Deliverable 3: Report

Vision

Our current vision for the system we are building is to create a sports betting system for the National Basketball Association (NBA). The goal focuses on implementing statistical analysis through intensive research regarding sports betting algorithms. Through the evaluation of each algorithm, we will gather data to gauge which algorithm is most accurate with present day NBA outcomes. Specifically, maximizing the accuracy for sports betting players over under betting. While our original plan was to build a sports betting algorithm that was ideally applicable for sports betting outside of strictly the NBA, such as football and soccer, we decided to narrow our vision to perfect the precision of the player over under calculations for the NBA. Through specifying our goals for the algorithm, we are able to divert our main focus to maximize performance of one form of sports betting. In terms of evolutions within the system, we originally wanted to construct our own file reader through creating tables based on CSV files from self-coded functions. However, after implementing columns and data table functions, it was very difficult to debug, so we decided to use OCaml CSV to read and square the data. After extensive research, we also decided on using OCaml Owl for as our numerical library to assist in the statistical analysis. Overall, we have adjusted our goals through trial and error to improve the accuracy of our statistical analysis for NBA sports betting.

Summary of Progress

Between MS1 and MS2, Eric used python code to web scrape score data from official NBA statistics that will be used to create a CSV with data from multiple seasons. Tim implemented a simple average, weighted average, simple moving average, and weighted moving

average for statistical analysis. Tanish created functions to input a CSV file path to create results when the structures are utilized in models. Anthony created a data table structure to manage the processed CSV file to filter through relevant data. Lastly, Selena focused on algorithmic research and debugging with OCaml CSV functions to ensure the data is operational.

Activity Breakdown

For each team member, give a bulleted list of the responsibilities that team member had, the activities in which they participated, the features they delivered, and the number of hours they spent working.

Eric

- Found sqlite database from Kaggle with NBA game data, used python to create a CSV with only the relevant data
- Created python code to web scrape advanced box score data from official NBA statistics website with functionality to combine data from multiple seasons
- Hours Worked: 5

Tim

- Implemented simple average and weighted average for column data structure (essentially mutable arrays, using fold left function)
- Implemented simple moving average and weighted moving average for both column data structures (arrays) and float lists (most likely going to be used due to CSV parsing)
- Hours Worked: 5.5

Tanish

- Developed OCaml functions to take in file path for a CSV and convert to string list list and float list list – now we have the ability to plug these structures into models for results

- Researched using linear regression to regress upon player box score stats with points scored as the output feature – found Owl OCaml library which will be helpful

- Hours Worked: 5

Anthony

- Made major progress in creating the data table structure and managing how we process CSVs

- Worked on pipeline from an entire CSV data file to filtering down to the relevant data for each model (particular player, a certain time period, playing a particular team, etc.)

- Hours Worked: 6

Selena

- Researched sports betting algorithms to consider best approaches to evaluate player over under bets

- Read OCaml CSV documentation to create the data into "square" to apply operations on each data point

- Hours Worked: 4.5

Productivity Analysis

We were a very productive and organized group. We had multiple work sessions a week since MS1, and came back to work well together. At the start of each work session, we set goals for each subgroup (algorithm development and data engineering) and did not end the session until the goal was reached.

Our goals were quite accurate for the most part – we made sure to do our research first about a specific task before setting an actual goal for it. We were able to achieve most if not all of the goals that we set for ourselves in terms of web scraping, setting up infrastructure for loading CSV data, and developing initial moving average models.