

This project is a classification problem to predict tweets from the 45th president of the United States Donald J. Trump and the prime minister of Canada Justin Trudeau. Formally, we want to classify a text using natural language processing and machine learning to predict who tweeted. The project has one clean dataset that's collected via the Twitter API. The first method that will be used is to vectorize the tweets using the CountVectorizer and TFIDF vectorizer. Then, a naive Bayes model was trained with both the CountVectorizer and the TFIDF vectorizer to see which one will perform better. The TFIDF model had an accuracy that is better than the count-based approach. Next, I have tried another classifier to compare different machine learning algorithms. By applying the linearSVC model, I have found that its accuracy is much higher than the naive Bayes model. The main result we got from the project is that Trudeau seems to be tweeting more in French than Donald Trump does. These results can be used for fake news detection classifiers in the future.