

Abstract

Table of Contents

List of Figures

List of Tables

Introduction and the Objectives

League of Legends is a multiplayer online battle arena video game developed and published by Riot Games for Microsoft Windows and macOS. In this game, players assume the role of a champion with unique abilities and battle against a team of other player or computer controlled champions. The goal is usually to destroy the opposing team’s “Nexus”, a structure that lies protected by defensive structures. This dataset contains the first 10 minutes’ statistics of approximately 10000 ranked games. Players in the dataset are roughly from the same level. There are 19 features per team (38 in total) collected after 10 minutes in game. It will be an interesting task to predict the winner at an early stage as well as to analyze the game strategy.

Hence, the proposed objectives for this project will be,

1. To develop a predictive model to predict the winner of the game based on the facts from the first 10 minutes of the game. This could be used to give an early warning to the players so that they can make an attempt to change their strategies if their probability of winning is low.

2. To find the factors which have a higher impact on determining the victory of the game. These findings will be helpful for beginners and future players to play a good game.

About the dataset

**No. of observations:** 9879

**No. of variables:** 40

|  |  |  |
| --- | --- | --- |
| No. | Variable Name | Description |
|  | gameID | Unique RIOT ID of the game |
|  | BlueWins | 1 if the blue team has won, 0 otherwise |
|  | xWardsPlaced | No. of Warding totems placed by the team x on the map |
|  | xWardsDestroyed | Number of enemy warding totems the team x has destroyed |
|  | xFirstBlood | First kill of the game. 1 if the team x did the first kill, 0 otherwise |
|  | xKills | No. of enemies killed by the team x |
|  | xDeaths | No. of deaths of the team x |
|  | xAssists | No. of kill assists by the team x |
|  | xEliteMonsters | No. of elite monsters killed by the team x (Dragons and Harolds) |
|  | xDragons | No. of dragons killed by the team x |
|  | xHeralds | No. of heralds killed by the team x |
|  | xTowersDestroyed | No. of structures destroyed by the team x |
|  | xTotalGold | x team total gold |
|  | xAvgLevel | x team average champion level |
|  | xTotalExperience | x team total experience |
|  | xTotatlMinionsKilled | x team total minions killed |
|  | xTotalJungleMinionsKilled | x team total jungle minions killed |
|  | xGoldDiff | x team gold difference compared to the enemy team |
|  | xExperienceDiff | x team experience compared to the enemy team |
|  | xCSPerMin | x team CS (minions) per minute |
|  | xGoldPerMin | x team gold per minute |

Where x = Red, Blue. That is, same set of variables are repeated for the two teams.

Important Results of the Descriptive Analysis

1. There is no imbalance in the response variable “blueWins”
2. Winning team seems to have the higher total experience
3. Total Gold is High in the Winning team
4. The winning team seems to have done more number of kills
5. There are correlations present among several predictor variables

Suggested Models for the Advanced Analysis

1. Binary Logistic Regression
2. Logistic Regression with Ridge Penalty
3. Logistic Regression with Lasso Penalty
4. Logistic Regression with Elastic Net Penalty
5. Support Vector Machine
6. Random Forest
7. AdaBoost Classifier
8. XG Boost

Important Results of the Advanced Analysis

Summary of Models fitted using all variables.

|  |  |
| --- | --- |
| Model | Accuracy |
| Binary Logistic Regression | 0.721154 |
| Logistic Regression with Ridge Penalty | 0.722165 |
| Logistic Regression with Lasso Penalty | 0.717105 |
| Logistic Regression with Elastic Net Penalty | 0.721153 |
| Support Vector Machine | 0.717105 |
| Random Forest | 0.715081 |
| Ada Boost Classifier | 0.712551 |
| XG Boost Classifier | 0.693826 |

Confusion matrix for Logistic Regression with all variables.

|  |  |  |  |
| --- | --- | --- | --- |
|  | No | Yes | Class wise error rate |
| No | 730 | 269 | 0.2693 |
| Yes | 282 | 695 | 0.2886 |

Summary of Models fitted using 10 most important variables.

|  |  |
| --- | --- |
| Model | Accuracy |
| Binary Logistic Regression | 0.718117 |
| Logistic Regression with Ridge Penalty | 0.716093 |
| Logistic Regression with Lasso Penalty | 0.717105 |
| Logistic Regression with Elastic Net Penalty | 0.716599 |
| Support Vector Machine | 0.713563 |
| Random Forest | 0.697874 |
| Ada Boost Classifier | 0.709514 |
| XG Boost Classifier | 0.685729 |

Confusion matrix for Logistic Regression with 10 most important variables.

|  |  |  |  |
| --- | --- | --- | --- |
|  | No | Yes | Class wise error rate |
| No | 723 | 276 | 0.2763 |
| Yes | 281 | 696 | 0.2876 |

Summary of Models fitted using 4 most important variables & interaction terms.

|  |  |
| --- | --- |
| Model | Accuracy |
| Binary Logistic Regression | 0.716559 |
| Logistic Regression with Ridge Penalty | 0.715587 |
| Logistic Regression with Lasso Penalty | 0.715587 |
| Logistic Regression with Elastic Net Penalty | 0.715587 |
| Support Vector Machine | 0.715587 |
| Random Forest | 0.681680 |
| Ada Boost Classifier | 0.716093 |
| XG Boost Classifier | 0.707490 |

Confusion matrix for Logistic Regression with 4 most important variables & interaction terms.

|  |  |  |  |
| --- | --- | --- | --- |
|  | No | Yes | Class wise error rate |
| No | 727 | 272 | 0.2723 |
| Yes | 288 | 689 | 0.2948 |

Issues Encountered and Proposed Solutions

Discussion and Conclusion

Appendix