Using Data Mining to Predict the Outcome of a League of Legends Professional Match

Jia Ming Chen Chen
The Lesley H. and William L. Collins College of Professional Studies
St. John's University
CUS610 Data Mining Research Paper

***Abstract*—This research will use data mining to analyze win probabilities of a League of Legends professional match according to the champions played by each team. It will also go a step further and try to target a specific team and find out if it is possible to make a strategy against a team based on findings of data mining analysis.**

Keywords*—*data mining, league of legends, data analytics, prediction

I. Introduction

Esports have been in a constant accelerated growth in the last decade with more than 470 million viewers[a] around the world and a market revenue of over 1 billion U.S. dollars according to Statista. These figures are projected to increase to 570 million viewers and 1.6 billion U.S. dollars by 2024[b]. This goes to show how serious esports is as a competition and an industry. As players and staff look to improve as much as they can, they try to get an edge in the game everywhere they can get it not only in training, but top tier teams even have therapists and nutritionists. In this research, we will look at the possibility of leveraging data mining to analyze professional League of Legends matches and use it to get some sort of informational advantage.

II. Literature Review

Data Mining has been used in the esports industry, but it is still new and has not been used widely to analyze the information in front of us.

A. A Data Mining Study on League of Legends [1]

In A Data Mining Study on League of Legends, the researchers focus on in-game interactions, choices, and events to predict the outcome of the game. However, there is already information being analyzed when teams are choosing their champions. In the "champion select", teams take turns to ban 5 champions and to pick their 5 champions. This gives a lot of room for analysis when choosing which champions are better to be picked in the first turns and which champions are better in the latter turns when you have an idea of how both teams are looking.

B. Profiling Successful Team Behaviors in League of Legends [2]

This article takes a look at in-game statistics such as kills and assists and creates performance metrics based on how the statistics the winning team had. Then, the authors utilized clustering to group teams according to the win rate certain statistics had. The study takes an objective look at the game by analyzing the statistics of matches and tells us which are more important and had better results in making a team win. However, it does not take into account the skill level of the players as the data was gathered from random League of Legends player's match history without filtering their rank.

C. Data Mining for Item Recommendation in MOBA Games [3]

In this study, the authors tackle the problem of item recommendation and create a solution to recommend the players items based on the context of the game such as champions played. The same model could be used for champion recommendation based on previous champion picks.

D. Logic Mining in League of Legends [4]

Logic Mining in League of Legends shows a great example on leveraging data mining to analyze information and target it in the competitive scene. The article analyzes different in-game statistics such as which team gets the first kill, first dragon, etc., and uses is to find out which were the most relevant for deciding the outcome of the game for different regions of professional League of Legends.

E. Champion Recommender System For League of Legends [5]

In this paper, the authors create a recommendation system that follows the ban and pick order when selecting champions to maximize the probability that a team will win. The system takes into account champion popularity and interaction with other ally and enemy champions (synergy and counter)

[a] Gough, Christina. "eSports audience size worldwide from 2019 to 2024, by type of viewers." Statista, 2021, 1st June 2021, https://www.statista.com/statistics/490480/global-esports-audience-size-viewer-type/

[b] Gough, Christina. "eSports market revenue worldwide from 2019 to 2024." Statista, 2021, 6th Aug 2021, https://www.statista.com/statistics/490522/global-esports-market-revenue/

F. *Using Machine Learning to Predict Game Outcomes Based on Player-Champion Experience in League of Legends [6]*

In this article, the authors research a system to predict a game outcome based on player skill calculated from the amount of games per season, win rate, and champion mastery which is an in-game metric of number of matches with a specific champion plus game scores.

III. TERMS

A. *Champions*

This is the name of the playable characters in League of Legends

B. *Top, Jungle, Mid, Bot, Adc/Marksman, Support*

These refer to the in-game roles and positions. Top, jungle, and mid refer to their specific position in-game. Marksman and support are both played in the bot lane together. Marksman and Adc (Attack Damage Carry) are interchangeable.

C. *MonkeyKing*

Refers to the champion Wukong

IV. DATA

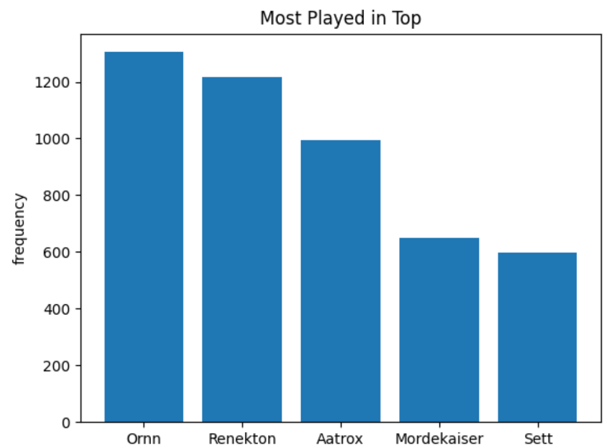
For our study we used a database [7] that included professional League of Legends matches from 2020 with information about the league where the match was played, the teams playing, the champions played, and the result of the game.

In order to preprocess the data, multiple datasets were created by deleting and merging different columns. We used 2 datasets to analyze, one where we merged a specific role disregarding the team for an overall analysis of the roles, and a dataset targeting a specific team's matches. The league and id row were deleted as they did not fit in our studies. To keep the analysis short, we only looked at the "top" role and the "bot" role which includes marksmen and supports. However, the following analysis can be applied to every role in the game. For our target team, we chose DAMWON Gaming as the champions of the world tournament that year. Both datasets were further preprocessed to include only matches where the champions were played enough times to be of significance.

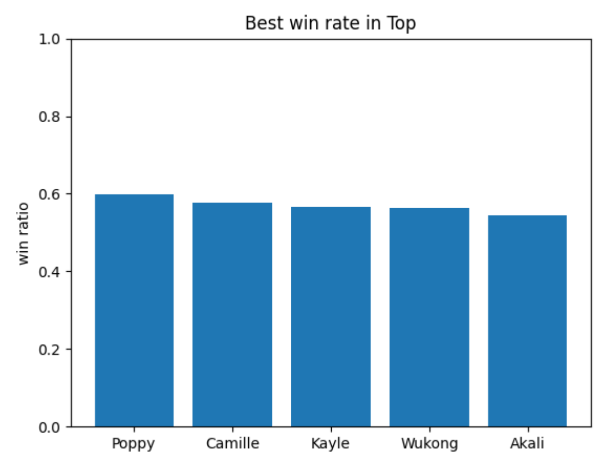
V. METHODOLOGY

We applied the same analysis to both datasets we were using. Since the champions were organized as categorical data under their respective role column, we grouped it and made basic analysis of how popular and effective they were. Note that the original dataset had 5612 entries and ended with 8878 after merging roles and preprocessing. We fist gathered the data for top lane most picked and best champion according to win rate. Then, we created a heat map to showcase bot lane duos with their respective win rate. Lastly, we used the apriori algorithm [8] to find rules of bot lane matchups (blue team versus red team)

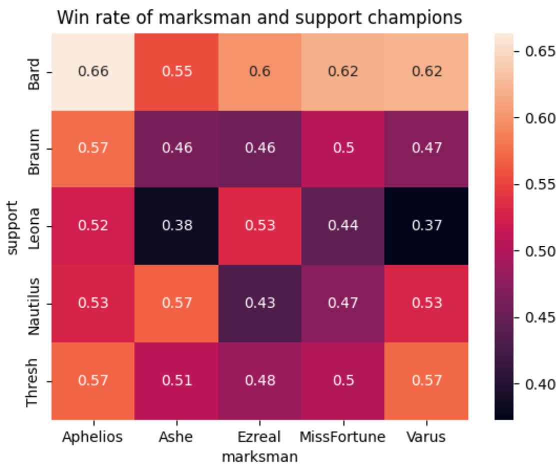
VI. RESULTS

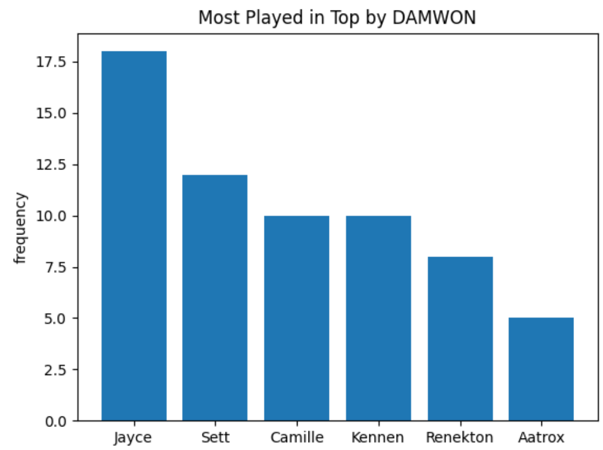


In the previous graph, we can see the most popular champions to be picked for the top role. These are champions that were good at the time, or that were good blind picks as in when you don't know who your opponent will be.



This time, we see the champions in the top lane that had the best win rates with at least 50 matches. We see that these champions were not necessarily the most popular, although Wukong and Camille were picked 561 and 450 respectively. On the other hand, Poppy was picked only 102 times meaning that although she is not as popular, when picked in the right scenarios, she was the most effective champion to win the game.





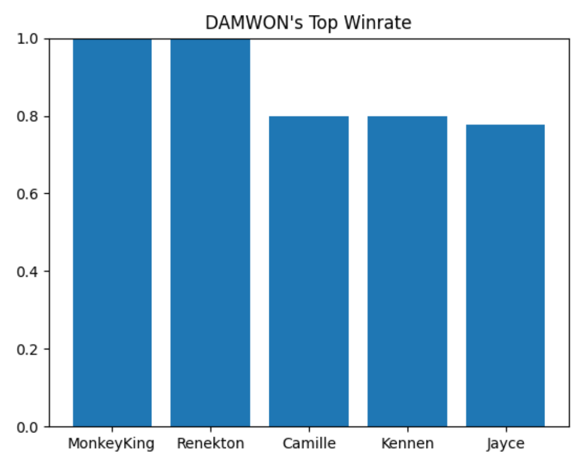
In this graph, we see a heat map of bot lane combinations (marksman and support) and what are the win rates of some popular duos. Because of the nature of marksmen and supports going to the same position, both picks heavily rely on synergizing with the other.

Lastly, we used the apriori algorithm to find rules of bot lanes matching against each other, and the KNN algorithm to predict a match outcome based on champions in each team. Note that 1 means blue team win and 0 means red side win, and that a support of 0.002 in this dataset means that the rule had a support count of 9. We see that this can be effective to known which specific champions are better against others while the KNN model could be effectively used for match prediction.

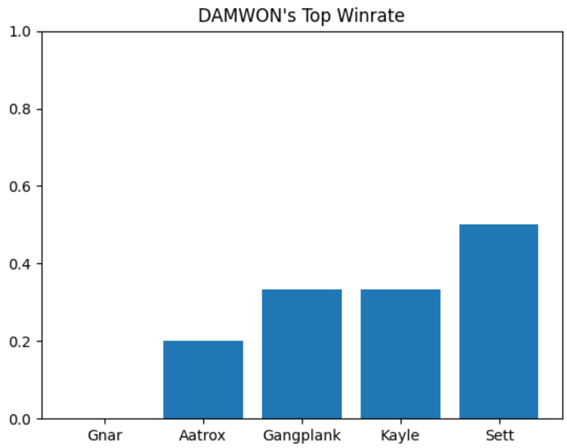
```
{blueadc_Aphelios, bluesup_Braum, redadc_MissFortune, redsup_Leona} -> {1} (conf: 0.917, supp: 0.002, lif
{blueadc_MissFortune, bluesup_Nautilus, redadc_Aphelios, redsup_Thresh} -> {0} (conf: 0.750, supp: 0.003,
{blueadc_Aphelios, bluesup_Nautilus, redadc_MissFortune, redsup_Leona} -> {1} (conf: 0.812, supp: 0.003,
{blueadc_Ezreal, bluesup_Yuumi, redadc_Aphelios, redsup_Thresh} -> {0} (conf: 0.688, supp: 0.002, lift: 1
{blueadc_Aphelios, bluesup_Leona, redadc_MissFortune, redsup_Nautilus} -> {1} (conf: 0.769, supp: 0.002,
{blueadc_Aphelios, bluesup_Nautilus, redadc_Ashe, redsup_Thresh} -> {1} (conf: 0.706, supp: 0.003, lift:
{blueadc_Aphelios, bluesup_Nautilus, redadc_MissFortune, redsup_Thresh} -> {1} (conf: 0.688, supp: 0.002,
{blueadc_Aphelios, bluesup_Nautilus, redadc_Ezreal, redsup_Braum} -> {1} (conf: 0.667, supp: 0.002, lift:
{blueadc_Aphelios, bluesup_Braum, redadc_MissFortune, redsup_Nautilus} -> {1} (conf: 0.647, supp: 0.002,
{blueadc_Aphelios, bluesup_Thresh, redadc_Ezreal, redsup_Nautilus} -> {1} (conf: 0.625, supp: 0.002, lift
{blueadc_Aphelios, bluesup_Thresh, redadc_MissFortune, redsup_Nautilus} -> {1} (conf: 0.567, supp: 0.004,
K-nearest neighbors classifier
Accuracy KNN:  0.8463180362860192
              precision    recall  f1-score   support

           0       0.81      0.86      0.84       429
           1       0.88      0.83      0.85       508

    accuracy                           0.85       937
   macro avg       0.85      0.85      0.85       937
weighted avg       0.85      0.85      0.85       937
```

For analyzing DAMWON, we tried to keep the same methodology. However, the match count of a single team is highly reduced with DAMWON having played only 96 matches even though they were the champions and playing 21 different top champions. Because of this, different filter numbers were used such as champions being played at least 3 times.

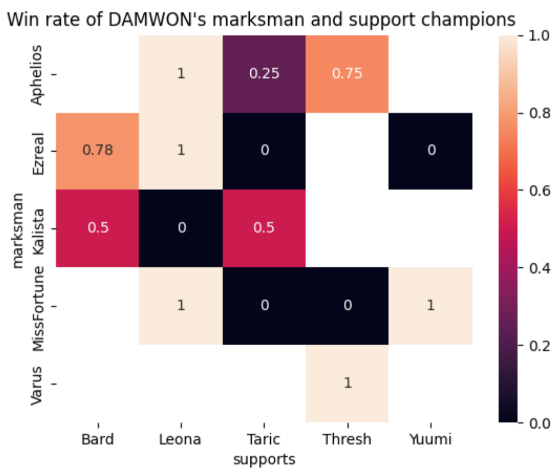
We can see that in the top lane DAMWON has a clear preference to pick Jayce with whom they also have a high win rate letting us know that a good idea against this team would be to ban Jayce.



On the other side of the spectrum, DAMWON does not usually pick champions with which they have a bad win rate. However, with this information a team could strategize and force them into picking one of these champions with specific bans. For example, their second most picked champion Sett is also their fifth worst win rate with around 50%. Note that there were only 12 different champions that DAMWON played in at least 3 games. This analysis could also be used to find which champions DAMWON struggles the most against.



Unfortunately, heat map was not useful as with the low count of matches, most bot lane combinations were not played enough to be significant.



Likewise, applying the apriori algorithm on the bot lane was not significant as with a count of 51 matches, there was not enough data to cover the amount of possible marksman and support combinations. With a support of 0.073, the highest support rule had a match count of 4.

```
{Bard, Ezreal} -> {1} (conf: 0.778, supp: 0.073, l
{Thresh, Varus} -> {1} (conf: 1.000, supp: 0.052,
{Maokai, Senna} -> {1} (conf: 1.000, supp: 0.042,
{Aphelios, Taric} -> {0} (conf: 0.750, supp: 0.031
```

VII. Conclusion

Data mining did not seem to be useful when used to analyze champions against specific team in League of Legends because of the high number of combinations between champions possible and the reduced number of games a single team has. However, it can still be used on a more general perspective to analyze professional matches as a whole instead of targeting a team giving us an idea of what champions are trending and which have been more effective in the competitive scene in League of Legends.

VIII. Limitations and Future Work

Data mining did not work when focusing the champions played by and against a specific team because of the low number of matches and the high amount of champion combinations. However, it could still be used to analyze a more generalized characteristic of the game that has less categorical values such as in-game statistics or to use in a more general dataset such as all professional matches, professional matches of a region, or matches played non-professionally through the game client.

LIBRARIES USED

- Pandas dataframe for data manipulation. McKinney, Wes. "Data structures for statistical computing in python." Proceedings of the 9th Python in Science Conference. Vol. 445. 2010. https://zenodo.org/record/5774815#.Yb6bzGiZOUk
- Sci-kit learn for KNN neighbor algorithm. Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830.
- Seaborn for heatmap graph. Waskom, Michael L. "Seaborn: statistical data visualization." Journal of Open Source Software 6.60 (2021): 3021.
- Matplotlib for statistical graphing. Hunter, John D. "Matplotlib: A 2D graphics environment." Computing in science & engineering 9.03 (2007): 90-95.

REFERENCES

[1] E. Tanyıldızı, G. Yıldırım and O. C. Kaya, "A Data Mining Study on League of Legends," 2019 1st International Informatics and Software Engineering Conference (UBMYK), 2019, pp. 1-4, doi: 10.1109/UBMYK48245.2019.8965530.

[2] Jr, Fernando & Melo, Allan & Costa, Igor & Marinho, Leandr. “Profiling Successful Team Behaviors in League of Legends.” 261-268. 2017 10.1145/3126858.3126886.

[3] Parra, Denis & Rios, Felipe & Araujo, Vladimir. (2019). Data Mining for Item Recommendation in MOBA Games. 10.1145/3298689.3346986.

[4] Kho, Liew Ching, et al. "Logic Mining in League of Legends." Pertanika Journal of Science & Technology 28.1 2020.

[5] Kim, Young Seok, Pradeep Prabhakar, and Aniruddha Dutta. "Champion Recommender System For League of Legends.”

[6] Do, Tiffany D., et al. "Using Machine Learning to Predict Game Outcomes Based on Player-Champion Experience in League of Legends." The 16th International Conference on the Foundations of Digital Games (FDG) 2021. 2021.

[7] Ahmed Morra, “League of Legends 2020 Esports Match Data.” https://www.kaggle.com/xmorra/lol2020esports

[8] Tommyod, “Efficient-Apriori” https://github.com/tommyod/Efficient-Apriori