

ESPN

X



reddit

How well can we Predict the Subreddit?

Author : Matthew Hill

Table of Contents

- 01 Information Gathering
- 02 NBA vs NFL Top 10 phrases
- 03 Model Comparison
- 04 Accuracy Comparison
- 05 Recommendation
- 06 Conclusion



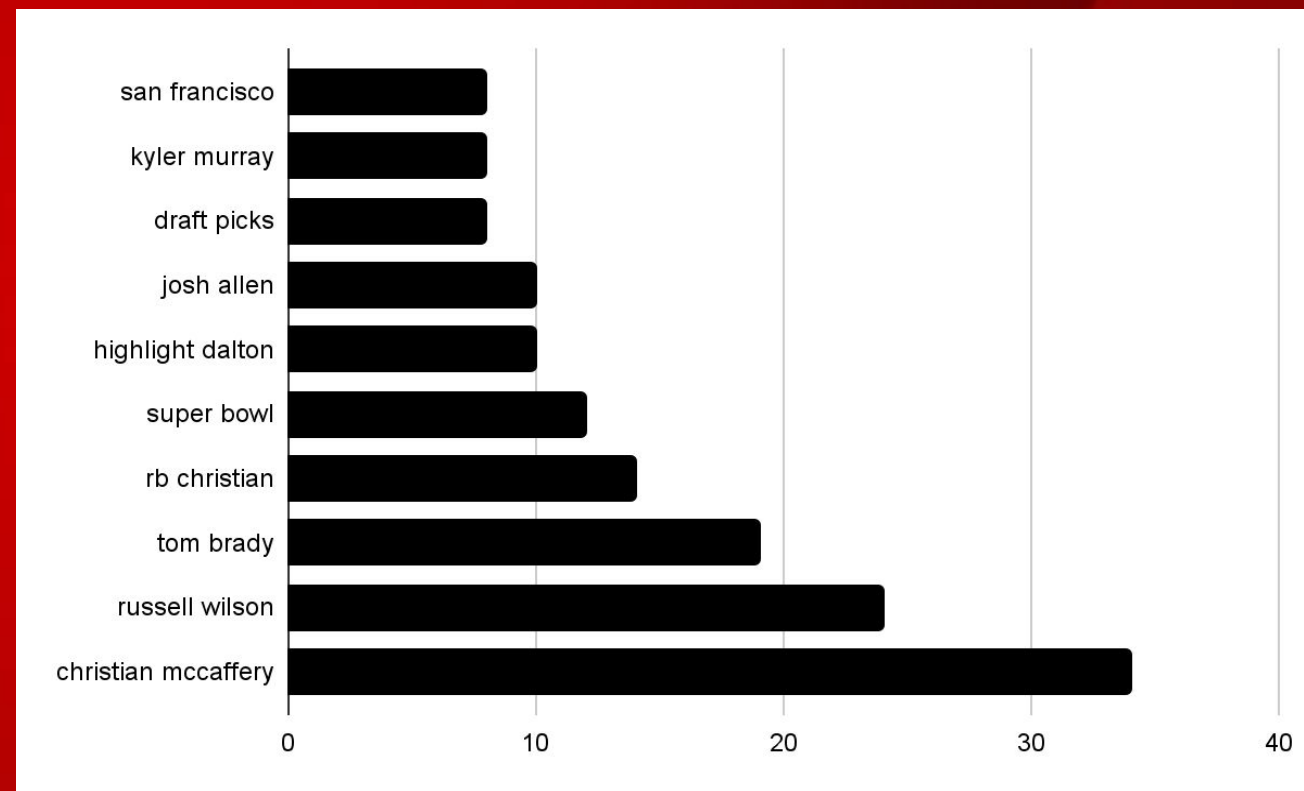
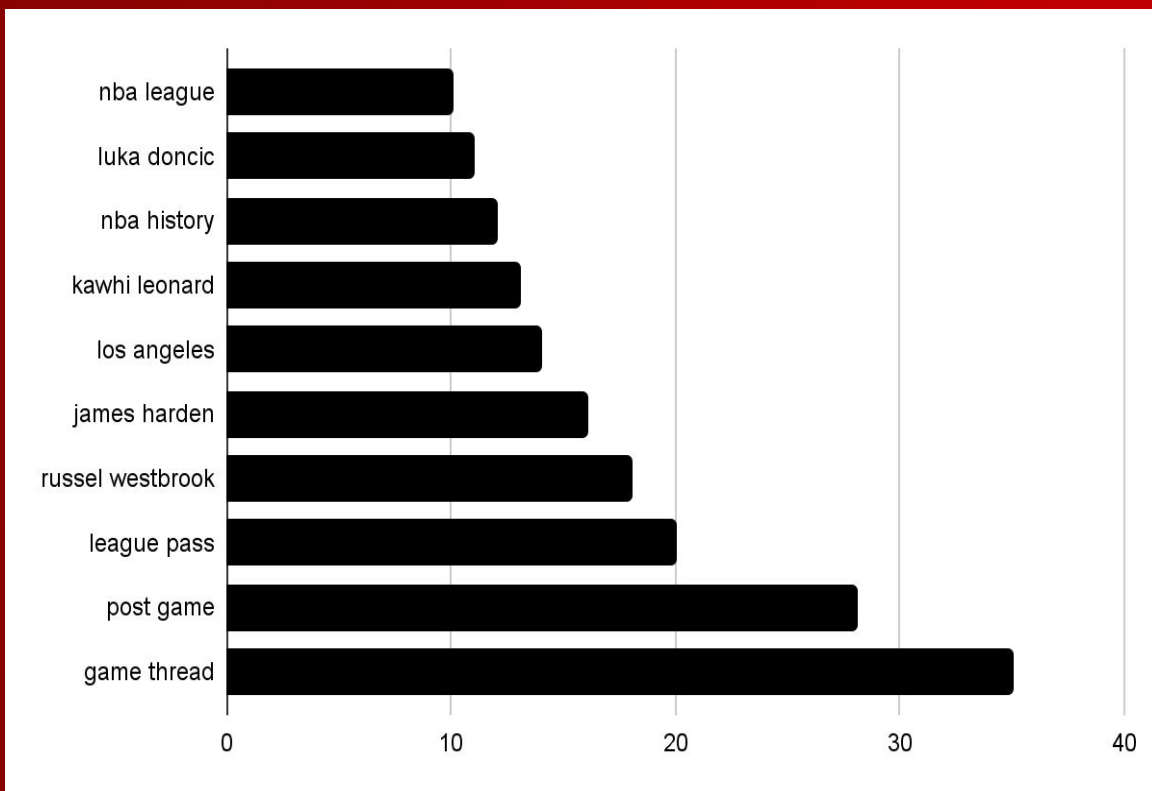
Information Gathering

- NBA & NFL Subreddit
- Pushshift's API
- JSON
- Parameters





Top 10 Phrases



Model Comparison



Naive Bayes Pros

- Classification
- Simple implementation
- Fast training & small training data

Cons

- Not suitable for numerical data
- Assumption of independent features



Random Forest Pros

- Classification & Regression
- Decision trees
- Not over fit

Cons

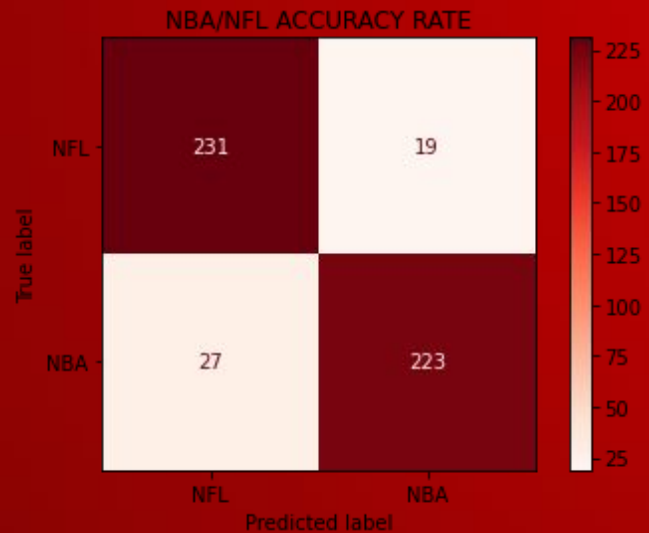
- Interpretability
- Slow training



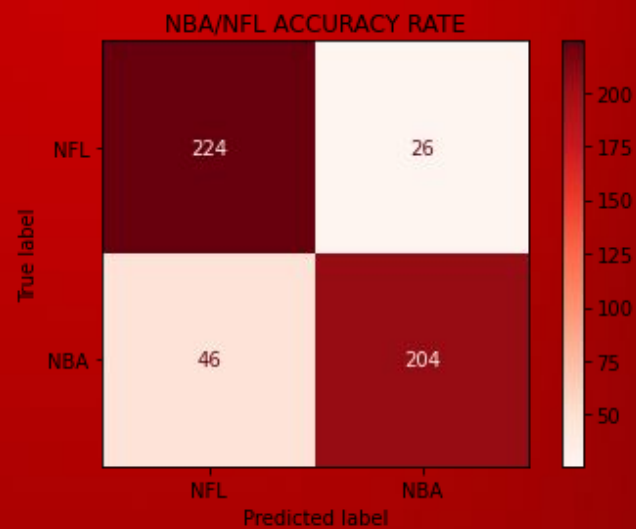
Accuracy Comparison



- Accuracy Rate
90%



- Accuracy Rate
86%



Recommendation

NAIVE BAYES



Thank you!

Any Question?

