Apply Machine Learning

## How do you frame your main question as a machine learning problem?

Since this dataset can be used to augment my own knowledge of the game based on personal observations of other players, this data can be used to learn more about player behavior in the game. As a machine learning problem, this dataset would be best fit in a clustering model that can be analyzed via k-means or a similar approach.

## Is it a supervised or unsupervised problem?

This is an unsupervised problem where there is likely no correct answer given the values we have. If we observe the data we are given and have cleaned, then we see trends rather than any specific output. There are many categories in the datasets and as such we can classify this data as a clustering problem since there are inherent groupings and behaviors in the data that can be observed.

## What are the main features (also called independent variables or predictors) that you’ll use?

* K/D (kill/death ratio)
* Win ratio
* Wins (by total value)
* Kill streaks (which would be confined to a single best in a round)

The rest of the data has potential to be supporting aspects. I say this because for example, K/D by itself may say something about the player, however if we compare it to the statistic known as “Longest Kill” we could learn more about the player’s behavior in game. If a player has a longer range then it could show they prefer a very safe approach to combat (via safer engagement distance) and/or their skill in long range combat. On the other hand, a shorter range could indicate a preference as well. Shorter ranges could indicate comfort with close range combat that could range from house to house fighting or perhaps significant use of flanking manuevers. Although this cannot be used to certainty, it can be an indicator toward for players that are outliers. Outliers tend to show a higher degree of skill (or just downright cheating), and thus we can perhaps infer a subset of skills they possess based on other statistics.

## Which machine learning technique will you use?

For now I have decided upon clustering and K-means as the main method. I will subject this to change if I realize there is a better algorithm/technique to utilize.

## How will you evaluate the success of your machine learning technique? What metric will you use?

I believe the success of the technique will fall toward two factors. The identification of significant outliers (which could indicate an exceptional level of skill thus up for further analysis) and significant shifting points in K via the elbows. These factors could help determine whether or not these players at those areas are truly skilled or just playing too many custom games or cheating. With those specific points identified we can cross check the statistics of those pool of players and perhaps identify behavioral patterns that we can learn from. Since I have noticed some of the players in the dataset record their games, this would help in analysis.