

# Market Research: Sports Fans in the U.S. and England

Riley Robertson



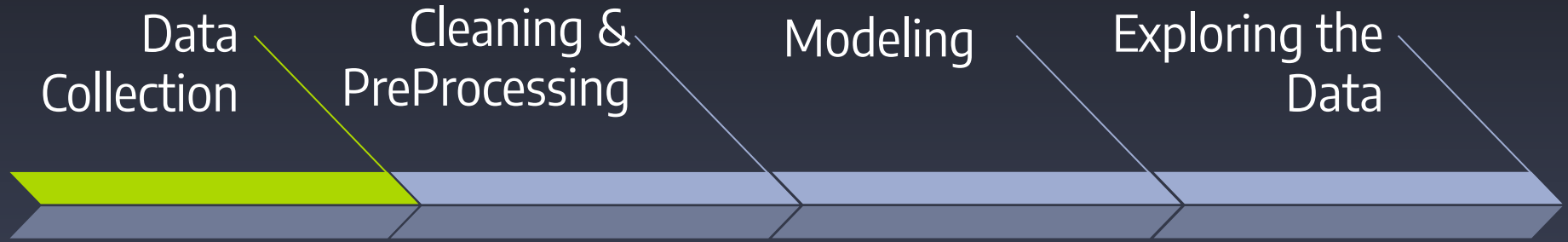
Arguably the most popular sports in each of their respective countries, the National Football League (NFL) and the English Premier League (EPL) boast some of the largest fan-bases in North America and Europe, making them prime targets for advertising.



# OverArmor's Requests:

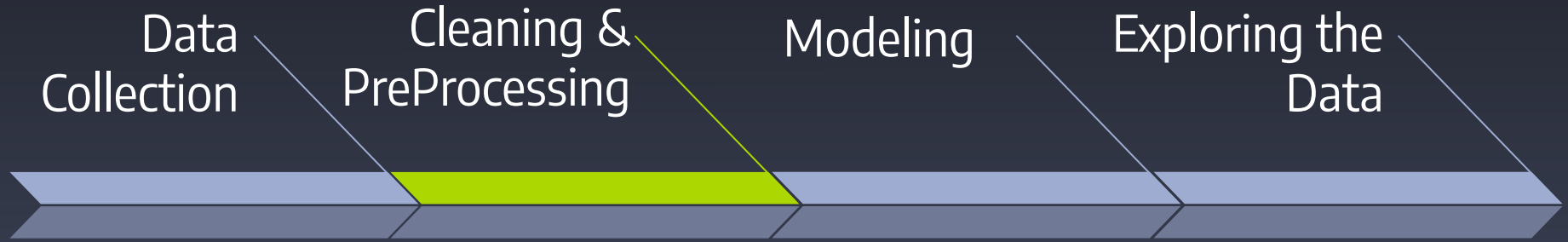
1. Collect data from each league's respective Reddit community to build a clean database they can use to study the language used by each fan-base
2. Create a machine learning model that can quickly identify the source of a post if the source is unknown.
3. Provide insights into the vernacular of each fanbase so OverArmor can more effectively reach potential customers

# Workflow



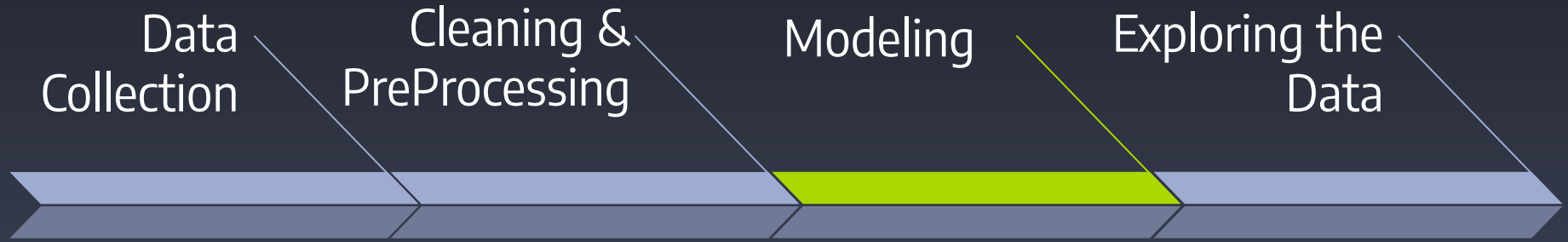
- ▷ 100k Posts Per Subreddit
- ▷ Lots of Cleaning Required

# Workflow



- ▷ Decreased to 13k total posts with a 56% / 44% split
- ▷ Lots of Cleaning Required

# Workflow

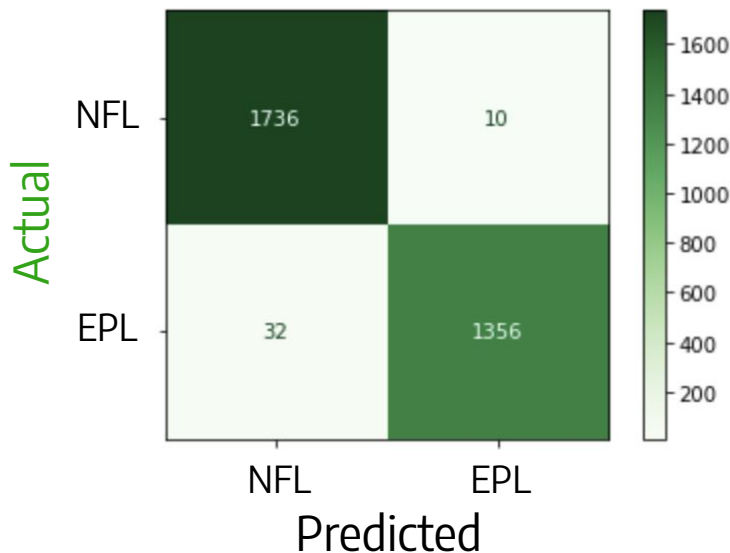


- ▷ Logistic Regression and Random Forest both returned excellent results

# Model Comparison

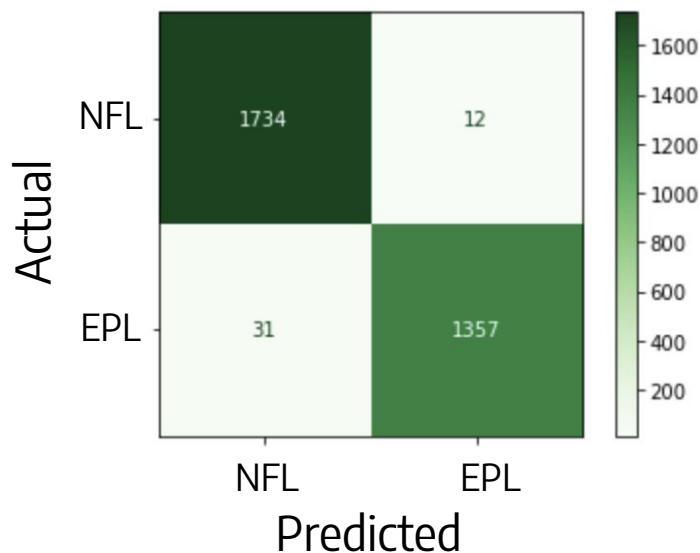
## Logistic Regression

98.6% Accuracy

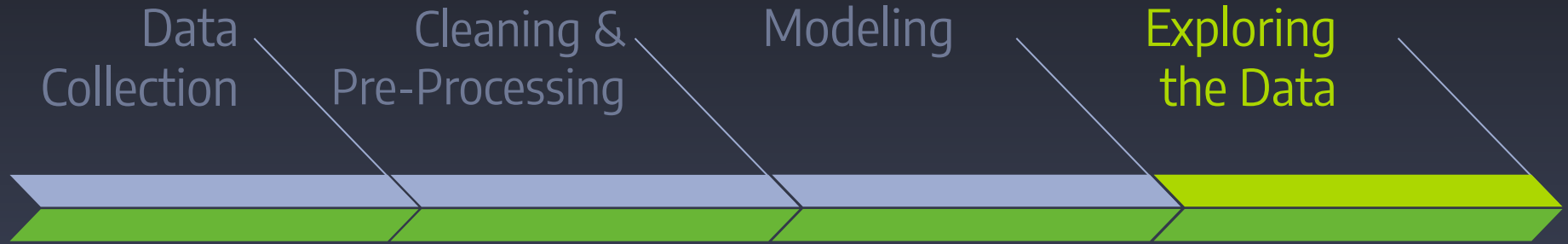


## Random Forest

98.8% Accuracy



# Workflow



▷ EDA



# EDA - Misclassified posts

title

subtitle

title

subtitle

# EDA - Token Prominence Comparison

r/nfl

subtitle

r/PremierLeague

subtitle

# EDA - Sentiment Analysis?

r/nfl

subtitle

r/PremierLeague

subtitle

# EDA - Post Length by Subreddit?

r/nfl

subtitle

r/PremierLeague

subtitle

# Conclusions

- 1.
- 2.
- 3.
- 4.
- 5.