

140XP FINAL PROJECT

Vanda Suklar, Laura Ngo, Ryan Largo, Suraj Rajan, Odeya Russo, Masato Ishizuka

INTRODUCTION

In the evolving landscape of social media discourse, the intersection between political communication and platform moderation policies has increased in importance. Using a dataset consisting of web-scraped tweets from Donald Trump between 2009 and 2021, our study aims to predict through a sentiment analysis which tweets are more likely to be flagged by Twitter’s moderation system.

PROBLEM STATEMENT

Unmasking Incentives: Flag Analysis of Trump’s Twitter Communication We will be using Sentiment analysis to conduct a detailed investigation of Donald Trump’s Twitter that compares his communication strategies and controversies prior to and during his candidacy.

FEATURE SELECTION

Should probably include a section that lists the features used to build the models.

METHODS

Firstly, the dataset was initially filtered to exclusively include Tweets posted by Trump himself, excluding any retweets. Secondly, the Tweets were cleaned by removing stop words and unnecessary text. Following this, stemming was applied to reduce each word to its base or root form, allowing for simplification in the language representation. To address the class imbalance in the flagged versus not flagged categories, a random oversampling technique was used to balance the class distribution of the dataset. To further optimize the dataset, Principal Component Analysis (PCA) was used for dimension reduction, effectively reducing the number of features to be used in the predictive models. The data was then split into training and testing sets. Lastly, a variety of models were trained and evaluated, including Naive Bayes Classification, Logistic Regression, Random Forest (assessed through Feature Importance Plot, Confusion Matrix, ROC and AUC Curve, and Precision-Recall Curve), and K-Nearest Neighbors. Testing accuracies were calculated to

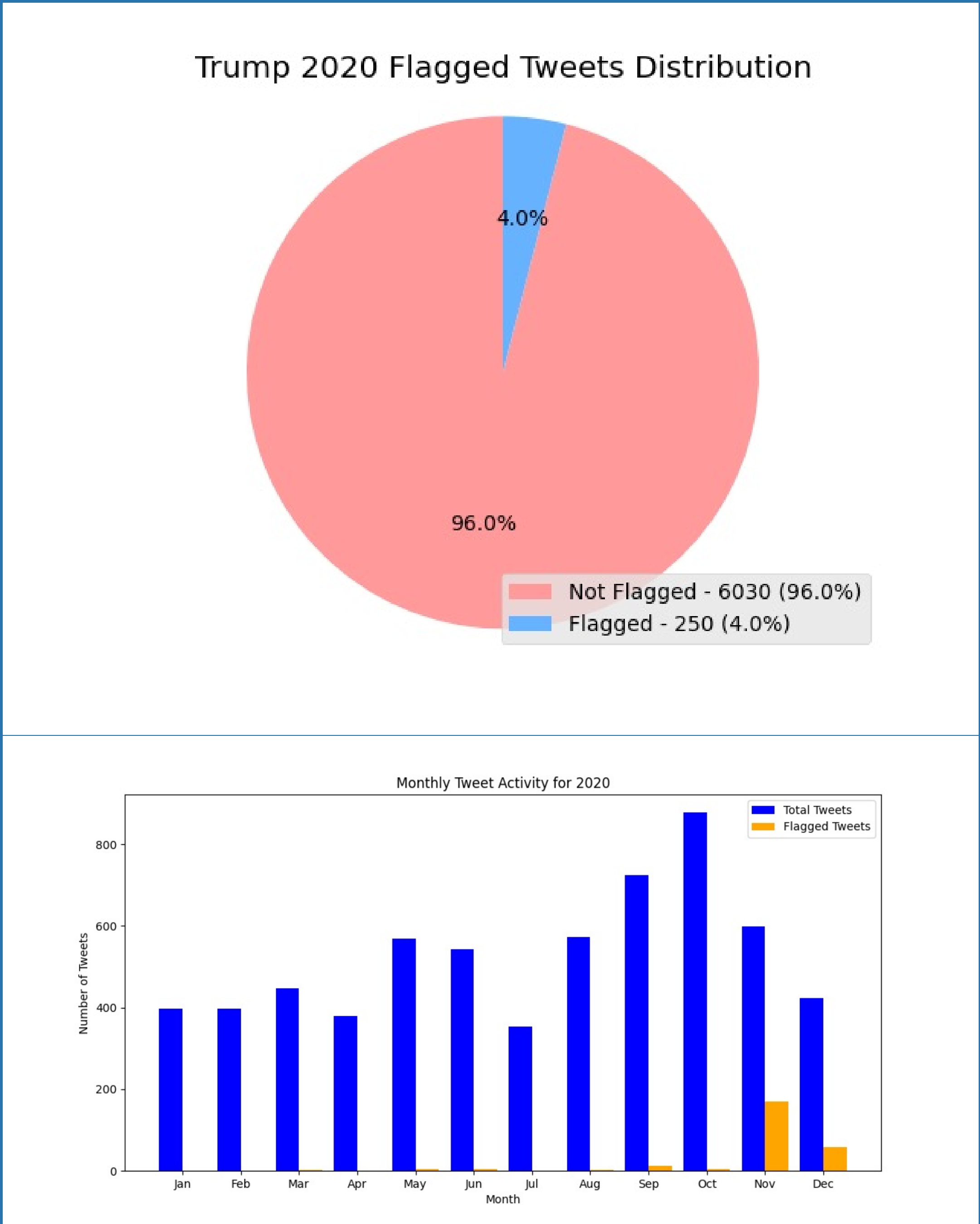
An Analytical Study: Predictive

Modeling for the Classification of

Flagged Twitter Posts: Insights

from Trump’s Tweets Dataset

(2009-2021)



RESULTS

- Naive Bayes
- K-Nearest Neighbors
- Random Forest
- Logistic Regression

images/tyre.png

CONCLUSION

REFERENCES AND APPENDICES

images/stats_140xp.JPG