accuracy: 63%

Precision: 0->65% 1->58%

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: 80%



Conclusion

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

Recommendation: BERT should be the production model.

Limitations: Only applicable for English language

Next Steps: Explore more contextual neural nets and techniques

Applications

- This analysis can be used to design intelligent chatbots
- Can help in adding emotions into voice detection systems
- E-commerce websites can use such it to have a better understanding of product/service reviews posted by users