ORIE 4741 Project Proposal

Jibran Gilani (jg793), Milan Shah (mrs282)

September 2019

1 Background

During each NFL game, we only see about 11 minutes of action during the 60 minute runtime. Even then, the action and results are only a fraction of the research and work that gets put in by coordinators and coaches to prepare for matchups. As a defensive coordinator, part of that research is being able to guess what the opposing offense will do for each play and provide the proper scheme and design against that play. This process depends heavily on their ability to correctly predict what the offense has planned.

Research question: Would it be possible to use machine learning techniques to help NFL defenses combat offenses by predicting the most likely play they will run?

2 Data

There is a detailed NFL play-by-play dataset that spans from 2009-2018 that has has 356,768 rows and 100 columns. Each row represents the outcome of a single play in a game and the columns provide information including game score, time, play type (run vs pass), game situation (down, distance, and yard line), result, and more. This data can be divided to analyze the tendencies of specific teams or offensive coordinators. Researchers from Carnegie Mellon University used an API managesd by the NFL to create this dataset (more information is on https://github.com/ryurko/nflscrapR-data.)

3 Goals

We plan on using the dataset to predict what play type will appear next given certain variables about the game conditions. As of now, we will approach this as a binary classification problem at first but may also expand to predict pass length (shallow, medium, deep) or field location (right, middle, left). Even more depth could be added if we include type or run or pass (e.g. pitch or play action) but the initial focus will be on predicting run vs pass. We also may add to the dataset by including data about the opponent that may factor into a coordinators decision. For example, playing against the top ranked run defense will make a team more prone to pass and will definitely factor into a coordinators game plan. Once a model is built, it can be tested against ongoing games during the 2019 season if there is a team, offensive coordinator and offensive player combination that is similar to the years in the dataset.