**League of Legends - Team Project**

Joey Gladys, [jgladys@bellarmine.edu](mailto:jgladys@bellarmine.edu)

Connor Goodman, [cgoodman2@bellarmine.edu](mailto:cgoodman2@bellarmine.edu)

**ABSTRACT**

For our project we decided to take a data set from the League of Legends Worlds event. Each player within the Worlds competition was tracked on how and what went on within the games. Each player had a different number of points, kills, deaths, assists, and more data to be able to track and we decided it would be cool to use machine learning to predict how many kills each player would get based off their stats in each game. This could allow players and coaches to draft players on how well they could have done in their games like those games that are tracked within the data set. The object of any sport or game is to win and having a kill amount within these games gets the team further to that objective of winning. Using machine learning predictability, we can see who has a higher chance of winning!

1. **INTRODUCTION**

The League of Legends Worlds 2021 data set can be found on Kaggle. This data set has tracked every professional League player’s stat in the competition event. In this project we decided to use “LR, SVR, DTR, and RFT”. We can use all 4 and compare results to see which are the most accurate.

1. **BACKGROUND**
   1. *Data Set Description*

Professional League player stats are tracked. Those stats include kills, deaths, assists, creep score, gold earned, champion damage, kill participation, wards placed, wards destroyed, dragons, and finally barons taken. Even if you are no League admirer and know nothing about the game this data set allows you to see who does what within the games. Players with higher stats have a much higher chance of winning and coaches and players can notice this data and see who has the better chances of winning.

* 1. *Machine Learning Model*

The machine learning models we used are Linear Regression, Support Vector Regression, Decision Tree Regression, and Random Forest Regression. One of the most accurate was Decision Tree Regression and we would recommend using that for tracking stats accurately within the doc when making an analysis. All 4 models were taken and used within our project so new viewers can see all the predictions and calculations and decide on which they prefer.

1. **EXPLORATORY ANALYSIS**

This data set contains 215 samples with 20 columns. Six columns are objects, 12 columns are integer data, and two columns are float data. There is not a single missing data value. Anything unusual would be string data since we did not use player names in the data.

**Table 1: Data Types**

|  |  |
| --- | --- |
| *Variable Name* | *Data Type* |
| Deaths | int |
| Assists | int |
| Creep Score | int |
| Gold Earned | int |
| Champion Damage Share | float |
| Kill Participation | float |
| Wards Placed | int |
| Wards Destroyed | int |
| Dragons For | int |
| Barons For | int |

1. **METHODS**
   1. *Data Preparation*

Due to their being no missing data in the null check, we did not have to fill in any missing data or use an imputer. We had string columns that did not contribute to our predictions, so we did not include them in our X-axis. The columns could have been dropped but were kept in the dataset.

* 1. *Experimental Design - Attached*

Table X: Experiment Parameters

|  |  |
| --- | --- |
| **Table:** | Table is attached in GitHub as an excel file named: Experiment Parameters! |

* 1. *Tools Used*

The following tools were used for this analysis: Python running the Anaconda environment for Apple Macintosh computer was used for all analysis and implementation. In addition to base Python, the following libraries were also used: Pandas 0.18.1, Numpy 1.11.3, Matplotlib 1.5.3, Seaborn 0.7.1, SKLearn 0.18.1,. Using these tools allows us to run our code and find the expected outcomes based on the data inputted from the League of Legends Worlds dataset.

1. **RESULTS**
   1. *Mean square Error and R-Square calculation*

SVR- R-Square: 0.9343

MSE: 1.5096

DTR- R-Square: 0.9343

MSE: 2.4824

RF- R-Square: 0.9343

MSE: 0.7022

LR- R-Square: 0.8139

MSE: 1.1820

* 1. *Discussion of Results*

The DTR gave us the most concise answers since they were whole numbers. The worst model was Linear Regression. The accuracy was poor since the R-squared value was farthest from 1.0. RFT was another model that was poor due to the same reasons as the Linear Regression model.

* 1. *Problems Encountered*

The length of the dataset brought on complications due to confusion with the number of answers. Many of the string columns also provided little to no help contributing to predictions.

* 1. *Limitations of Implementation*

Predictions are limited to the data, and anything could happen within the game. The use of this data is strictly for prediction purposes. The DTR model worked the best in giving us these predictions.

* 1. *Improvements/Future Work*

One improvement we could make is to use more in-game statistics in this project. This would allow us to find even more detailed answers and outcomes. Predictions would be even more precise. Adding these variables would allow for expected outcomes to be very close to the actual outcomes.

1. **CONCLUSION**

For this project we are happy with the results and predictions of the dataset and machine learning predictability. DTR and SVR gave well received and concise predictions based on the stats tracked within the data set. This example of machine learning can be applied to players and coaches when interpreting the data within the League worlds event. Overall, the project was smooth and with little troubleshooting the code was clear and working properly. We will use this data set for all League Worlds predictions in the future although more data can be applied.

**REFERENCES**

<https://www.kaggle.com/datasets/braydenrogowski/league-of-legends-worlds-2021-playin-group-stats>

League of Legends Worlds 2021 Play-In Group Stats

Game stats for all matches in the League of Legends Worlds 2021 Play-in Groups