Group - 23

Exploratory DOTA Analysis

Data Mining
(CS-685) Project
Odd Semester
2021-22

Under the guidance of
Dr. Arnab Bhattacharya

Department of
Computer Science
And Engineering
IIT Kanpur(IITK)

Aditya Loth (21111004) - adityal21@iitk.ac.in
Ashutosh Patel (21111018) - ashutoshp21@iitk.ac.in
Harsh Agarwal (21111030) - harshmag21@iitk.ac.in
Varun Vankudre (21111064) - varunsv21@iitk.ac.in

Overview

Analysis of popular game **DOTA 2** to find interesting patterns and results in tournament data.

Predicting likelihood of a side winning based on heroes picked by each side and items built during gameplay.

INTRODUCTION

- In the past decade, there has been unforeseen increase in no of people playing Esport games.
- Leading to generation of mammoth size raw data, ready to be processed and analyzed to unveil interesting trends and habits of players like most sought after heroes , strategies followed in different scenarios and much more.
- So riding this wave of Esports popularity, we have chosen DOTA2 as center of our attention to gather data , analyze to find astonishing results and to predict outcomes of matches by taking in two subsets of data gathered.

What is DOTA 2?

- Dota 2 is a famous free-to-play multiplayer online battle arena (MOBA) video game.
- Match is played between 2 teams. One side is radiant another is Dire.
- Objective of each team is take control of opponent base and destroy their Ancient.
- Each player has to pick a hero that synergizes with their team and counters the opponent’s hero combination.
- Once a player picks the hero, that hero cannot be picked by any other player in that match.
- A large pool of hero choices and a wide range of talents and abilities make each game unique.

Proposed Tool

The tool **Exploratory DOTA Analysis** is designed by :

- Gathering data using BeautifulSoup library and Steam Web API
- Building a Neural Network model for Prediction Engine.
- Performing Analysis on subsets of gathered data and visualising result as graphs and tables and storing most result as csv files.

Dataset

- Two Dataset were created, one for analysis and other for prediction

- For analysis -
  - Liquipedia website was scraped to get match ids.
  - Match ids from 5 The International tournaments was scraped
  - Tournaments of year 2016 - 2021 except 2020 , as not conducted due to COVID-19 pandemic.

Dataset

- For prediction-
    - Collected real time data from DotaBuff website and Steam API
    - Data was collected over the course of 5 days ~ 100 hours
    - 26000 match ids were collected with conditions:
        - Skill level of players involved is *'very high'*
        - Duration of Match is greater than 600 sec.
        - All Players involved are humans and no bots.
- Match details were fetched using Steam API

Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	Match_id	Player_1_playerid	Player_1_hero_id	Player_1_item1	Player_1_item2	Player_1_item3	Player_1_item4	Player_1_item5	Player_1_backpack1	Player_1_backpack2	Player_1_backpack3	Player_1_kills	Player_1_deaths	Player_
2	2557955371	93119769	44	168	50	116	51	208	0	0	0	17	5	
3	2558033556	93119769	73	235	48	116	147	119	0	0	0	5	5	
4	2558113366	94155156	90	244	214	36	188	46	0	0	0	0	5	
5	2558190331	129585121	6	263	63	108	116	0	0	0	0	9	3	
6	2558255566	87382579	23	46	151	50	244	131	0	0	0	2	3	
7	2558290213	129585121	93	117	63	252	36	152	0	0	0	5	0	
8	2559987699	41231571	17	41	77	98	63	121	0	0	0	3	3	
9	2560069057	41231571	65	0	116	1	214	102	0	0	0	4	4	
10	2560158271	86727555	112	180	265	0	244	46	0	0	0	1	4	
11	2560226228	139876032	101	1	0	131	254	46	0	0	0	5	11	
12	2562482317	129585121	6	263	63	52	147	164	0	0	0	6	3	
13	2562521800	129585121	8	1	208	145	160	147	0	0	0	14	2	
14	2562582896	86727555	68	180	104	102	39	46	0	0	0	4	10	
15	2562671370	134276083	2	1	100	214	125	188	0	0	0	2	6	
16	2566725198	129585121	15	116	154	112	149	236	0	0	0	6	1	
17	2566800064	129585121	59	151	29	236	36	164	0	0	0	7	4	
18	2558332885	85805514	75	0	92	244	0	214	0	0	0	6	2	
19	2558387942	87278757	20	0	214	36	46	90	0	0	0	1	12	
20	2558473213	4281729	75	214	36	218	244	0	0	0	0	0	4	
21	2558534849	82262664	62	11	180	244	202	79	0	0	0	3	5	
22	2560320998	94155156	30	187	63	0	36	218	0	0	0	1	10	
23	2560377671	85805514	50	92	79	46	63	218	0	0	0	1	8	
24	2560446360	87382579	23	100	50	151	34	254	0	0	0	0	2	
25	2560502724	98887913	38	194	102	110	1	48	0	0	0	3	4	
26	2562091541	41231571	9	119	176	1	48	116	0	0	0	6	5	
27	2562207201	89871557	41	147	63	196	116	81	0	0	0	5	2	
28	2562303634	82262664	23	131	42	0	151	50	0	0	0	5	6	
29	2562388622	139876032	87	43	214	46	0	16	0	0	0	0	7	
30	2562425844	139876032	91	41	214	218	34	46	0	0	0	0	7	
31	2564211031	85805514	50	90	218	188	180	46	0	0	0	0	5	
32	2564310434	87382579	103	178	1	100	214	30	0	0	0	4	4	
33	2564408925	87382579	16	102	36	1	214	0	0	0	0	0	3	
34	2564485350	89871557	48	147	63	116	156	236	0	0	0	11	1	
35	2564547149	82262664	68	79	180	46	102	36	0	0	0	6	8	
36	2564619977	134276083	16	1	244	29	232	46	0	0	0	1	6	
37	2564666119	87382579	84	0	102	229	244	180	0	0	0	0	1	
38	2566482748	89871557	34	1	96	108	48	232	0	0	0	9	1	
39	2566594115	93119769	9	119	48	176	1	204	0	0	0	8	5	
40	2566866066	89871557	9	1	50	176	116	166	0	0	0	6	4	
41	2566934520	87382579	23	250	151	1	46	0	0	0	0	4	4	

Prediction of Matches

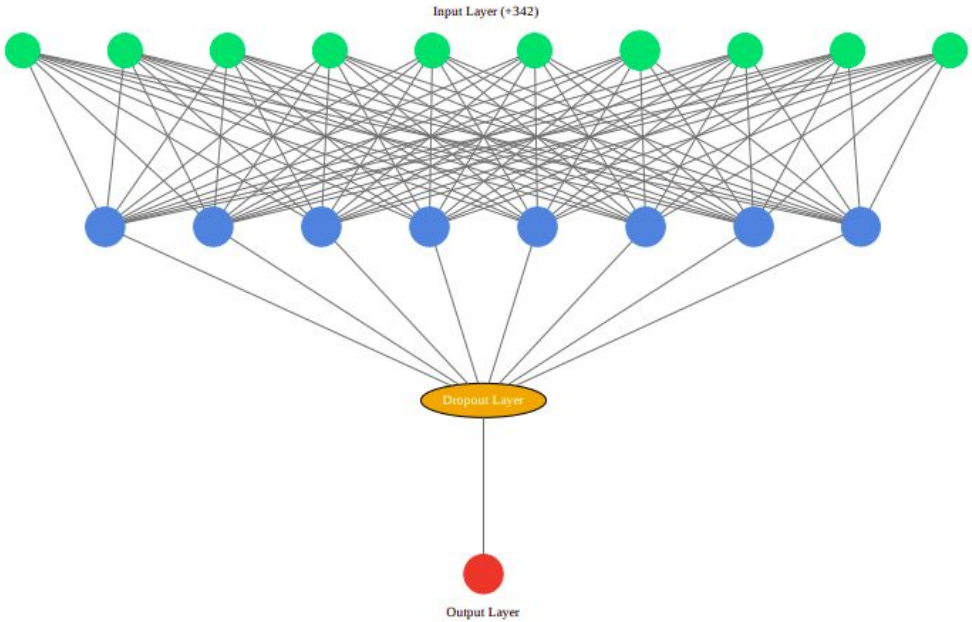
PART 1

Using only Hero IDs

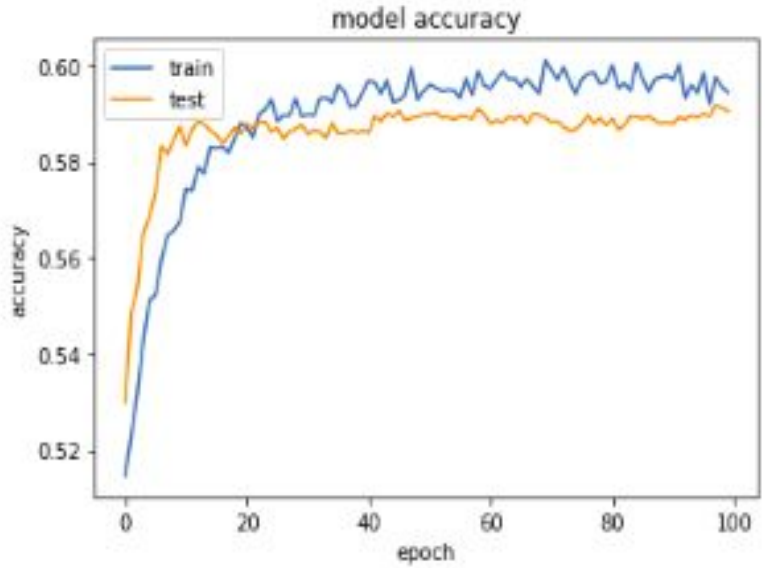
PART 2

Using Hero IDs, Items and Backpacks

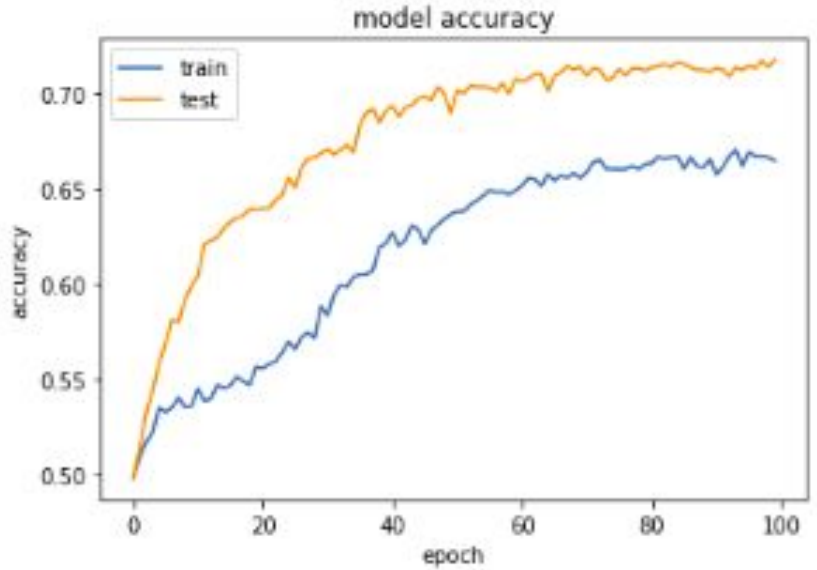
Neural Network Architecture



Results & Comparisons

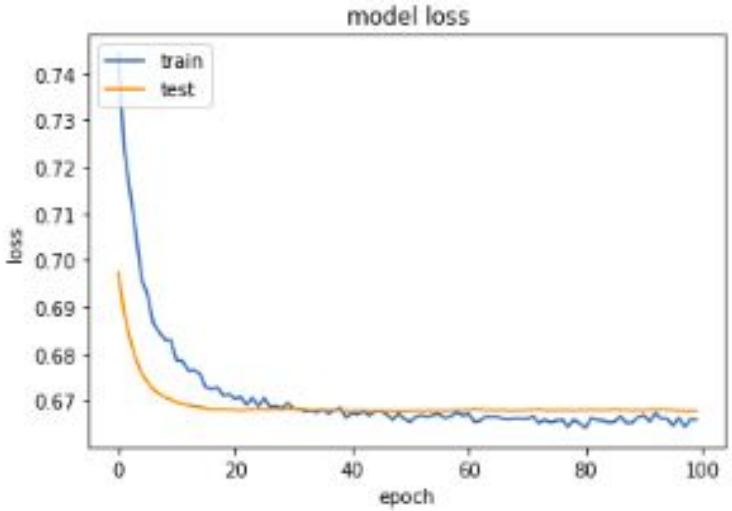


Model_1: Epoch VS Accuracy

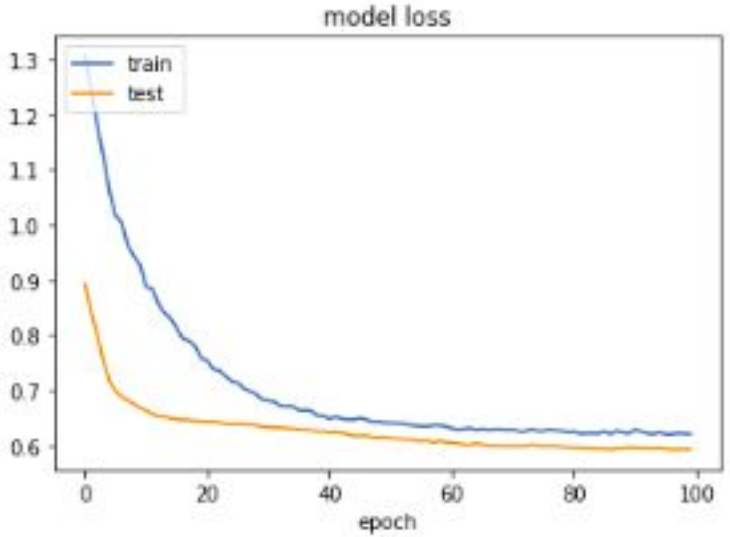


Model_2: Epoch VS Accuracy

Results & Comparisons



Model_1: Epoch VS Loss



Model_2: Epoch VS Loss

Results & Comparisons

- Model 1 Accuracy on test data

```
testloss,testacc=model.evaluate(xtest,ytest,verbose=0)
testacc
```

```
0.591559112071991
```

- Model 2 accuracy on test data

```
testloss,testacc=model.evaluate(xtest,ytest,verbose=0)
testacc
```

```
0.7115384340286255
```

- Prediction of Winning Odds

```
For match id: 132
Before the game begins:
Radiant has 60.86% chance of winning.
After choosing items:
Dire has 60.74% chance of winning.
```

```
For match id: 755
Before the game begins:
Radiant has 52.04% chance of winning.
After choosing items:
Radiant has 71.33% chance of winning.
```

Analysis

The analysis is partitioned into 4 categories:

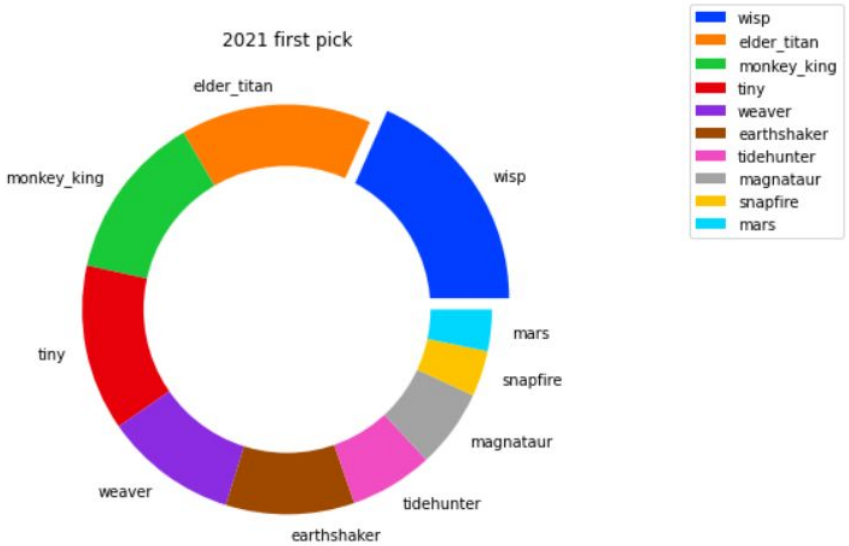
1. Analysis on Heroes : No of Times Picked, No of Times Banned, Win Rate, etc
2. Analysis on Players : Highest Kills , Highest Assist Average, Highest GPM Average, etc
3. Analysis on Matches of Tournament : No of Unique Heroes Picked, Shortest Duration, etc
4. Analysis on Teams : Most Different Heroes Picked, Fewest Different Heroes Picked, etc

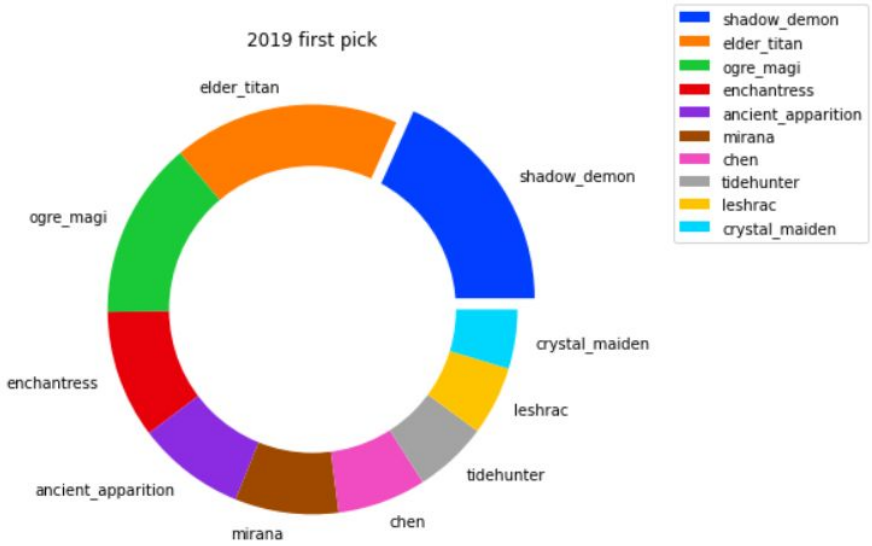
Analysis

General Analysis Procedure :

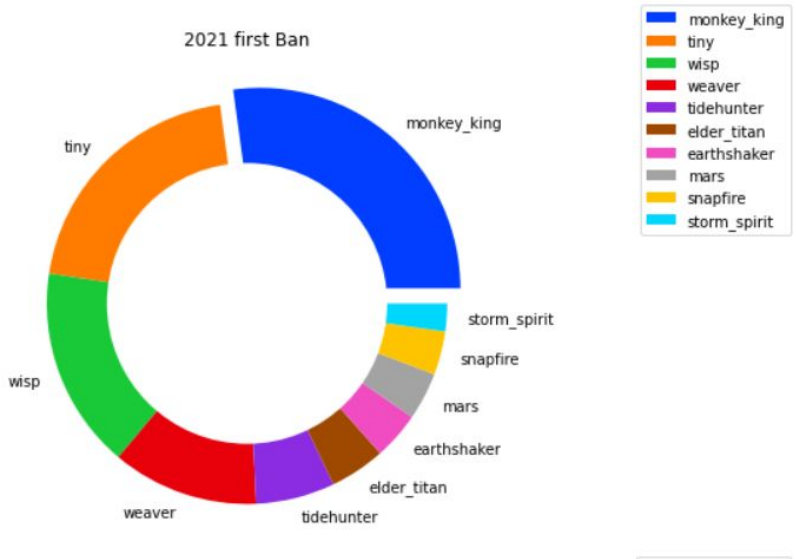
1. Read the CSV for a particular year tournament into DataFrame.
2. Subset required columns for current analysis.
3. Perform Data Aggregation.
4. Visualize analysis result in graphs like Bar Graph, Line Plot etc, if required.
5. Store the result as CSV or text file.

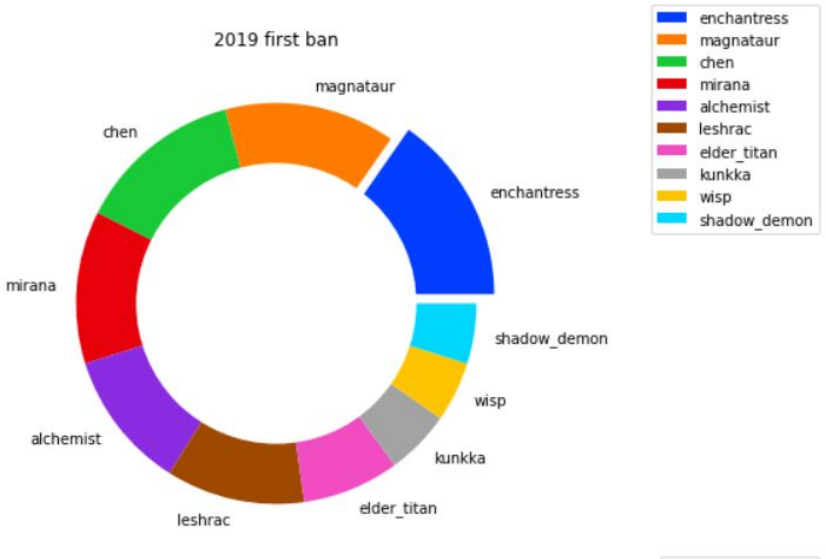
2021 V/S 2019 FIRST PICK





2021 V/S 2019 FIRST BAN





Conclusion and Further Works

- Model accuracy can be increased by providing diverse and large amount of data, to better predict winning odds of a side in a match.

- Currently Prediction engine does not take into account variation in hero's abilities across patches and therefore is most accurate and useful for current patch v7.30e

- Currently for analysis purposes, items for each player and as a whole were not considered due to sheer huge amount of possible permutations and combinations possible for data which as a result requires very large dataset of matches to get some sensible patterns to make analysis possible which was not possible at the moment due to computing and time constraint for the project and thus in future to extract huge set of matches to perform analysis on items for each player

Thank You.