

# Improving DraftKings targeted marketing

NATHAN PERRY

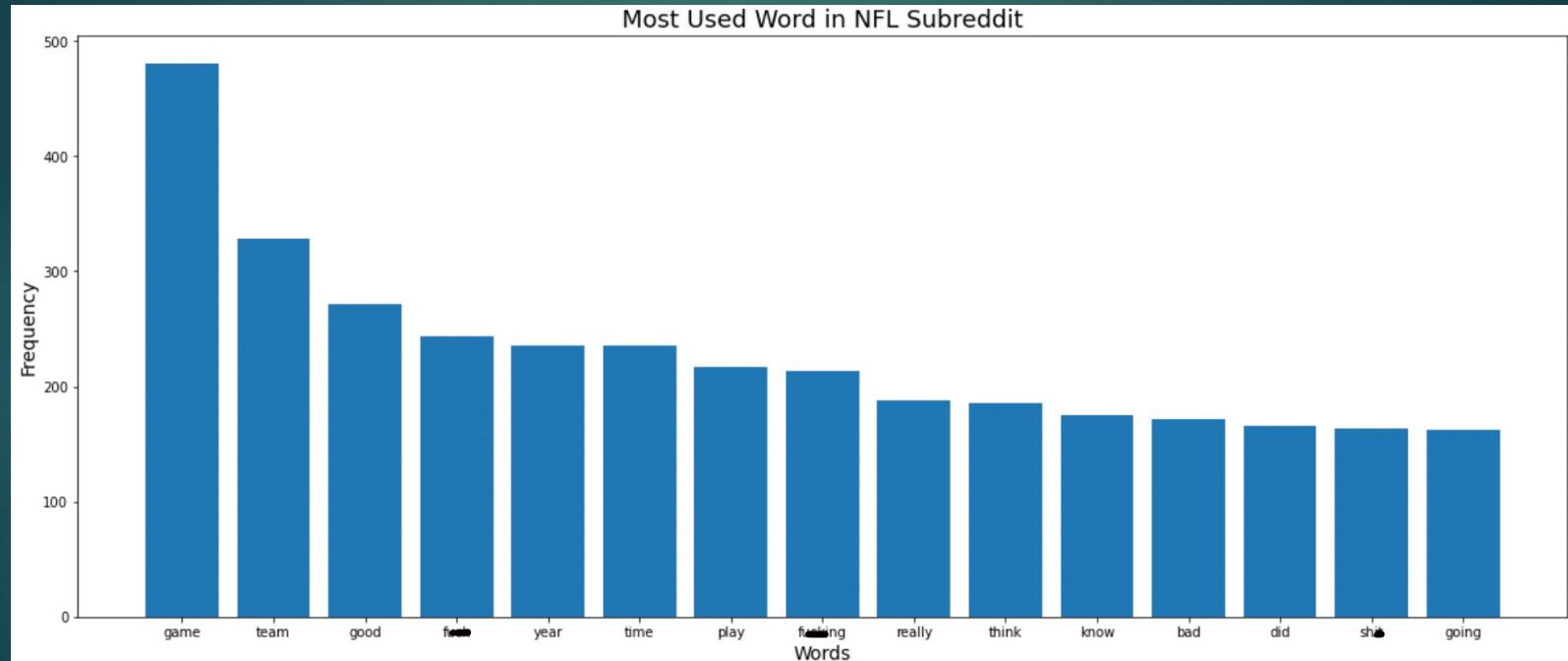
# Problem Statement

- ▶ DraftKings marketing is looking to improve their targeted advertising on social media
- ▶ My goal is to create a model that can predict if a football fan is interested in fantasy football based on the words they use
- ▶ The two metrics I care about are accuracy and recall
- ▶ Recall is important because we want to reach as many potential customers as possible
- ▶ Accuracy is important because although our main concern is reaching as many customers as possible, false positives are a waste of advertising money

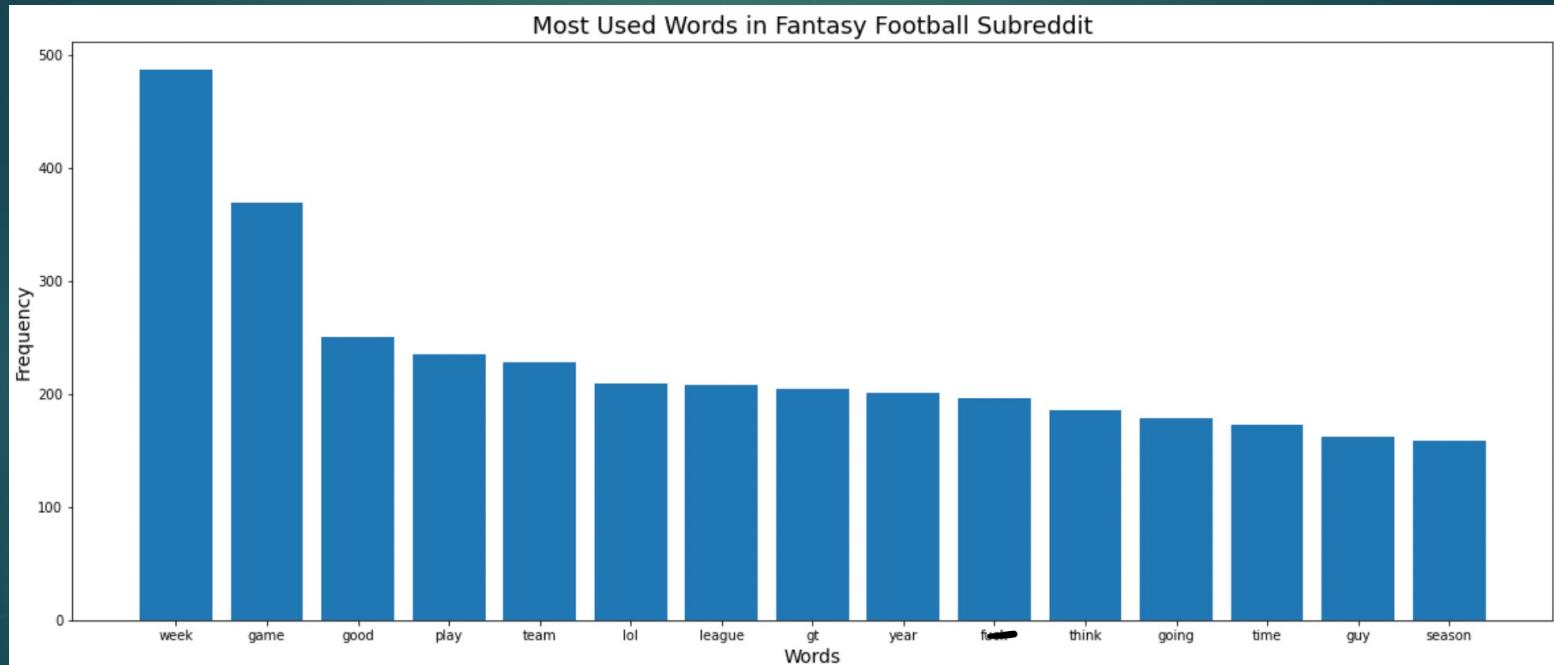
# Data

- ▶ The data was acquired by scraping the nfl and fantasy football subreddits using the pushshift api
- ▶ The top 100 comments in each subreddit per day were scraped for October and November 2020
- ▶ The data I scraped for each comment was, subreddit, time posted, score and body

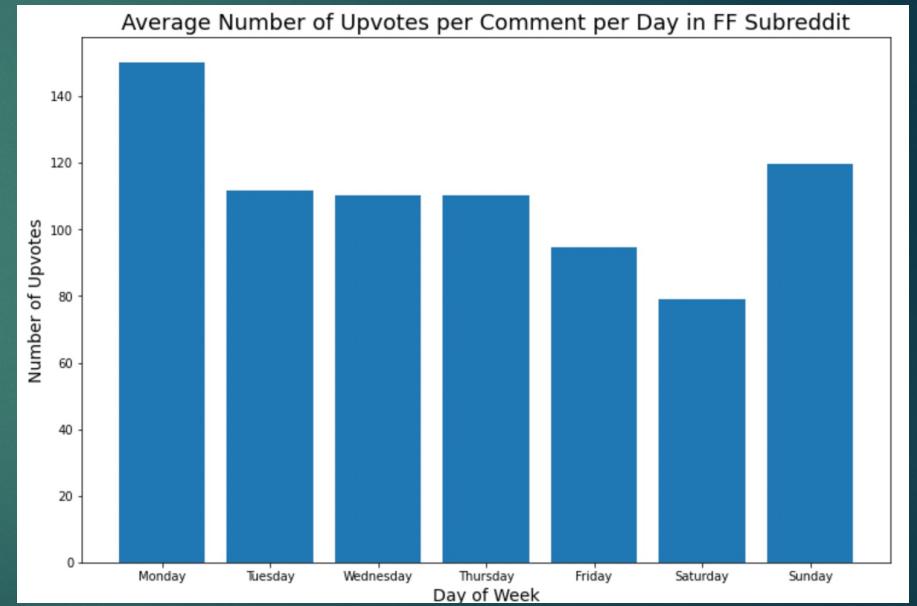
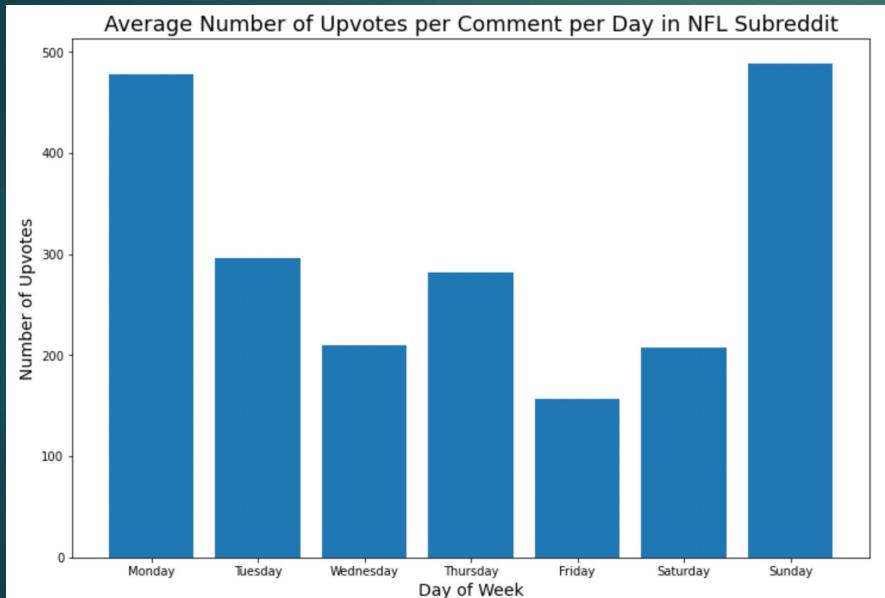
# Most popular words in NFL Subreddit



# Most popular words in Fantasy Football Subreddit



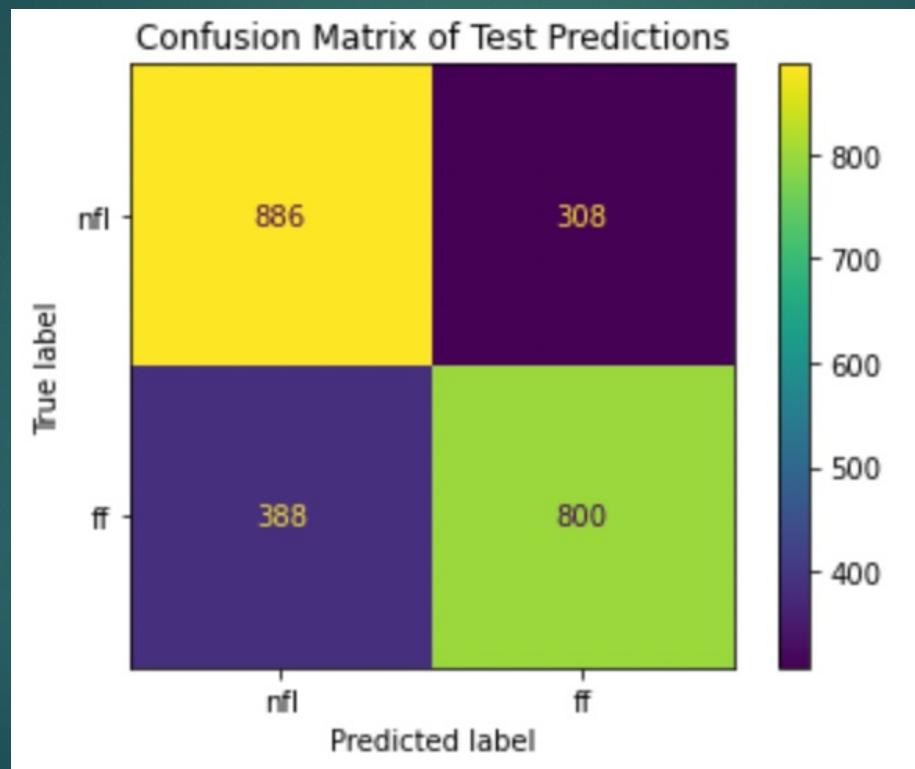
# What days of the week get the most traffic



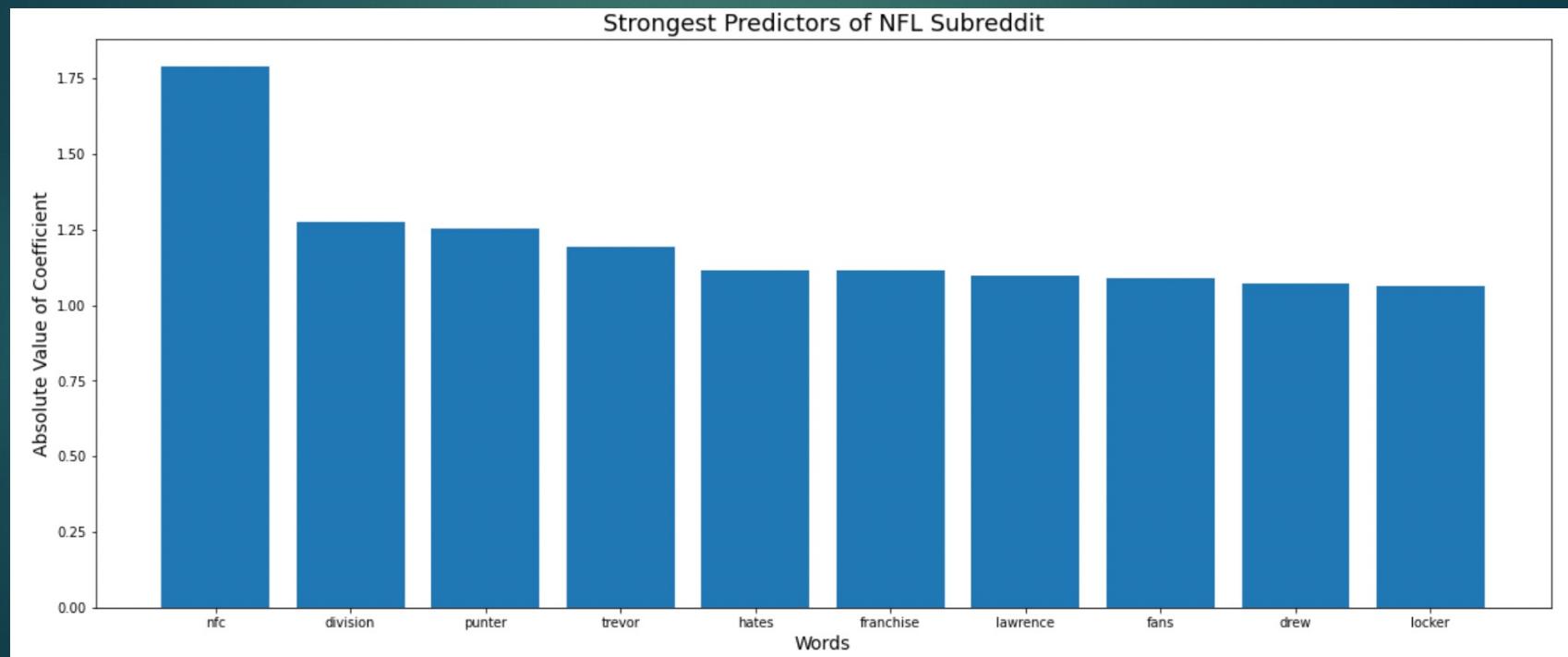
# Models used

- ▶ I created 3 models, one using KNN Classification, one using Logistic Regression, and one using Random Forest Classification
- ▶ The baseline model predicts 'nfl' for each comment and had an accuracy of 50.1%
- ▶ The KNN Model has an accuracy of 56% and a Recall of 79%
- ▶ The Random Forest Model has an accuracy of 72% and a recall of 63%
- ▶ The Logistic Regression Model has an accuracy of 71% and a recall of 67%
- ▶ I selected the Logistic Regression model because it has a better recall than Random Forest and the KNN accuracy is too low.

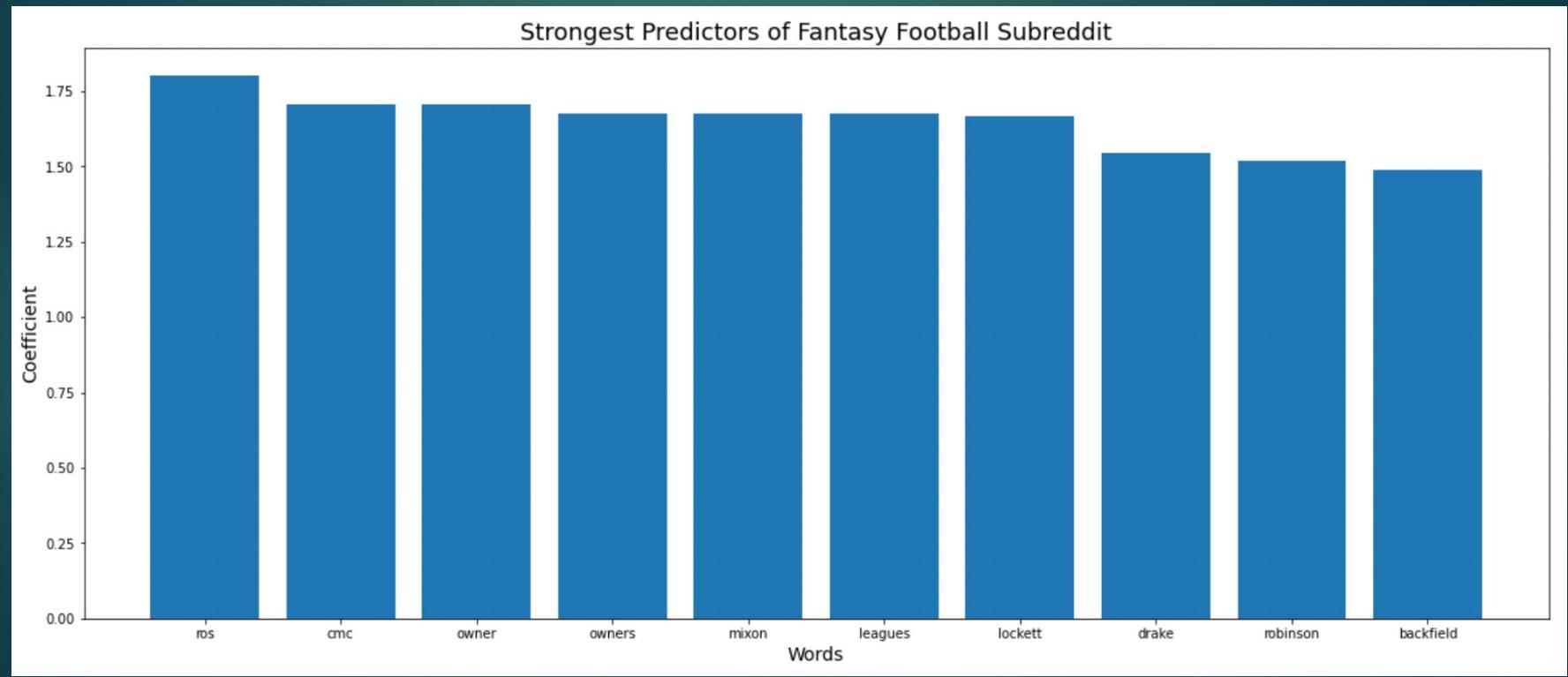
# Evaluation of model



# Which words are the strongest predictors for nfl commenters



# Which words are the strongest predictors for the fantasy football subreddit



# Conclusions

- ▶ General football fans are less active on reddit than fantasy football fans on days games are not being played
- ▶ While both the Logistic Regression and Random Forest models perform well, I recommend the Logistic Regression
- ▶ Fantasy football fans talk about individual players with strong statistical performances at a higher rate
- ▶ Football fans talk more about college prospects and special teams players than fantasy football players
- ▶ There are some terms that fantasy football players are more likely to use than general football fans when evaluating players.