Jonathan Cao L30045431

Task 1: Initial Proposal

Master’s Thesis

**Fantasy Football Team Generator Proposal**

My project proposal is to create a fantasy football team generator that will directly compete in a live draft against human participants. The model will be using various machine learning techniques to train, utilizing data from past fantasy seasons. Significant features will be picked out, and each player in the league will be given a rating based on their cost to value ratio. During the live draft as the rounds progress, the generator will attempt to output the next best player according to your current team’s composition. The generator will readjust itself depending on the remaining players in the pool. I would also like to implement a way for the model to readjust depending on the composition of the other teams participating in the league, however, I am still unclear on how to proceed with that.

I drew inspiration for this project from a video that a professor in one of my earlier classes showed me about not necessarily having to be knowledgeable in a specific area to be able to build a prediction model in it. According to AI researcher Tapiwa Chiwewe (2018) “You do not have to be a domain expert to solve big problems”. I personally have no knowledge of football at all, and it will be interesting to utilize what I have learned during this program to try to compete against my friends who are so-called “experts” at fantasy football. When Jeremy Howard, the president of Kaggle, was asked whether or not ‘specialized subject experts’ ever enter their data prediction competitions, Jeremy (2012) interestingly replies “Oh yes. Every time a new competition comes out, the experts say: “We’ve built a whole industry around this. We know the answers.” And after a couple of weeks, they get blown out of the water.”

**Methodologies, Data Collection/Training**

I will be using python with various libraries such as pandas to build the generator. As stated earlier, player statistics and data will be gathered from the 2015, 2016, 2017, 2018, and 2019 football seasons. I will then gather data from the top performing fantasy teams in their respective leagues for their corresponding seasons; perhaps 5 winning fantasy teams from each season? The teams will be compared to see what specific statistics they all have in common, and I will attempt to figure out what significant features make these teams successful so that I can throw them into a machine learning model. More research is needed into which exact model will be used, however, several will be explored to see which will provide the best outcome towards our target goal including but not limited to: XGBoost, ARIMA, and Recurrent Neural Network Long Short-Term memory.

The data will be obtained from various websites including <https://www.fantasyfootballdatapros.com/csv_files>, <https://www.pro-football-reference.com/> , kaggle.com and other sources that I have not yet researched. The data will be parsed using various techniques in natural language processing including but not limited to: normalization, text classification, and naive bayes.

This project is seemingly straight forward, but will require me to explore various ML models and techniques to build something that is usable. Please let me know whether or not this proposal topic is robust enough to proceed as a thesis project.

**References**

Chiwewe, T. (2018, February 23). *You don’t have to be an expert to solve big problems* [Video]. TED Talks. <https://www.ted.com/talks/tapiwa_chiwewe_you_don_t_have_to_be_an_expert_to_solve_big_problems?language=en>

Aldhous, P. (2012, December 8). *Specialist Knowledge Is Useless and Unhelpful*. Slate Magazine. <https://slate.com/technology/2012/12/kaggle-president-jeremy-howard-amateurs-beat-specialists-in-data-prediction-competitions.html>

King, J. (2020, June 22). *Using Machine Learning to Predict Fantasy Football Points*. Medium. <https://towardsdatascience.com/using-machine-learning-to-predict-fantasy-football-points-72f77cb0678a>

Parthasarathy, S. (2021, February 12). *Top 5 Predictive Analytics Models and Algorithms | Logi Analytics Blog*. Logi Analytics. https://www.logianalytics.com/predictive-analytics/predictive-algorithms-and-models/