Spring 2024 Erdos Institute Project Summary

Title: Valorant Champions 2023 Tournament
Team: Ahram Lim, Amogh Parab, Yuqing (Frank) Lin, Alex Elchesen
Github: https://github.com/ahrlim/vct-erdos-project

Overview:

Valorant is a first-person tactical hero shooter video team game created by Riot Games. Throughout the year professional Valorant teams participate in various tournaments to claim the title of Valorant Champions. We want to create a model that predicts the results of 2023 Valorant championship matches using the regular season data. Also, we want to figure out which aspects of the game matter the most in such models.

Stakeholders: Riot Games, Valorant Teams, Online Betting, Broadcasters

Data Collection:

We have a vast collection of data from professional tournaments played over the years 2021 to 2023. The data includes results, economy data, individual player and team statistics, and team compositions; at the levels of both individual rounds and an entire game (which usually lasts 13 to 24 rounds). We collected this data from Kaggle.com.

The data was spread over various files with varying amounts of overlapping and missing data relative to each other. We put substantial effort into unifying the data into a single dataframe while preserving key features. We indexed our data by games, with around 160 features per game after feature engineering.

Model:

We looked at this problem from a number of different angles, focusing primarily on player performance, team synergy (kills), team economy, and team performance in clutch rounds.
We looked at player/role composition in a team on a map. A team will have four role ratings/stats depending on the player playing that role.

In the economy approach, we used the data credits spent by each team and their opponents to come up with a rating system. This rating gives higher advantage to the team which beats worse odds. For example, a team winning with more credits spent than their opponent will earn less credits than the team winning with less credits spent than their opponent. The goal of this rating is to summarize the economic data in a number. Using economy data alone, logistic regression gave up to 60% accuracy.
We also trained KNeighborsClassifier, DecisionTreeClassifier, RandomForestClassifier, and XGBClassifier on individual players' stat, such as, rating, average combat score, kills per round, etc. Most of them gave accuracy less than 60% which implied our own economy rating is significant.

Finally, we combined all above approaches to get a better model and also to compare these approaches. We used XGBoost to find the most important features out of more than 160 features we have. Then we trained KNeighborsClassifier, DecisionTreeClassifier, RandomForestClassifier, and XGBClassifier on those important features. With the default hyperparameters, they gave 52-64% accuracy, with the XGBClassifier performing the best. By turning hyperparameters, we were able to increase the accuracy of the XGB classifier to 69-70%.

Future Direction: In our analysis we realized that in a match certain rounds matter more than others. Also, round win/loss momentum has a role in the match result. We would like to incorporate this in our model.