Business Analytics 2020 Fall

To the Top of League of Legend

--- LOL Match Prediction and Champion Clustering

Annan Chen ac4619
Jiepeng Lian jl5521
Jing Xiao jx2422
Qingyuan Ren qr2130
Zezuan Zhang zz2705

COLUMBIA UNIVERSITY

Contents

Introduction

COLUMBIA UNIVERSITY

About League of Legend

League of Legends is a PC-side MOBA game popular all over the world. Players can accumulate levels and money by killing monsters, minions and opponent heroes in the game, and finally winning the game by destroying the enemy base.

About LOL Business Market

There are mass user base and rich business chains, including worldwide professional competitions and regional leagues (LCS, LEC, LCK, LPL, etc). The winners of the competition could receive up to millions of dollar award. Various betting companies or live stream activities closely following the detailed results of each competition. The outstanding team and players could attract various advertisement endorsement and commercial activities.

About our research plan

Originally all 152 champions are in 6 categories: Tank, Mage, Marksman, Assassin, Fighter and Support. In this project, we want to dig deeper into clustering champions based on their performance in a professional match, and provide valuable insights by suggesting champions selection and predicting games results at an early stage.

Data Description and Preprocessing

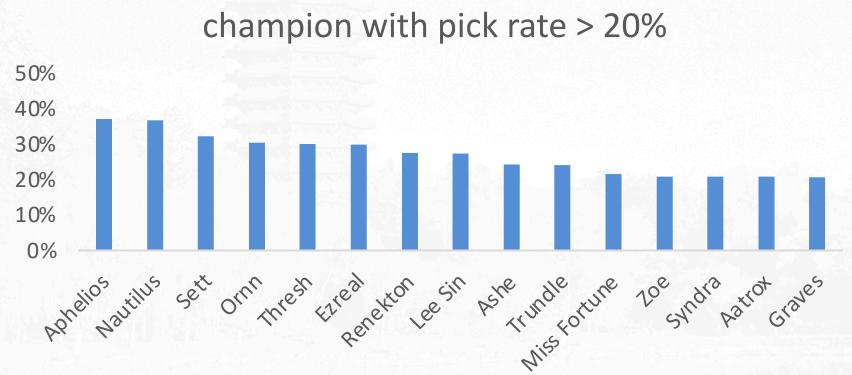
The datasets (https://oracleselixir.com/tools/downloads) we used include in-game features for all professional games during 2014 to 2020. More specifically, we pick all professional match data in 2020 because of timeliness - the game and champions might change after updates and new versions.

This dataset contains 77424 rows with 105 columns. Each 12 rows indicates 1 match (10 player-level data & 2 team-level data). We also do several data preprocessing steps to make the data more structural as desired (separating team level and champions level data, removing collinear and non-numerical features, combining each two team-level observations to one row . After data preprocessing we come to 6369 rows with 102 columns. We would use team-level dataset for predicting results, and use champion-level dataset for clustering.

<table>
<tr><th colspan="2">Observations N = 77,424</th><th>Features (k=105)</th></tr>
<tr><td rowspan="3">6,452 games in 2020 professional matches</td><td>Individual level: Blue sides players (5 in each game, N = 32,260)</td><td rowspan="2">Champion selected, game result
kills/deaths/assists, gold difference, experience, creep score, vision score, etc. ……</td></tr>
<tr><td>Individual level: Red sides players (5 in each game, N = 32,260)</td></tr>
<tr><td>Team level integrated performance (2 in each game, N = 12,904)</td><td>5 banned champions, game result
team kills/deaths/assist, gold, experience, etc. ……</td></tr>
</table>

Data Description and Preprocessing

Figure: Champion pick rate (Who is popular?)



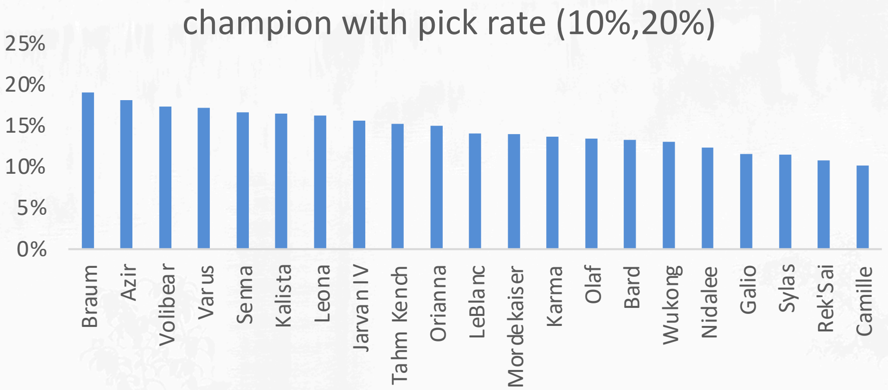
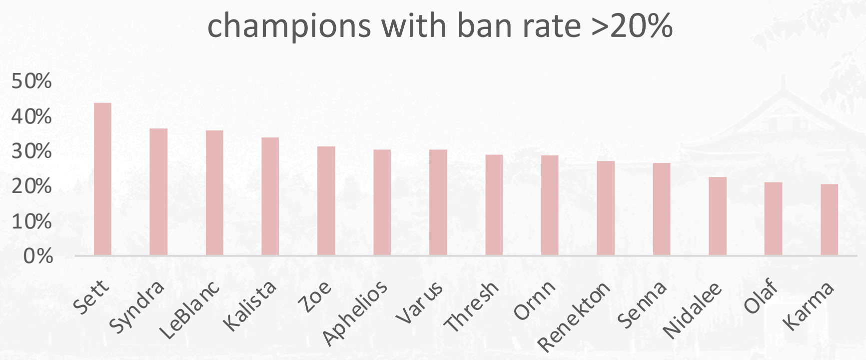
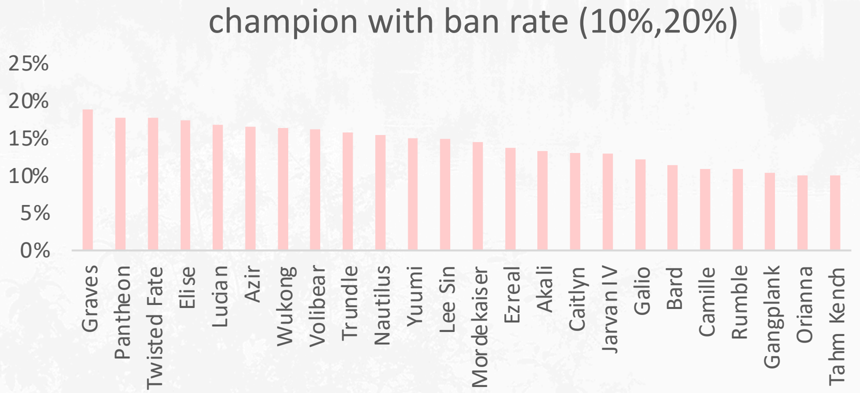


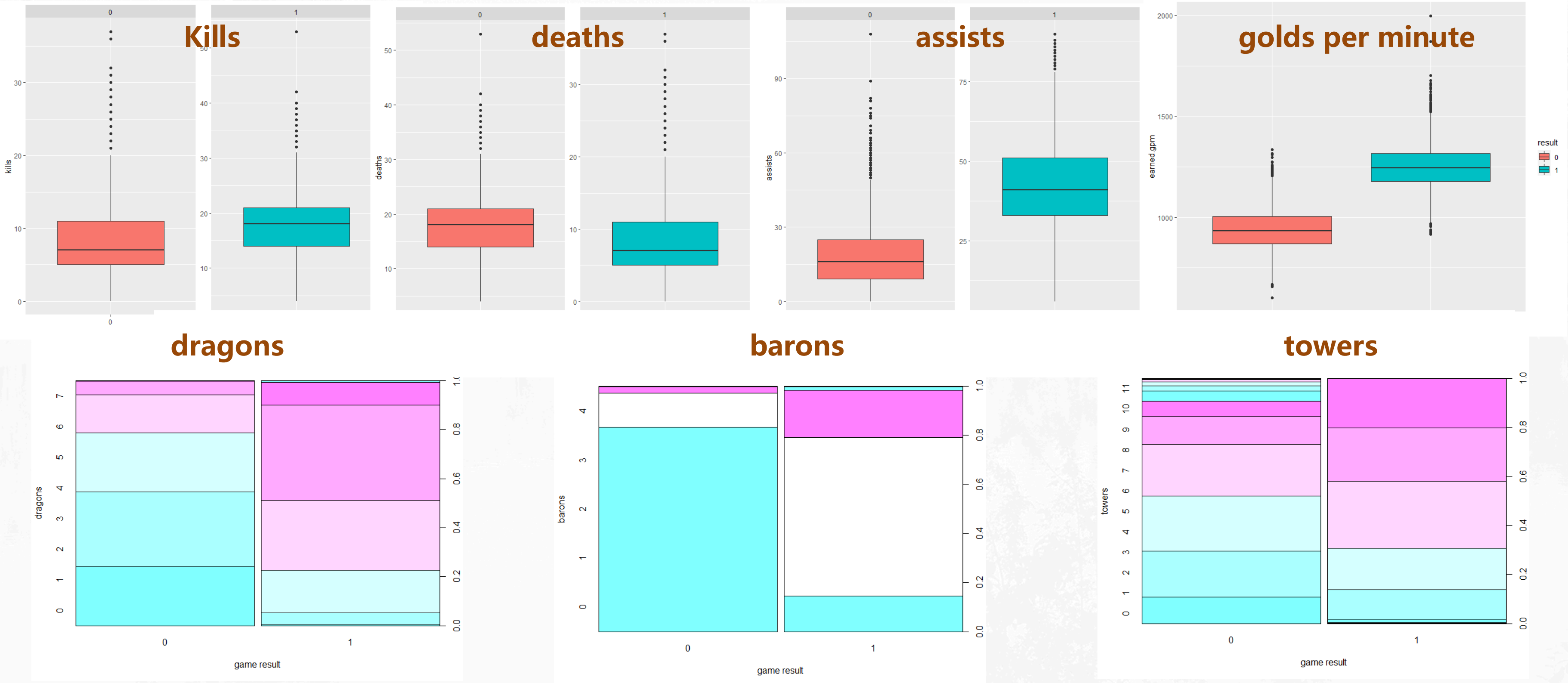
Figure: Champion ban rate (Who is threatening?)





Data Description and Preprocessing

Figure: The overall performance difference between win and loss team



Champions Clustering

Purpose: Using player level data, clustering the champions for further prediction.

Step 1 Aggregate mean performance of players for each champion

Grouping by champion and getting the mean of every feature, we got the 54 features for each champion.

Step 2 Calculating mean "win" and "loss" performance for each champion

Split the whole data into the "win" and "loss" part to see how these champions performed in win or loss games. Thus, we got total 108(54 x 2) features for each champion and did PCA based on this dataset.

Step 3 Decomposing the features using the PCA method

Letting the cumulative proportion of variance explained go over 90%, we chose 4 principle components, and the final cumulative var explained is 94.05%. We have loadings of the four factors as shown in next page.

Champions Clustering

Result: Four feature factors

damage		Economy-experience		Win-loss difference		enemy	
damagetochampions	0.653	damagetochampions_	0.506	damagetochampions	0.532	earnedgold_	0.527
damagetochampions_	0.521	earnedgold	0.325	damagetochampions_	0.465	earnedgold	0.486
totalgold	0.188	totalgold	0.295	totalgold_	0.311	xpat15	0.227
goldspent	0.146	opp_xpat15	0.258	goldspent_	0.300	xpat15_	0.227
totalgold_	0.142	xpat15_	0.256	earnedgold_	0.295	opp_goldat15	0.224
opp_goldat15_	0.142	xpat15	0.255	opp_xpat15	0.202	opp_goldat15_	0.213
opp_xpat15_	0.141	opp_xpat15_	0.243	opp_xpat15_	0.185	opp_goldat10	0.178
opp_goldat15	0.130	damagetochampions	0.238	xpat15	0.145	opp_xpat15	0.170
opp_xpat15	0.130	goldspent	0.233	opp_xpat10	0.133	opp_xpat15_	0.159
goldspent_	0.129	opp_xpat10	0.191	xpat15_	0.133	opp_goldat10_	0.148

Results: Clustering result when k = 20

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
'Senna' 'Twitch' 'Jinx' 'Taliyah' 'Vayne'	'Karma' 'Lux' 'Nunu & Willump' 'Volibear' 'Xin Zhao' 'Bard' 'Blitzcrank' 'Morgana' 'Soraka'	'Fiora' 'Irelia' 'Twisted Fate' 'Jax' 'Riven' 'Tristana' 'Tryndamere' 'Camille' 'Aatrox'	'Jayce' 'Quinn' 'Corki' 'Zoe' 'Vladimir' 'Syndra' 'Rumble' 'LeBlanc' "Vel'Koz" 'Viktor' 'Swain'	'Thresh' 'Trundle' 'Skarner' 'Lulu' 'Jarvan IV' 'Taric' 'Leona' 'Tahm Kench'	'Sejuani' 'Pyke' 'Lee Sin' "Rek'Sai"	'Lucian' 'Dr. Mundo' 'Talon' 'Akali' 'Ekko' 'Aurelion Sol' 'Diana' 'Yorick' 'Cassiopeia' 'Sylas' 'Ahri' 'Ryze' 'Qiyana' 'Orianna' "Cho'Gath"
Cluster 7	**Cluster 8**	**Cluster 9**	**Cluster 10**	**Cluster 11**	**Cluster 12**	**Cluster 13**
'Sivir' 'Varus' 'Caitlyn' 'Aphelios'	'Malphite' 'Mordekaiser' 'Gnar' 'Neeko' 'Yuumi' 'Singed' 'Lissandra' 'Fiddlesticks' 'Zilean' 'Kennen' 'Maokai' 'Malzahar'	'Zyra' 'Karthus' 'Xerath' 'Heimerdinger' 'Ziggs' 'Illaoi'	'Sion' 'Urgot' 'Kled' 'Shen' 'Ornn' 'Renekton' 'Galio' 'Poppy' 'Wukong' 'Darius' 'Sett'	'Draven' 'Miss Fortune' 'Ashe' 'Jhin' "Kai'Sa" 'Kalista' 'Xayah'	'Janna' 'Vi' 'Anivia' 'Sona'	'Garen'
Cluster 14	**Cluster 15**	**Cluster 16**	**Cluster 17**	**Cluster 18**	**Cluster 19**	**Cluster 20**
'Gangplank' 'Evelynn' 'Azir' 'Lillia' 'Zed' 'Kayle'	'Nocturne' 'Olaf' 'Fizz' 'Zac' 'Yasuo' 'Hecarim' 'Pantheon'	'Veigar' 'Kassadin'	"Kha'Zix" 'Graves' 'Nidalee' 'Gragas' 'Kayn' 'Elise' 'Kindred'	"Kog'Maw" 'Shyvana' 'Ezreal'	'Alistar' 'Rakan' 'Ivern' 'Braum' 'Nami' 'Nautilus'	'Master Yi' 'Rengar' 'Amumu' 'Brand' 'Katarina' 'Shaco' 'Teemo' 'Warwick' 'Rammus' 'Annie'

Game Result Prediction

Game Result Prediction Based on In-game Features

Methods: logistic regression with LASSO, decision tree, KNN, random forest, SVM, neural network
Train : Validation : Test Set = 0.6 : 0.2 : 0.2

	Logistic with Lasso penalty	Decision Trees	SVM	KNN	Random Forest	Neural Network
Accuracy in Test set	98.81%	98.18%	99.13%	96.86%	99.06%	98.74%

The overall accuracy based on in-game features is very high! Why?
It is not surprising to see such a high prediction accuracy since some features are collected at the end of the game, which have already indicated which team would win.

"snowballing" effect: For example, one feature that shows up huge in both logistic regression and tree-based models is the **earned gold per minute**. During a game, a team with advantage would try to keep the enemy team away from minions and objects, preventing them from obtaining money and buffs. The team that falls behind is typically getting more and more difficult to come back behind.

Game Result Prediction

Game Result Prediction Based on Champions Clustering

Methods: logistic regression with LASSO, KNN, random forest, SVM, neural network

Train : Validation : Test Set = 0.6 : 0.2 : 0.2

	Logistic with Lasso penalty	SVM	KNN	Random Forest	Neural Network
Accuracy in Test set	53.74%	51.85%	52.79%	54.17%	56.17%

The overall accuracy based on champions clustering is better than random guessing (50%)

Apply the clusters we found using PCA to form features and create new predicting models. Using the results of the clustering algorithm, we create a one-hot 21 length vector for every champion. For each team we sum its vectors, then we concatenate two vectors of the matching teams. Using the 42-length vectors as input, we get the following result.

Further improvement: finding better cluster numbers and applying weights to the features in clustering, etc.

Recommendation and Insight

Topic 1: Recommendation on Resource Allocation

Tower and Inhibitor is always the focus, don’ t forget the goal is to destroy the opposed base.

Tower

Inhibitor

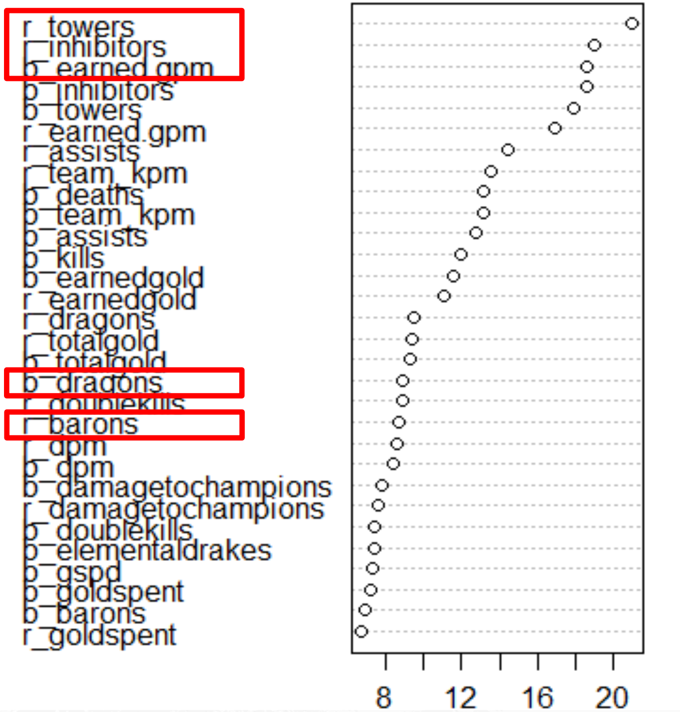
Gold Per Minute

The importance of dragon is larger than baron

Dragon

Baron

Figure: Feature importance obtained from the Random Forest model

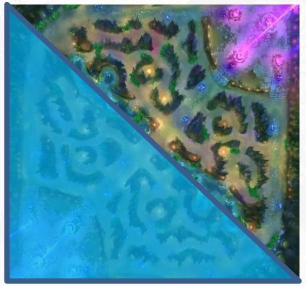


Recommendation and Insight

Topic 2: Recommendation on Pick Strategy

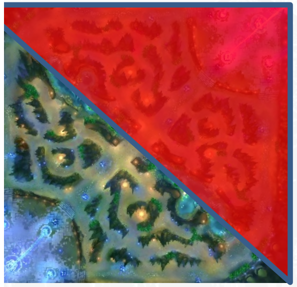
For the blue side and red side, if the opponent chooses a champion from one cluster, how can the team maximize the chance of winning by picking the natural enemy from another (or same) cluster?

Figure: Combination of the top 5 winning rates for each side (441 possibilities in total)

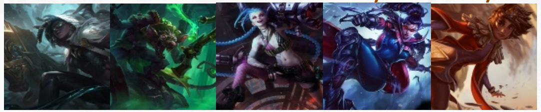


Blue Side Pick Cluster	Red Side Pick Cluster	Winning Rate **(Blue)**
0	14	61.7%
19	0	60.8%
2	0	60.7%

Blue Side Pick Cluster	Red Side Pick Cluster	Winning Rate **(Red)**
9	6	56.0%
8	7	54.6%
8	0	54.5%

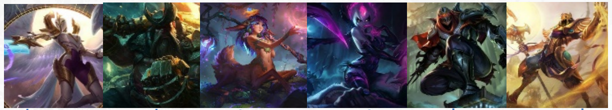


Cluster 0: 'Senna' 'Twitch' 'Jinx' 'Vayne' 'Taliyah'



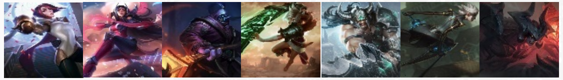
ADC (Attack Damage Carry) JNG

Cluster 14: 'Kayle' 'Gangplank' 'Lillia' 'Evelynn' 'Zed' 'Azir'

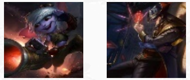


TOP JNG MID

Cluster 2: 'Fiora' 'Irelia 'Jax' 'Riven' 'Tryndamere' 'Camille' 'Aatrox' **'Tristana' 'Twisted Fate'**



TOP (Fighters)



ADC/MID (Marksman/Mage)

Thanks for your time

Annan Chen ac4619
Jiepeng Lian jl5521
Jing Xiao jx2422
Qingyuan Ren qr2130
Zezuan Zhang zz2705

COLUMBIA UNIVERSITY