

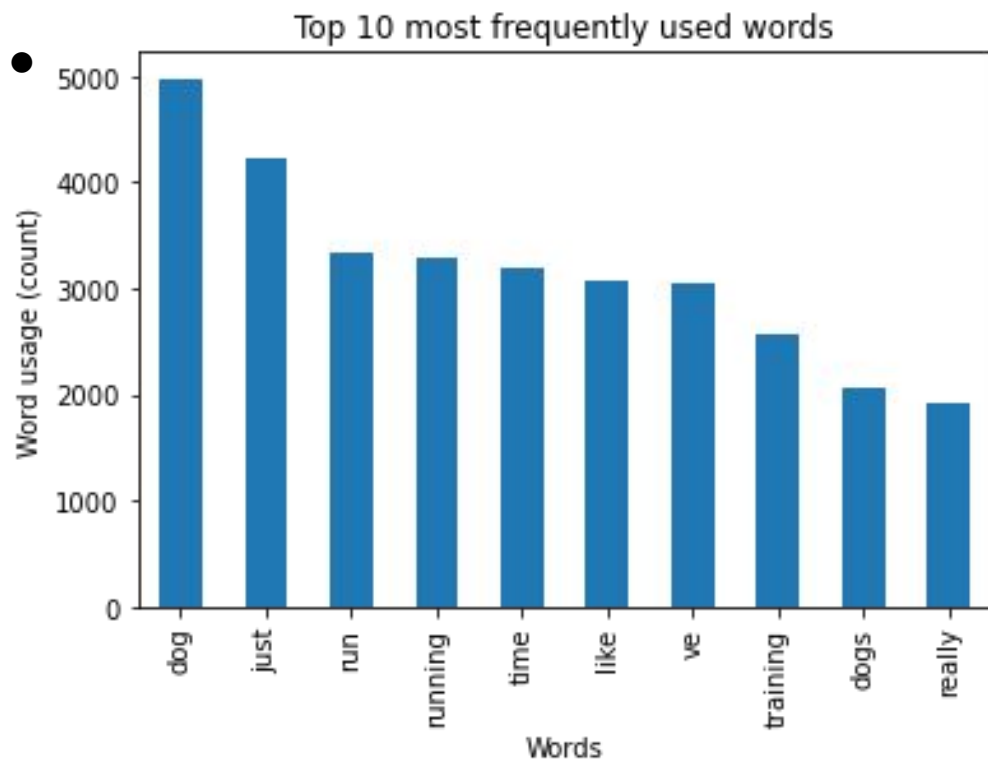


Subreddit Identification

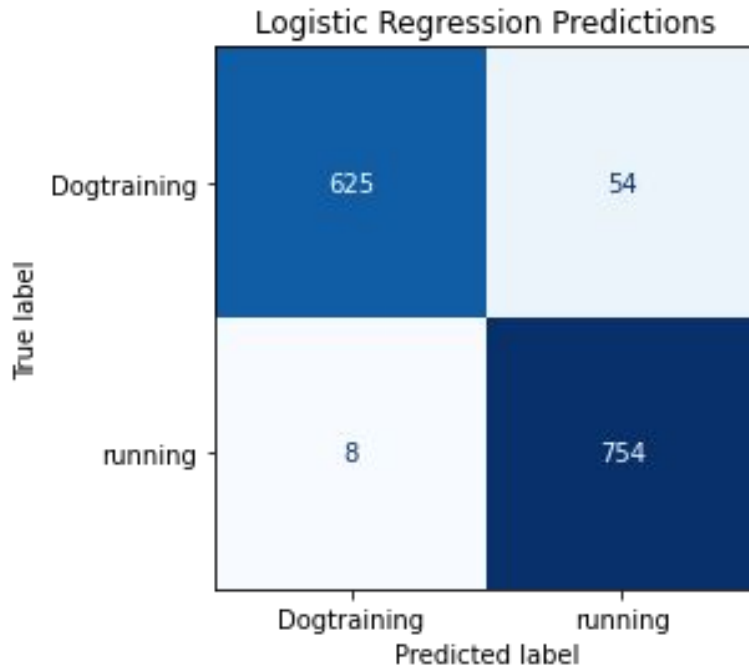
Dog Training vs. Running



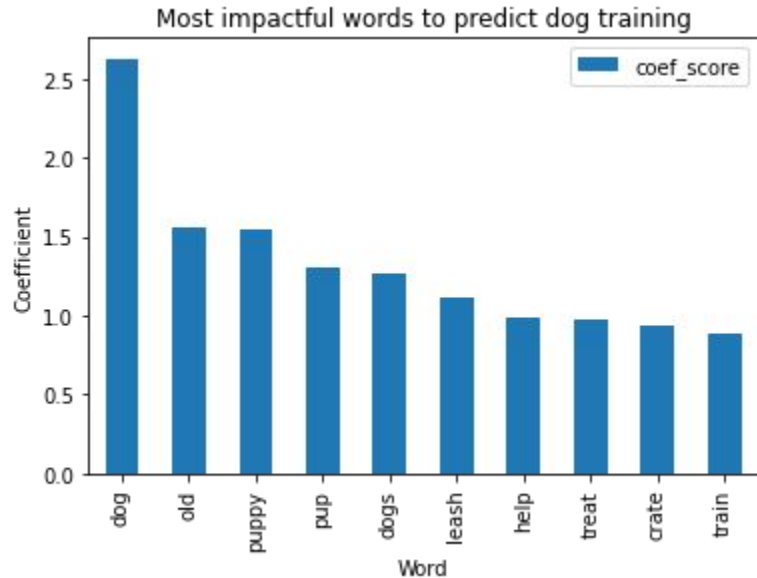
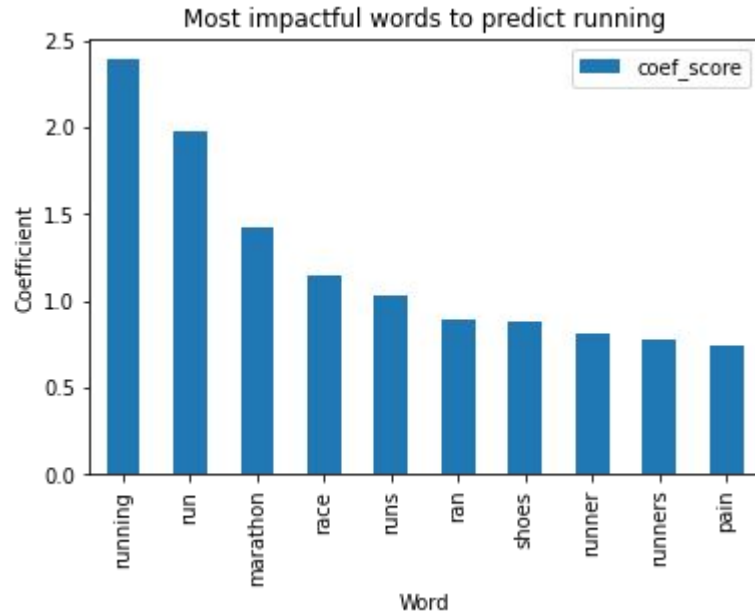
Subreddit Selection



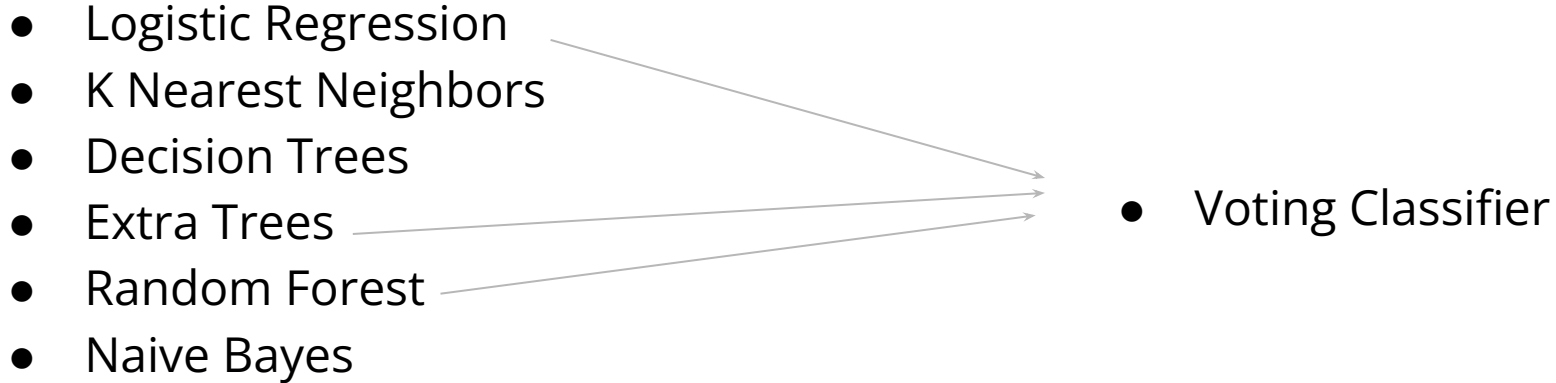
Model Performance



Removing common words



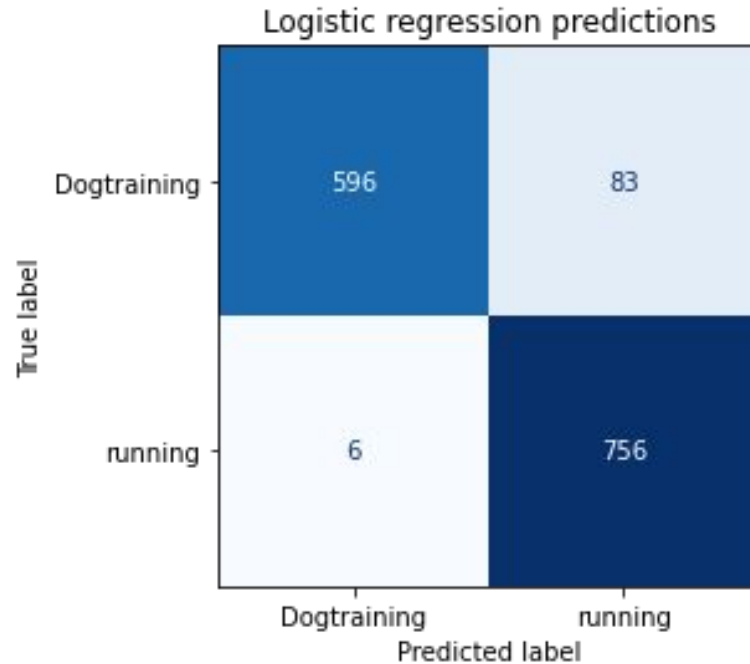
Exploring Models



Tuning the model

- Grid Search on Countvectorizer & Logistic Regression
 - Ngrams
 - Min_df
 - C strength
- Grid Search on Voting Classifier
 - Optimal Ngrams and Min_df from above
 - C strength
 - Max depth
 - N estimators

Evaluating Performance



Next Steps

- Try it out on more data
- Deep dive into the misclassified posts