Cameron Grande, MD Hossain, Evan Cheney

Fantasy football report

Abstract:

Fantasy sports have become an industry filled with large prize pools that are up for the taking by those who prove best to defeat their competition. Overall, the fantasy sports industry awarded 6.18 billion dollars to players in 2019 alone. 2.91 was awarded through the more niche version of fantasy sports, which are Daily Fantasy Sports (DFS). These contests involve selecting players weekly (or daily, dependent on the sport) basis to win competitions that come with potentially large cash prizes. The advantage that DFS has over traditional sports betting is that the contests are a peer-to-peer (P2P) format, rather than against the “House” or a Sportsbook. Players tend to have a more significant advantage over their other humans than against a sharp Sportsbook. Previous research on the topic has involved creating projections or best predictions of how many fantasy points a real-life player will have in a game to see if the researcher(s) can predict the outcomes of games better than their peers in the contests. However, this previous research does not take into account that DFS is a zero-sum game and must be played as such. Thus, it is necessary to create a system that can make high expected value (EV) decisions that can outperform peers using a combination of traditional projections and optimal game theory.

We propose that it is possible to create player clusters based on their leading form statistics that allow us to create a player-based environment. Using historical contest data will allow our model to learn policies based on the contest environment. While this may be better performed in a reinforcement learning environment, we propose that it is possible under the conditions of unsupervised and supervised learning structures.

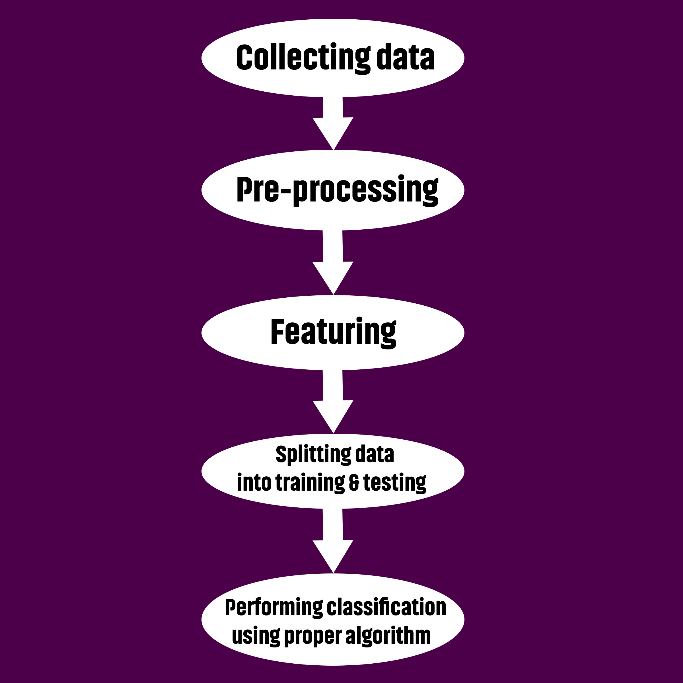
Goals:

* Create a model to find the optimal strategy for weekly DraftKings fantasy football contests based on the given weekly environments and slate conditions

Objectives:

* Create optimal strategies for the use of weekly DraftKings fantasy football contests

Class Block Diagram:



Contribution Plan:

• Data Collection-Evan

• Interpretation of Data (what does it mean, what can we predict from it) -Evan

• Formatting data to be assessed. -Cameron

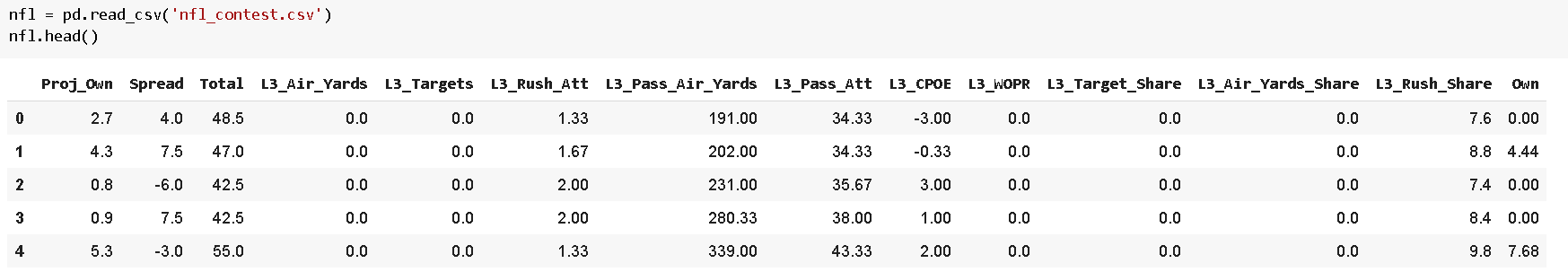
• Choosing tools/algorithms to use. -Md

• Running data through said algorithms. - Cameron

• Analyzing results -Md

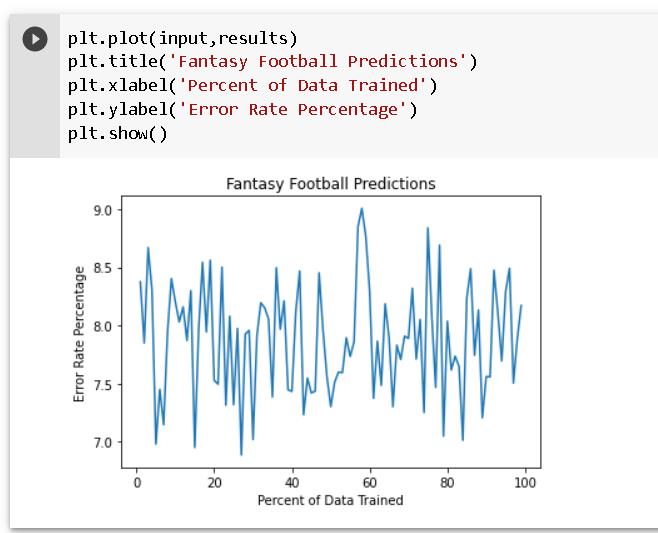
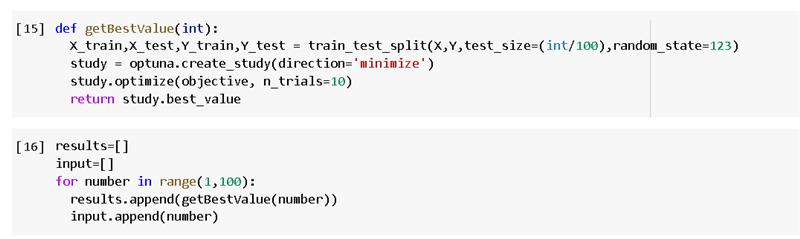
• Documenting each step of the way and writing out our results -Cameron

What our project looked like:

 We used Python to pull in and train data. Our algorithm focused on the different attributes and their relationship to the own column, which can be seen above. The own column is what we were primarily attempting to predict. Evan in a many way took lead on this project, as he was the one who had this idea, and he was the only one in our group that understood fantasy football on a deep level. MD and myself (Cameron) focused more on documentation, code functionality, and communicating with Evan to understand what this project should look like and making sure everything went smoothly. There was one sizable bug Evan was running into that took all of us looking into to solve. Cameron debugged it and noticed that it had to do with a line in our self-made function that helped retrieve the mean squared error, and we were able to fix and proceed from there. We ended up with a functional product that was able to make predictions with a relatively good accuracy!

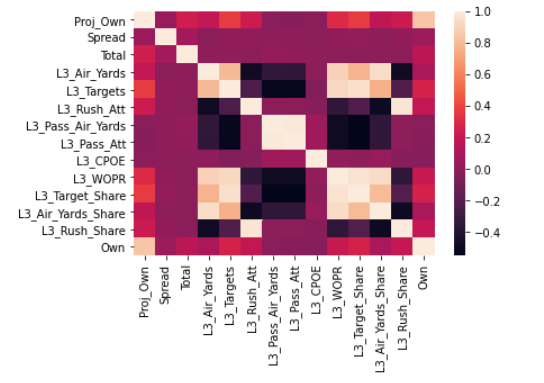
The function Evan created was as follows: 

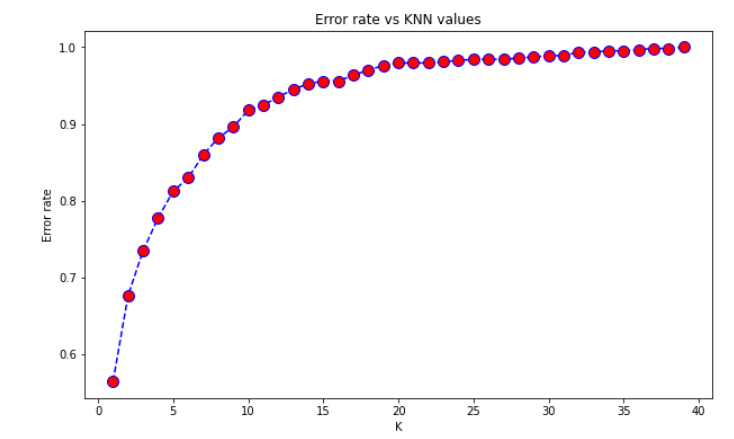
After we had this trial working, we were able to create some visuals to help us understand our data on a different level. Cameron created the following visual to help us understand what the best train-test split was for each integer and to see the accuracy of our algorithm in a range.



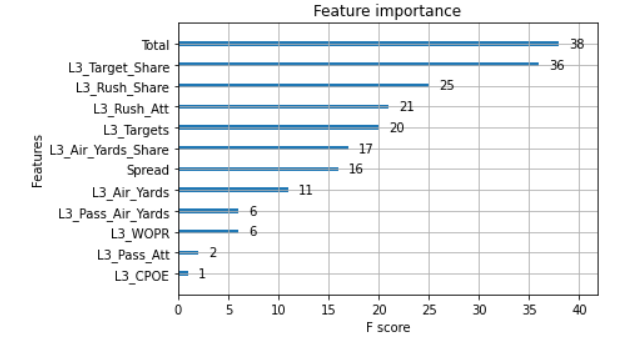
We can see from the above that we had an error rate that fluctuated from roughly 9 to 7 percent, which is overall a pretty good accuracy.

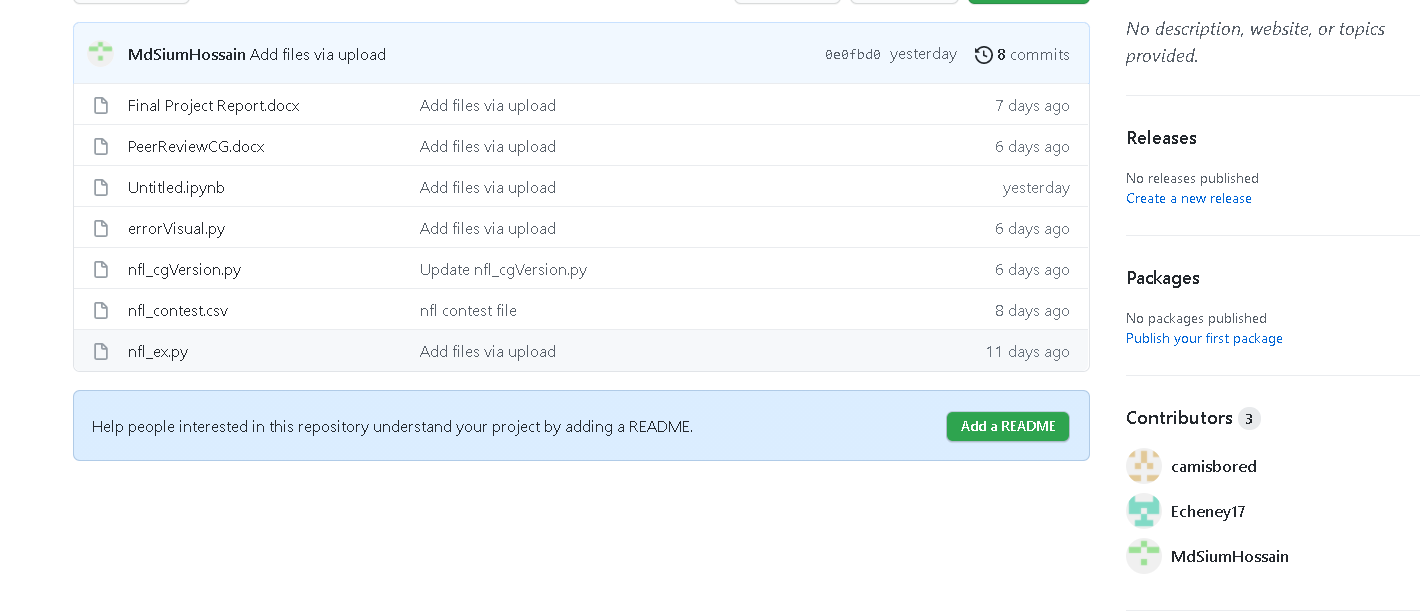
MD created the following visual to help aid in the interpretation of our data.



MD also used a KNN algorithm to predict for our model as well. The results of his efforts showed our model does much better with a lower K number.

Evan used the following visuals to describe the importance of the features in our model:



We maintained our project using GitHub, which we all contributed different files to which all became a part of our final product. 

We all took turns and did different pieces for this project. We enjoyed doing it along the way and our efforts resulted in an effective and well presentable product.