

Isaac Min

CPSC 392

6 May 2022

Variables (39 total)

- blueWins: 1 if won, 0 if not
- blueWardsPlaced: number of wards placed in total by blue team
- blueWardsDestroyed: number of wards destroyed by blue team
- blueFirstBlood: 1 if yes, 0 if no
- blueKills: number of total kills by blue team
- blueDeaths: number of total deaths by blue team
- blueAssists: number of total assists in blue team
- blueEliteMonsters: total number of elite monsters slain by blue team (dragons and heralds)
- blueDragons: total number of dragons slain by blue team
- blueHeralds: total number of Rift Heralds taken by blue team
- blueTowersDestroyed: total number of turrets taken by blue team
- blueTotalGold: total gold balance of blue team
- blueAvgLevel: average level of blue team
- blueTotalExperience: total experience of blue team.
- blueTotalMinionsKilled: total number of minions killed by blue team
- blueTotalJungleMinionsKilled: total number of jungle monsters killed by blue team
- blueGoldDiff: difference in gold between blue and red team
- blueExperienceDiff: difference in experience between blue and red team
- blueCSPerMin: creep score per minute of blue team
- blueGoldPerMin: gold per minute of blue team
- redWardsPlaced: number of wards placed by red team
- redWardsDestroyed: number of wards destroyed by red team
- redFirstBlood: 1 if yes, 0 if no
- redKills: total number of kills by red team
- redDeaths: total number of deaths by red team
- redAssists: total number of assists by red team
- redEliteMonsters: total number of elite monsters slain by red team (dragons and heralds)
- redDragons: total number of dragons killed by red team
- redHeralds: number of Rift Heralds slain by red team
- redTowersDestroyed: number of turrets destroyed by red team
- redTotalGold: total gold of red team
- redAvgLevel: average level of red team

- redTotalExperience: total experience of red team
- redTotalMinionsKilled: total number of minions killed by red team
- redTotalJungleMinionsKilled: total number of jungle monsters killed by red team
- redGoldDiff: gold difference between red team and blue team
- redExperienceDiff: experience difference between red team and blue team
- redCSPerMin: creep score per minute of red team
- redGoldPerMin: gold per minute of red team

Question 1

Question: When predicting if blue team won, which model (Linear Regression, Logistic Regression, Decision Trees, Naïve Bayes, K-Nearest Neighbors) has the highest accuracy score (confusion matrix)?

Variables Involved: blueWins (binary), blueWardsPlaced (continuous), blueWardsDestroyed (continuous), bluefirstBlood (binary), blueKills (continuous), blueDeaths (continuous), blueAssists (continuous), blueEliteMonsters (continuous), blueDragons (continuous), blueHeralds (continuous), blueTowersDestroyed (continuous), blueTotalGold (continuous), blueAvgLevel (continuous), blueTotalExperience (continuous), blueTotalMinionsKilled (continuous), blueTotalJungleMinionsKilled (continuous), blueGoldDiff (continuous), blueExperienceDiff (continuous), blueCSPerMin (continuous), blueGoldPerMin (continuous), redWardsPlaced (continuous), redWardsDestroyed (continuous), redFirstBlood (binary), redKills (continuous), redDeaths (continuous), redAssists (continuous), redEliteMonsters (continuous), redDragons (continuous), redHeralds (continuous), redTowersDestroyed (continuous), redTotalGold (continuous), redAvgLevel (continuous), redTotalExperience (continuous), redTotalMinionsKilled (continuous), redTotalJungleMinionsKilled (continuous), redGoldDiff (continuous), redExperienceDiff (continuous), redCSPerMin (continuous), redGoldPerMin (continuous)

Cleaning: Missing values (in any) will be dropped from the dataset

Modeling/Computation: A train/split test with an 80/20 split will be utilized throughout all models. All continuous variables will be standardized via z-scoring. A linear regression model, a logistic regression model, a decision tree, a Naïve Bayes model, and a K-Nearest Neighbors model will be fit to predict if blue team won.

Graphs: A confusion matrix showing how many games the model was able to predict the blue team winning will be created for each model. Each accuracy score will be plotted in a bar chart to show how accurate each model is.

Brief Discussion of why analysis is effective at answering question: This analysis is effective because it compares different models that use all the predictors to see how the accuracy of predicting the blue team winning changes. The confusion matrices and bar graph will help

illustrate the numeric results and allow for a visual comparison of the accuracy across all of the models.

Question 2

Question: Is total experience related to the total gold (for both blue and red teams)?

Variables Involved: blueTotalGold (continuous), blueTotalExperience (continuous), redTotalGold (continuous), redTotalExperience (continuous)

Cleaning: Missing values (in any) will be dropped from the dataset

Modeling/Computation: Continuous variables will be z-scored.

Graphs: A scatterplot will be created to compare the blueTotalGold and blueTotalExperience. Another scatterplot will be created to compare the redTotalGold and redTotalExperience.

Brief Discussion of why analysis is effective at answering question: This analysis is effective because the graphs illustrate the relationship (if any exists) between gold and experience between both the blue team and red team. It will effectively show if there is a correlation between the two factors.

Question 3

Question: With total minions killed, total experience, and total kills in consideration, what clusters emerge and what are the characteristics of these clusters?

Variables Involved: blueTotalMinionsKilled (continuous), blueTotalExperience (continuous), blueKills (continuous), redTotalMinionsKilled (continuous), redKills (continuous)

Cleaning: Missing values (in any) will be dropped from the dataset

Modeling/Computation: Scatterplots will be constructed to evaluate the rough relationship between the variables. After evaluating the scatterplot, a clustering method (KMeans, EM, DBSCAN, or Hierarchical Clustering) will be selected for clustering. Continuous variables will be z-scored.

Graphs: A graph for the respective clustering (e.g., dendrogram for hierarchical clustering) will be created to view the clusters that emerge from the variables.

Brief Discussion of why analysis is effective at answering question: This analysis is effective because it evaluates the clusters that emerge from minions killed, total experience, and total kills in both teams. It will help visualize characteristics of each cluster and how they may contribute to a team winning a game.

Question 4

Question: When comparing a model predicting if blue team won using PCA on the continuous variables of this dataset and retaining enough principal components to keep 85% of the variance, to the models in question 1, what is the difference in R^2 and Mean Absolute Error (MAE)?

Variables Involved: blueWardsPlaced (continuous), blueWardsDestroyed (continuous), blueKills (continuous), blueDeaths (continuous), blueAssists (continuous), blueEliteMonsters (continuous), blueDragons (continuous), blueHeralds (continuous), blueTowersDestroyed (continuous), blueTotalGold (continuous), blueAvgLevel (continuous), blueTotalExperience (continuous), blueTotalMinionsKilled (continuous), blueTotalJungleMinionsKilled (continuous), blueGoldDiff (continuous), blueExperienceDiff (continuous), blueCSPerMin (continuous), blueGoldPerMin (continuous), redWardsPlaced (continuous), redWardsDestroyed (continuous), redKills (continuous), redDeaths (continuous), redAssists (continuous), redEliteMonsters (continuous), redDragons (continuous), redHeralds (continuous), redTowersDestroyed (continuous), redTotalGold (continuous), redAvgLevel (continuous), redTotalExperience (continuous), redTotalMinionsKilled (continuous), redTotalJungleMinionsKilled (continuous), redGoldDiff (continuous), redExperienceDiff (continuous), redCSPerMin (continuous), redGoldPerMin (continuous)

Cleaning: Missing values (in any) will be dropped from the dataset

Modeling/Computation: Continuous variables will be z-scored. The eigenvector and eigenvalues of the covariance matrix will be calculated to determine the principle components of the dataset. These values will then be ordered to determine the principle components in the order of significance. A PCA model will be fit and the R^2 and MAE will be pulled. R^2 and MAE values will be pulled to construct

Graphs: A bar graph will be made comparing the R^2 and MAE values for each of the models made in question 1 and the PCA model.

Brief Discussion of why analysis is effective at answering question: This analysis is effective because it compares full models (all predictors) with a model with dimensionality reduction to see the differences in R^2 and MAE values. The plot will help visualize the numeric values and demonstrate the differences in the R^2 and MAE values.

Question 5

Question: Looking at the coefficients for each model created thus far, which variable was the strongest predictor of winning a game?

Variables Involved: blueWins (binary), blueWardsPlaced (continuous), blueWardsDestroyed (continuous), bluefirstBlood (binary), blueKills (continuous), blueDeaths (continuous), blueAssists (continuous), blueEliteMonsters (continuous), blueDragons (continuous), blueHeralds (continuous), blueTowersDestroyed (continuous), blueTotalGold (continuous), blueAvgLevel (continuous), blueTotalExperience (continuous), blueTotalMinionsKilled (continuous), blueTotalJungleMinionsKilled (continuous), blueGoldDiff (continuous),

blueExperienceDiff (continuous), blueCSPerMin (continuous), blueGoldPerMin (continuous), redWardsPlaced (continuous), redWardsDestroyed (continuous), redFirstBlood (binary), redKills (continuous), redDeaths (continuous), redAssists (continuous), redEliteMonsters (continuous), redDragons (continuous), redHeralds (continuous), redTowersDestroyed (continuous), redTotalGold (continuous), redAvgLevel (continuous), redTotalExperience (continuous), redTotalMinionsKilled (continuous), redTotalJungleMinionsKilled (continuous), redGoldDiff (continuous), redExperienceDiff (continuous), redCSPerMin (continuous), redGoldPerMin (continuous)

Cleaning: N/A

Modeling/Computation: Coefficient values will be obtained for each variable in each model created. Each of these values will be stored on a separate data frame.

Graphs: A bar graph will be made that compares the coefficients values for each variable across all of the models that were made.

Brief Discussion of why analysis is effective at answering question: This analysis is effective because it compares the different coefficient values that each predictive model produces. It will help visualize the different numeric values as well as see which variables (if any) hold a consistent coefficient value.

Question 6

Question: Are the models in #1 more accurate for blue teams who had over 15,000 gold at the 10-minute mark than those who had less than 15,000 gold?

Variables Involved: blueWins (binary), blueWardsPlaced (continuous), blueWardsDestroyed (continuous), blueFirstBlood (binary), blueKills (continuous), blueDeaths (continuous), blueAssists (continuous), blueEliteMonsters (continuous), blueDragons (continuous), blueHeralds (continuous), blueTowersDestroyed (continuous), blueTotalGold (continuous), blueAvgLevel (continuous), blueTotalExperience (continuous), blueTotalMinionsKilled (continuous), blueTotalJungleMinionsKilled (continuous), blueGoldDiff (continuous), blueExperienceDiff (continuous), blueCSPerMin (continuous), blueGoldPerMin (continuous), redWardsPlaced (continuous), redWardsDestroyed (continuous), redFirstBlood (binary), redKills (continuous), redDeaths (continuous), redAssists (continuous), redEliteMonsters (continuous), redDragons (continuous), redHeralds (continuous), redTowersDestroyed (continuous), redTotalGold (continuous), redAvgLevel (continuous), redTotalExperience (continuous), redTotalMinionsKilled (continuous), redTotalJungleMinionsKilled (continuous), redGoldDiff (continuous), redExperienceDiff (continuous), redCSPerMin (continuous), redGoldPerMin (continuous)

Cleaning: Missing values (in any) will be dropped from the dataset

Modeling/Computation: Matches will be filtered and divided into blue teams that had over 15,000 gold at the 10-minute mark and those who had less than 15,000 gold at the 10-minute

mark. Continuous variables will be z-scored. Models in question 1 will be recreated for these two groups and the accuracy score will be obtained.

Graphs: A confusion matrix showing how many games the model was able to predict the blue team winning will be created for each model. Each accuracy score will be plotted in a bar chart to show how accurate each model is.

Brief Discussion of why analysis is effective at answering question: This analysis is effective because it compares how different total gold values may affect a model's accuracy. The confusion matrices will show how many match outcomes the model was able to predict correctly and incorrectly. The bar plot will help visualize the numeric results to compare it to accuracy scores from the models in question 1.

Question 7

Question: Are there differences in how accurate each of the models in #1 are for the different teams (blue/red)?

Variables Involved: blueWins (binary), blueWardsPlaced (continuous), blueWardsDestroyed (continuous), blueFirstBlood (binary), blueKills (continuous), blueDeaths (continuous), blueAssists (continuous), blueEliteMonsters (continuous), blueDragons (continuous), blueHeralds (continuous), blueTowersDestroyed (continuous), blueTotalGold (continuous), blueAvgLevel (continuous), blueTotalExperience (continuous), blueTotalMinionsKilled (continuous), blueTotalJungleMinionsKilled (continuous), blueGoldDiff (continuous), blueExperienceDiff (continuous), blueCSPerMin (continuous), blueGoldPerMin (continuous), redWardsPlaced (continuous), redWardsDestroyed (continuous), redFirstBlood (binary), redKills (continuous), redDeaths (continuous), redAssists (continuous), redEliteMonsters (continuous), redDragons (continuous), redHeralds (continuous), redTowersDestroyed (continuous), redTotalGold (continuous), redAvgLevel (continuous), redTotalExperience (continuous), redTotalMinionsKilled (continuous), redTotalJungleMinionsKilled (continuous), redGoldDiff (continuous), redExperienceDiff (continuous), redCSPerMin (continuous), redGoldPerMin (continuous)

Cleaning: Missing values (in any) will be dropped from the dataset

Modeling/Computation: A new column "redWins" will be created to demonstrate if the red team won. Models created in question 1 will be recreated to predict redWins. The accuracy score will then be pulled for each of the models.

Graphs: Confusion matrices will be created for each of the new models to illustrate how many games the model was able to correctly/incorrectly predict that red team won. The accuracy values will be plotted in a bar chart and compared to that of the accuracy of the models for blue team winning.

Brief Discussion of why analysis is effective at answering question:

This analysis is effective because it compares the models used to predict if blue team won with the models used to predict if red team won to see if one side yields a more predictable outcome.

The plots will help visualize and demonstrate a relationship (if any exists) between being able to predict the outcome of a match with which team a player is on.