Players' win rate predictor with multiple linear regression

Ana Dueñas Chávez – A01702080

Abstract – This document presents the implementation and explanation of a linear regression using sklearn framework to determine the win rate of League of Legends professional players.

I. INTRODUCTION

League of legends is one of the most played videogames all around the world, in which two teams of five members each try to destroy the rivals' nexus. Because of its continuous improvements and game mechanics, it has become one of most popular esports. League of legends develop various tournaments that take place worldwide. Because of it, before the start of a new season[1] every region starts building the best team they can to win the biggest tournament called Worlds. This competition team is built, mainly, based on the abilities of each person playing the game.

II. DATASET

The dataset used to develop this project was obtained from Oracle Elixir[2]. This compilation of data collected is build by statistics about every participant of each team in the main tournament.

It was split by year but the seasons from 2017 to 2020 are being merged to create a bigger dataset with about 357 records. Which include measurements taken from all the games played in the main tournament, Worlds.

Player	Team	Pos	GP	W%	CTR%	K	D	A	KDA
957	Team WE	Top	15	60%	60%	31	23	59	3.9
Albis	ahq eSports Club	Support	6	33%	-	0	7	26	3.7
Alphari	Misfits Gaming	Top	12	50%	60%	15	23	57	3.1
Ambition	Samsung Galaxy	Jungle	15	80%	0%	20	29	103	4.2
An	ahq eSports Club	ADC	6	33%	-	15	13	10	1.9
Archie	GAM Esports	Top	7	29%	-	6	35	34	1.1
Bang	SK Telecom T1	ADC	19	58%	60%	55	26	63	4.5
Bdd	Longzhu Gaming	Middle	9	67%	67%	31	12	55	7.2
Ben	Team WE	Support	15	60%	40%	7	21	107	5.4
Betty	Flash Wolves	ADC	6	17%	-	9	10	23	3.2
Biofrost	Team SoloMid	Support	7	43%	-	2	11	38	3.6

Figure 1 Portion of the dataset taken to do the project.

[1] Period in which a tournament take place, usually from February to November.

[2] https://oracleselixir.com/

III. APPROACH

The purpose of this project is to determine a player win rate based on the statistics about the performance.

This prediction can be performed by a linear regression to approximate the relationship between the selected data and the win rate.

In order to achieve the objective, some dataset fields that had no relation with the win rate were taken away. To do this selection, a correlation was applied to the data in order to see the relationship between each field (Figure 2). The attributes that were chosen were the ones that appear with a value equal or higher than 0.4 on the table and the ones that were not contain in other fields one, for example instead of using the field “Kills”, “Assist” and “Deaths”, the KDA[3] field was used.

The fields with the higher relation to the win condition was number of games played, gold, experience and the average creep score difference after 10 minutes of the game.

[3] Kills, assist and death relation provided by the dataset.

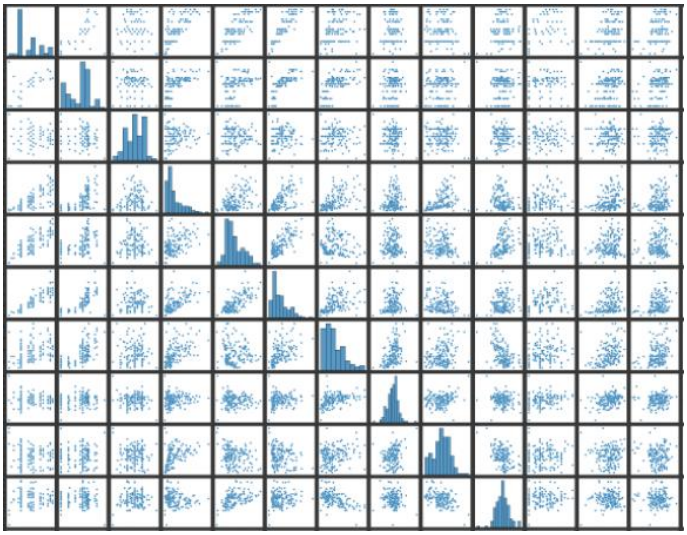

Figure 2 Graphical correlation between data fields.

	GP	W%	CTR%	K	D	A	KDA
GP	1.000000	0.609635	0.065749	0.726630	0.759542	0.876922	0.398879
W%	0.609635	1.000000	-0.006249	0.519342	0.352710	0.658369	0.656352
CTR%	0.065749	-0.006249	1.000000	0.052969	0.040655	0.036419	-0.013000
K	0.726630	0.519342	0.052969	1.000000	0.500416	0.532279	0.460779
D	0.759542	0.352710	0.040655	0.500416	1.000000	0.710412	-0.114674
A	0.876922	0.658369	0.036419	0.532279	0.710412	1.000000	0.469930
KDA	0.398879	0.656352	-0.013000	0.460779	-0.114674	0.469930	1.000000
KP	0.008060	0.196947	-0.040151	0.002601	-0.219589	0.122401	0.424269
KS%	-0.070123	-0.046604	0.053606	0.424362	-0.185986	-0.277905	0.112021
DTH%	0.043032	0.012142	0.027619	-0.145011	0.441480	0.088658	-0.519833
FB%	0.042188	0.109084	-0.013187	0.069840	-0.029236	0.108408	0.198980
GD10	0.255674	0.318837	0.003799	0.273082	0.115344	0.258939	0.327661
XPD10	0.290379	0.269050	-0.020349	0.252483	0.149465	0.267720	0.277642

Figure 3 Correlation between fields of the dataset.

IV. LINEAR REGRESION

To the develop the project the input data was divided in 67% for training and 33% for testing.

A multiple linear regression was applied to calculate the win rate, using the sklearn linear model.

V. RESULTS

Variance was used to measure how different were the observed values form the average of the predicted values, it had a value of 59%.

Since the dataset was limited, the cross-validation score was needs to estimate the skill of the model predicting the test data, it had a value of 53%.

Additionally, the mean squared error was 0.16 which means that model predictions were close to the real win rate calculation.

VI. CONCLUSION

Based on the results obtain the aim of the project was achieved. Even though, the cross validation and variance was a little bit low the model can successfully predict win rates based on the game statistics.

VII. REFERENCES

[1] BCM, "bcm blogs," 7 May 2018. [Online]. Available: https://www.bmc.com/blogs/mean-squared-error-r2-and-variance-in-regression-analysis/#:~:text=What%20is%20variance%3F,a%20value%20that%20is%20low.. [Accessed 11 November 2020].

[2] J. Brownlee, "Machine Learning Mastery," 23 May 2018. [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/#:~:text=Cross%2Dvalidation%20is%20a%20resampling,k%2Dfold%20cross%2Dvalidation.. [Accessed 11 November 2020].

[3] J. Frost, "Stadistics by Jim," 1 January 2018. [Online]. Available: https://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/#:~:text=The%20sign%20of%20a%20regression,variable%20also%20tends%20to%20increase.. [Accessed 10 November 2020].