Team members:
Simon Wang (ssw8641)
Mengya Qiu (mq583)
Bo Su (bs3957)
Binfang Ye (by2034)

Dataset link:
https://www.kaggle.com/bobbyscience/league-of-legends-diamond-ranked-games-10-min

League of Legends (hereinafter called LOL), which is a multiple online battle game with two different teams in one game, enjoys popularity in the world. There are many organizations holding LOL competitions but it is hard to tell who will win the game until one team takes down another team's Nexus. We are interested in predicting who will win the game in 10 mins based on some features such as "first blood", "kills", "deaths", and etc. The data is collected in the first 10mins of the game.

This dataset contains the first 10 min statistics of roughly 10k games. There are 19 features collected in 10min in-game for both red & blue teams respectively. The column 'blueWins' is the target variable we are trying to predict: A value of 1 means the blue team has won and otherwise 0. Besides these 39 features, there is a unique gameid assigned to each unique game. Hence there are 40 columns in total.

Our main goal is trying to predict the outcome of the game given by the above statistics of the first 10-min for each game. This could be modeled as a binary classification problem with a binary target value 0 or 1 and other 38 attributes for training and test. Hence, using a soft classifier such as a logistic regression model to obtain a probability that the blue team wins the game would be a reasonable choice. Beyond that, we also try to get more insights about which specific features are decisive for the final outcome. Therefore, typical regularizers like LASSO could be implemented for further feature selection.