Brett Mele

5/7/2023

MSDS458

Final Project Proposal

In the National Football League (NFL), anticipating what play an opponent may run on offense
can make a substantial impact on the eventual outcome of the game. Having the right play
called to defend a passing play, and having defenders anticipating a pass, will make it
increasingly difficult for the offense to find success. Conversely, using a play to defend the run
when the offense actually passes the ball gives the offense a significant advantage.

For the final project, I plan to build a model to predict the next offensive play type within an NFL
game using play-by-play data provided by the nflfastR package. The targets for prediction will
be all the possible play types within an NFL game, excluding kickoffs and extra points (field
goal, pass, run, punt). In addition to play type, for run plays I plan to also include the direction of
the running play (left, right middle) as this also has a meaningful impact on the potential
defensive play calls. Thus, the problem is a multi-class classification with six potential outcomes
(field goal, punt, pass, run left, run middle, run right). Given each play depends on plays
previously called, along with the game state, the final model architecture will be a recurrent
neural network (RNN).

The project will address all of the grading guidelines listed in the final assignment. Play
predictions would be a valuable tool for an NFL team, and clubs have likely already attempted
and implemented similar models. Public research using NFL data has proliferated in recent
years thanks to the NFL Big Data Bowl Kaggle competition, so there will be existing research to
draw from. Play type prediction models have already been constructed using tree based

methods, but these existing models only rely on the game state and not prior play sequences. Regarding research design, data is available via the nflfastR package for all NFL seasons dating back to 1999 and all necessary features are provided or can be easily engineered. The research problem should allow for substantial experimentation with different types of network architectures, with the final model likely using some kind of RNN variant.