NFL Play Predictor

By:Jayson Villena

Problem

A defensive coordinator gave our company a task of predicting what a play would be given the conditions at a specific point of a game. His team provided us a dataset of plays in the NFL for the past three years.

Problem Statement

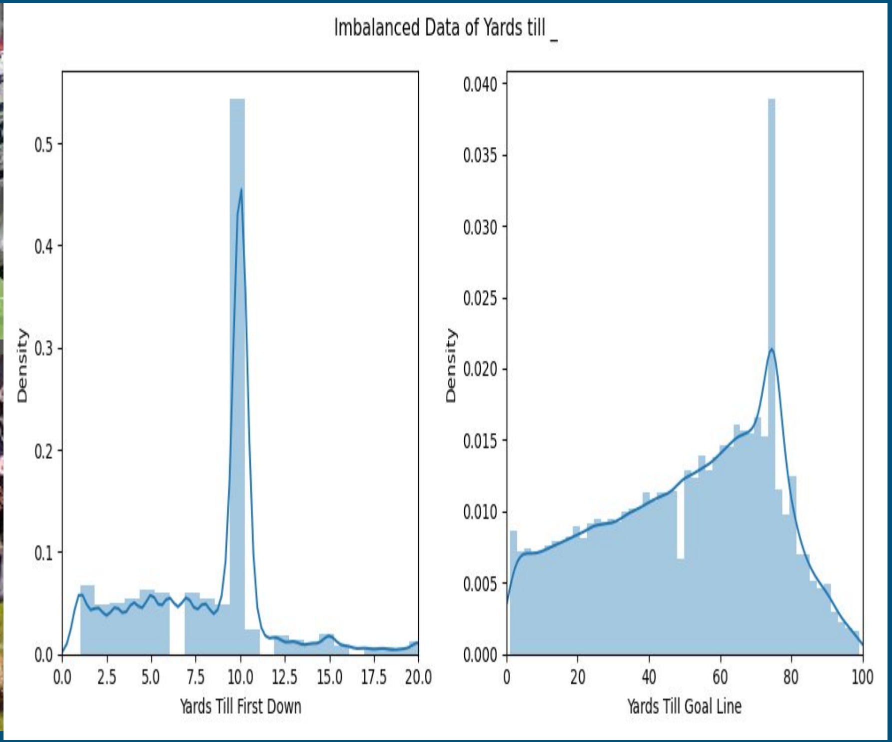
Create a predictive model that can predict the play type of an NFL play. Create an application that can be given new data and give a prediction of the play.

Methodologies

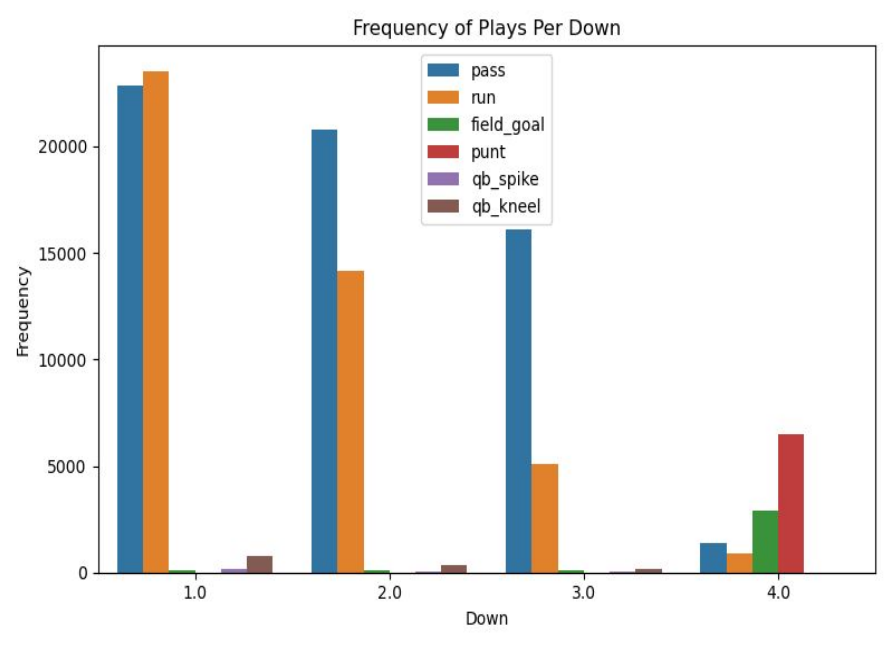
- Data Cleaning
- Processing
  - Feature Engineering (300+ columns)
  - Balanced or Imbalanced
  - PCA
- Model Selection
  - Neural Network or Decision tree?
- Model Evaluations

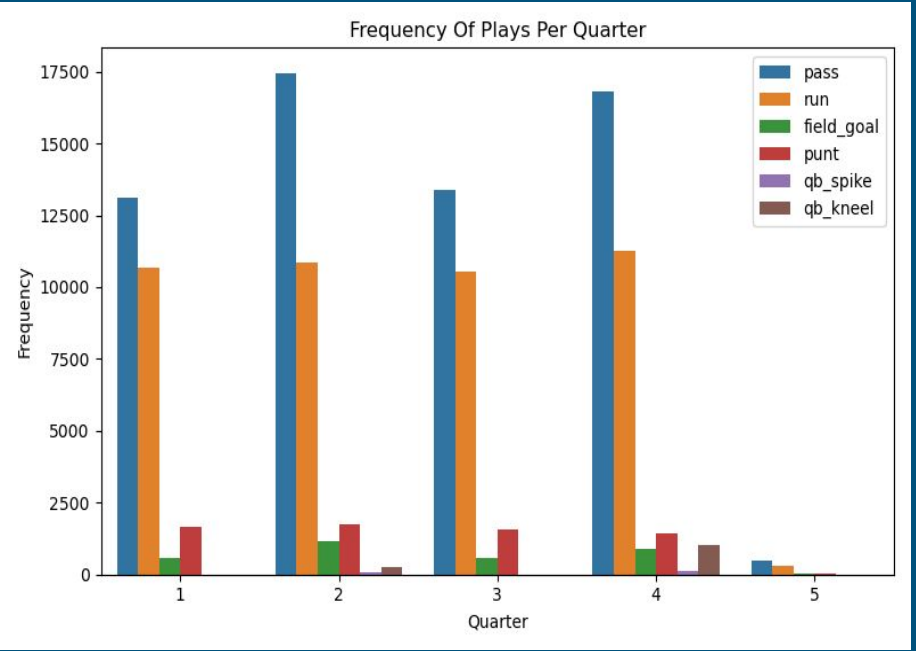
What types of plays

How's it look in our data?

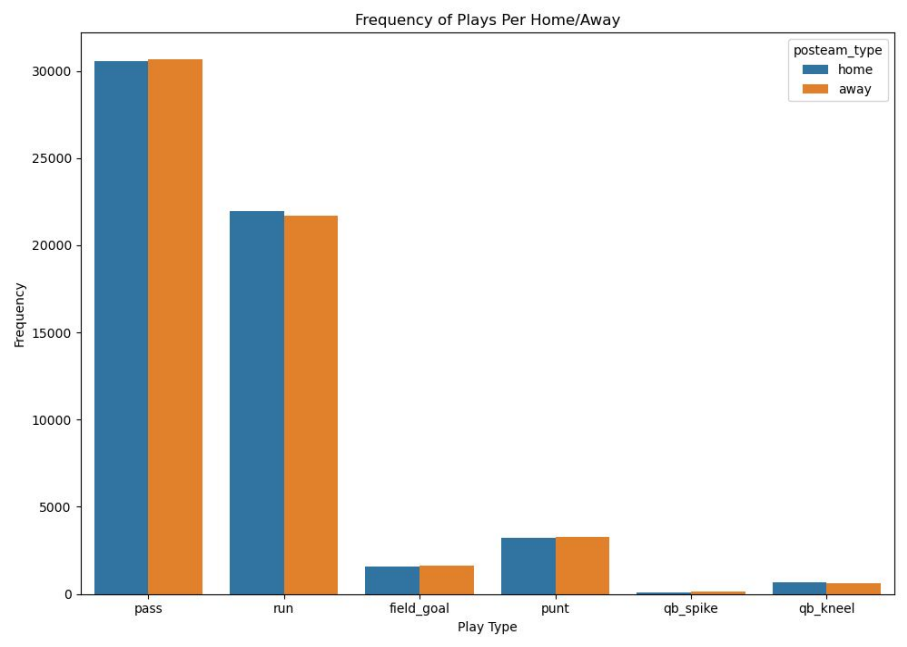


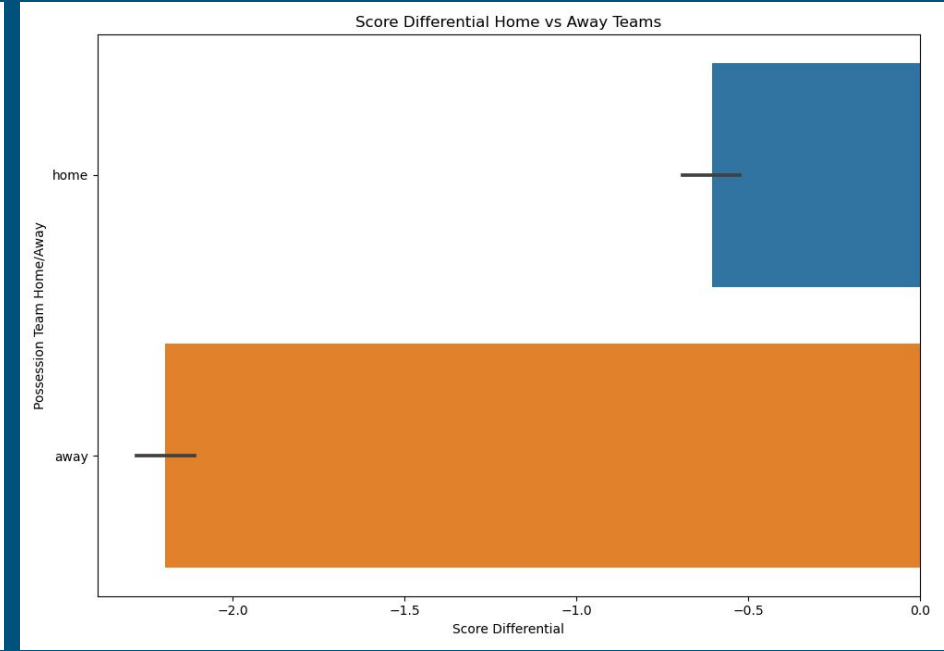
How do plays match with other features?





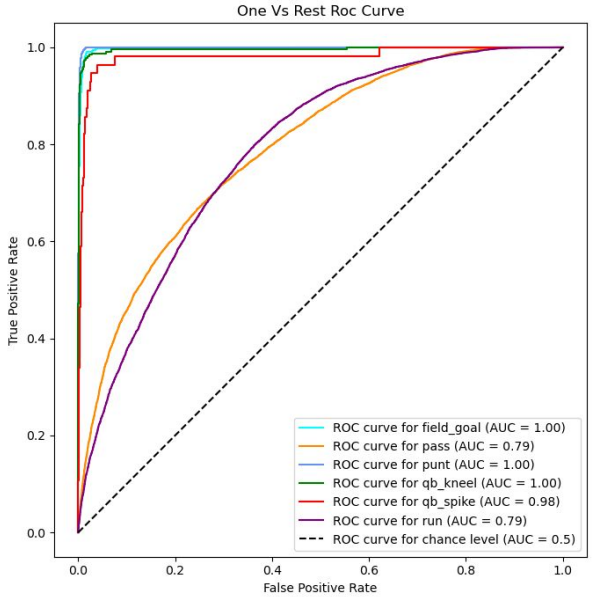
Home or not to home



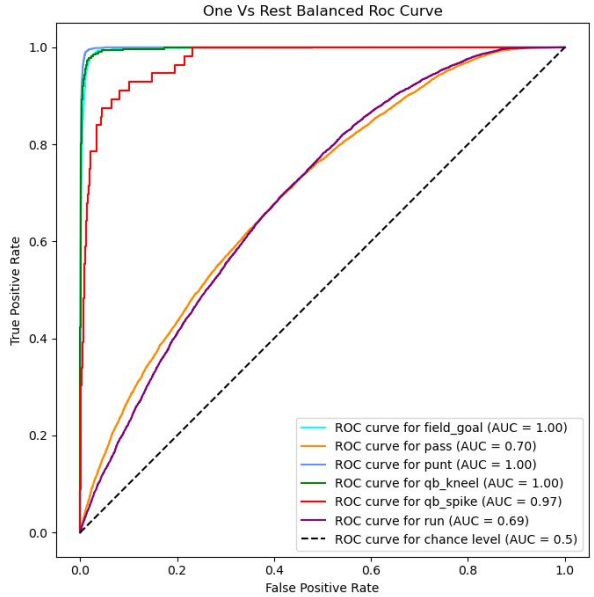


Results

Unbalanced



Balanced



- Play
- Qb Spike

Conclusions

- Our Xgboost imbalanced model achieved an accuracy of 70%, which is 18% better than our baseline accuracy.
- Our model had a slight difficulty classifying the pass and run play most likely due to the randomness of the NFL/ how past plays affect the next play/etc. However its main struggles was classifying qb spikes, because those are dependent on time outs a team has and if the clock is still running, how much time is left , which are features we do not have in our dataset.

Future

- Create a model where better imbalanced data replacement techniques were implemented
- Look for more data where more features are available, eg: Formation of offense, weather of the day.