Group 23: Exploratory DOTA analysis

Aditya Loth, Ashutosh Patel, Harsh Agarwal, Varun Vankudre
21111004 , 21111018 , 21111030 , 21111064
{adityal21, ashutoshp21, harshmag21, varunsv21}@iitk.ac.in
Indian Institute of Technology Kanpur (IIT Kanpur)

Abstract

In this project, we have analyzed the data of tournaments of a popular video game, DOTA 2. We analyzed the five latest annual competitions to find interesting patterns in them. We also have built two prediction models: one that takes heroes chosen by each side and the other that takes heroes and their respective generated item(s) to predict which side is likely to win.

1 Introduction

Dota 2 is a famous free-to-play multiplayer online battle arena (MOBA) video game. Every match is played between 2 teams of 5 players each. One team is the dire side, and the other is the radiant side. The objective of each team in a match is to destroy the other team's base. Each character is commonly known as heroes. Heroes have their strengths, weaknesses, and talents. Each player has to pick a hero to synergize with their team and counter the opponent's hero combination. Once a player chooses the hero, that hero cannot be selected 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. An effective hero pick can give a particular side an advantage before even the game begins.

Periodically, game developers release a new 'meta' that changes the game, such as adding new Heroes for players to choose from, tweaking current hero stats like giving them a unique ability or increasing/decreasing their health power etc. So periodically, pro players and teams have to either tweak or completely overhaul their strategy to incorporate the new meta. Analysis of the meta becomes crucial while forming a new startegy to win the game.

Owing to the sudden increase in popularity of Esports games in the past decade, has led many companies to build tools that analyze humongous amounts of data generated during these games to determine various winning factors like play style, hero combinations, etc and then using the analysed data as factors in predicting outcome of these matches. This service is in huge demand by pro players and e-sport team management companies to improve their teams performance in tournaments and to recruit undiscovered players with huge potential to play with their team perfectly to build a team which gives serious competition to its opponents. Pro players and pro teams prefer the analysis services provided by third-party to study their opponents strategy and give them a report through which they can devise a proper strategy to gain advantage during the matches.

2 Methodology

2.1 Data Collection

Data gathering was done mainly through two methods, that is through web scraping using BeautifulSoup library and through WebAPI developed by 'Steam' for extracting DOTA match history and details. We used Steam's API to first fetch 26000 match ids played in a period of about 5 days. Some constraints on the matches were placed to get high quality matches.

```
url = 'http://api.steampowered.com/IDOTA2Match_570/GetMatchHistory/v001/?key=&skill=3&min_players=10'
source = requests.get(url)
source.text.replace('\n','')
```

Figure 1: Match Id

Above code is to gather Match ids of recent matches being played outside of any tournament with some constraints. Constraints are :

- Skill determines the skill level of players involved in the match, it is set to 3 by us signifying a very high skill level so we can get top 8
- Min-players set the condition that all players in the match should be human and no bots, this was done as matches with bot do not provide consistent data and most of the times match or game is abandoned without result.

```
url = 'http://api.steampowered.com/IDOTA2Match_570/GetMatchHistory/v001/?key=&skill=3&min_players=10'
source = requests.get(url)
source.text.replace('\n','')
a = json.loads(source.text)
```

Figure 2: Match Details

Then we used the 26000 fetched match ids to get details about each match individually and store it in a csv file. Match details like result, score of each side, team name if such exists for each side. Also for each player in the match, we collected player id , hero id (hero he chose), items generated during gameplay, kills, assists, deaths and many more features and stored it in a csv. This was again done using Steam's API but with different query and parameters. We used the above 26000 match detail file to build our prediction models. The first model takes a set of heroes chosen by each side and predicts the likelihood of a side winning and the second model takes data generated during gameplay like heroes and their items to predict outcome of the match.

Attributes for each match:

- 'match_id'
- For each player :
  'playerid','hero_id', 'item1', 'item2', 'item3', 'item4', 'item5', 'backpack1', 'backpack2', 'backpack3', 'kills', 'deaths', 'assists', 'hits', 'denies', 'gold_per_min', 'xp_per_min', 'level', 'networth', 'aghanims_scepter', 'aghanims_shard','moonshard', 'hero_damage', 'tower_damage', 'hero_healing', 'gold', 'gold_spent', 'scaled_hero_damage', 'scaled_tower_damage','Scaled_hero_healing',
- 'Result',
- 'Duration',
- 'First_blood_time',
- 'Lobby_type',
- 'Game_mode',
- 'Radiant_score',
- 'Dire_score'
- 'Radiant_team_id'
- 'Dire_team_id'
- 'Radiant_team_name'
- 'Dire_team_name'

We have scraped the match ids of particular tournaments from the liquipedia website for the analysis. The match data for these match IDs is fetched using Steam API and stored in CSV for further analysis. For the analysis of tournaments, only main stage and group stage matches are considered.

2.2 Prediction

We have divided the prediction process into two parts: Before the game begins and in the mid-game. For the first part, we will look only at the Hero ids of each player because before the game starts, each player has to choose a hero. So our first prediction will be considering only hero ids. We predict the team's winning chances in the mid-game for the next part, i.e. when players have started building their items and backpacks. Here the assumption is made that all players playing a particular character have the same skill level. In the paper [1], they have used Logistic Regression and Augmented Logistic Regression but we decided to use Neural Networks for predictions.

2.2.1 Neural Network architecture

For predictions, we have used Neural Networks. Tensorflow and Keras libraries are used to build the classification model. As a base layer, Keras Dense layer is selected from the Keras sequential model. This layer has eight nodes and input dimensions as a number of column vectors. For this layer we tried multiple activation functions but, 'tanh' was giving best performance. To avoid overfitting of model, we have used Keras Dropout Layer on top of our base layer.Then final layer has a single node with activation function as 'sigmoid'. The Model is compiled using 'binary_crossentropy' as loss function, 'adam' as optimiser and 'accuracy' as the metric. This Neural Network Model was used for both types of predictions. The architecture can be visualised in the Figure 3.

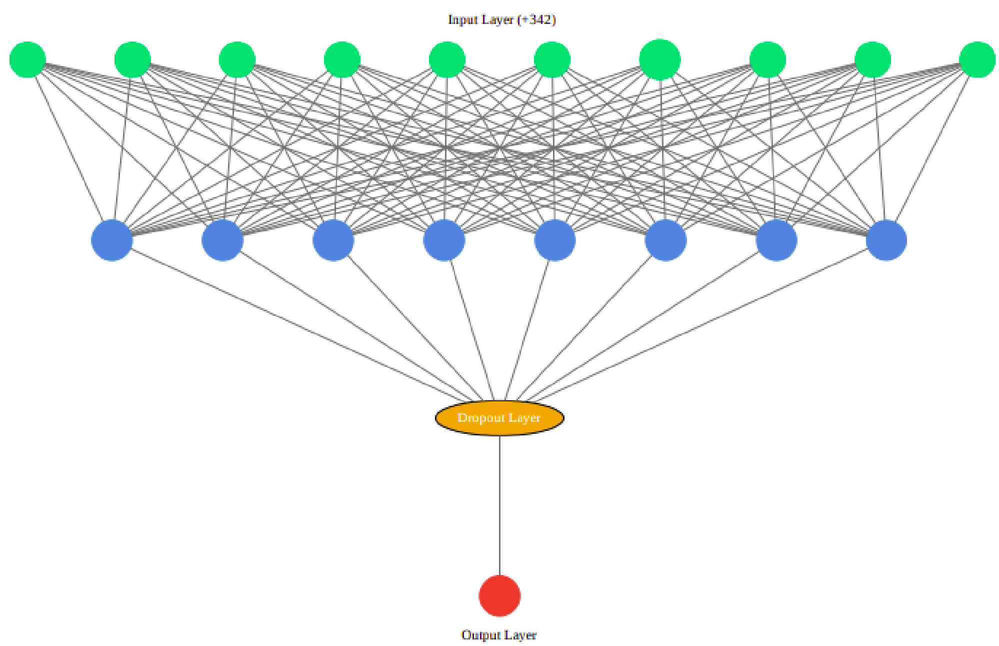

Figure 3: Neural Network Architecture

2.2.2 Pre-Game Prediciton

Each player from each team chooses a hero to play with before the game starts. That means we have ten features per match as heros. As Hero ID is a categorical attribute, and in total, the game contains 122 heroes, we converted them into One-Hot Encoding as done in the paper [2]. For each team, we used a vector of 122 columns. Each column represents if the hero is selected by the team or not. Similarly the heros chosen by team 2 are also converted to One-Hot Encoding. For labels,the "Result" column in collected dataset, which has entries of the winning team. The "Radiant" values were replaced with 0 and "Dire" one with 1. After that, the Neural Network model was trained using these features and label values. The epoch numbers were set to 100 and used batches of size 1000. The loss and accuracy graph is shown in the Figure 4 and Figure 5.

$$x_i = \begin{cases} 1 \text{ if a radiant player played as the hero with id i} \\ 0, \text{otherwise} \end{cases}$$

$$x_{136+i} = \begin{cases} 1 \text{ if a dire player played as the hero with id i} \\ 0, \text{otherwise} \end{cases}$$

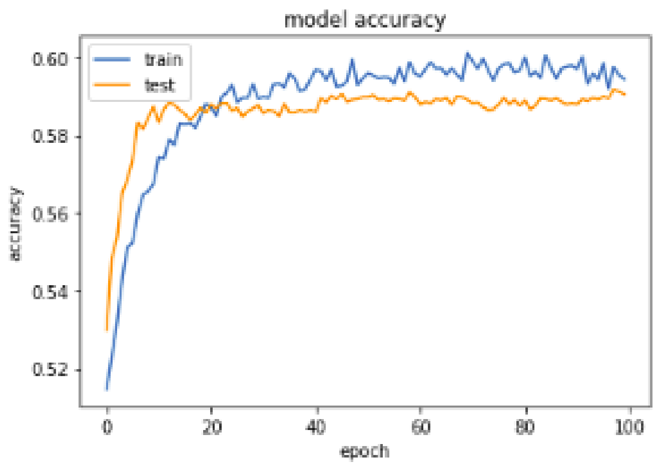

Figure 4: Accuracy v/s epochs for model 1

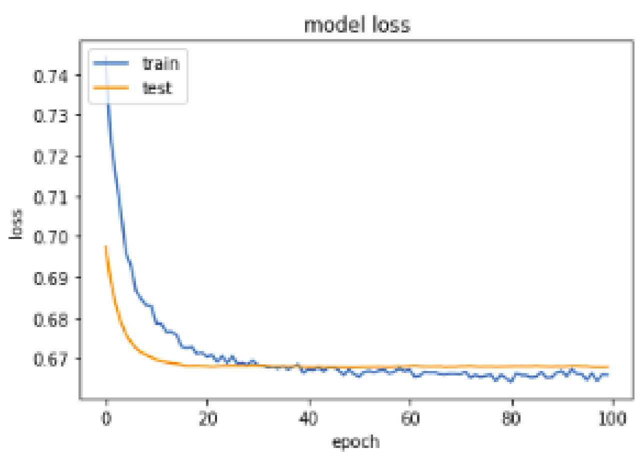

Figure 5: Loss v/s epochs for model 1

After training the model gives an accuracy score of 0.59 on both training as well as testing data. Testing data accuracy is shown in Figure 3.

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

0.591559112071991
```

Figure 6: Accuracy on test data

It shows that choosing good set of heroes has a slight effect on predicting who will win in the end. Carefully choosing the heros affects the winning chances of a team.

2.2.3 Mid-game Prediction

For the second part of prediction process, the model must predict the winning chances using hero ids, items picked by players and the backpacks chosen. A player choses items and backpacks after the game starts and progresses. He can choose five different items and three different backpacks. Considering the items and backpacks chosen as attributes, we now have additional eight attributes for each player. That makes an additional 80 new features. With 272 One Hot Encoded hero ids it makes the features to 352 in total. The labels are same as before, i.e. "Result" column. The same Neural Network architecture was used for building new model. Parameters like loss function and optimizer are kept same as first model. The new model is trained using 352 feature columns and their labels. The number of epochs are set to 100 and batch size to 1000. The accuracy and loss graphs are shown in the Figure 7 and Figure 8.

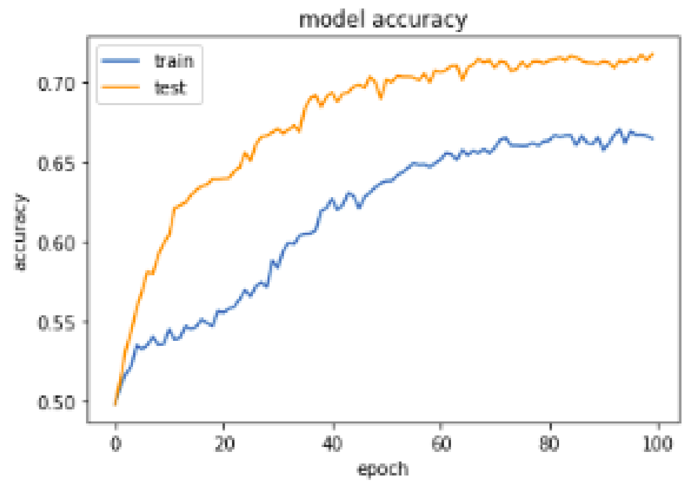

Figure 7: Accuracy v/s epochs for model 2

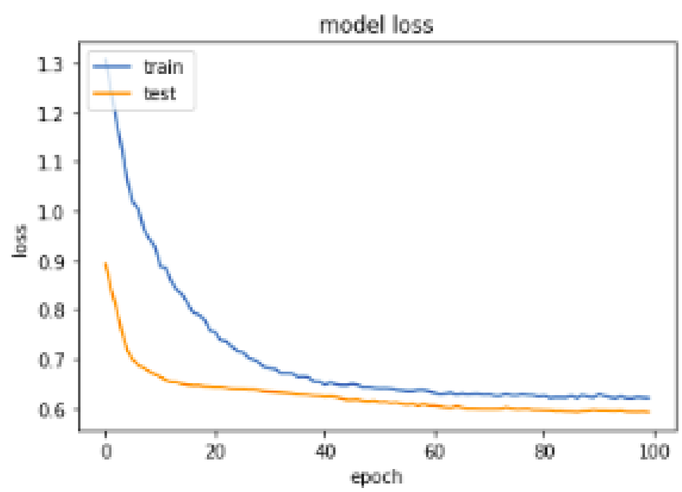

Figure 8: Loss v/s epochs for model 2

This time the model was giving accuracy score of 0.67 on training data and 0.71 on testing data. Testing accuracy is shown in the Figure 9.

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

```
0.7115384340286255
```

Figure 9: Accuracy on test data

Using items and backpacks clearly improved our prediction model. This shows that knowing the items and backpacks along with hero ids chosen for each player can help us in predicting the end result of the game.

2.3 Analysis

Process followed for Analysis:

- Read appropriate CSV file into pandas DataFrame
- Take Subset of columns required for Analysis
- Perform Aggregation of Data based on conditions
- Display top results and generate graphs (**if required**)
- Store analysis output in CSV with appropriate file name

Following Analysis were performed:

- Hero wise : Total picks, Total bans, Total Contested, Hero win rate, Hero Kill average , Hero xpm average, Hero last hit average
- Tournament Wise : No of Unique Heroes Picked , No of Unique Heroes Banned, Most Contested Hero, Most Kills in a Match , Longest Duration Match, Shortest Duration Match, Most Kills by a Hero in a Match, Most Deaths of a Hero in a Match, Highest Gold Per Minute by a Hero in a match
- Player wise : Average kills, Total kill, Lowest Death Average, Assists Average, Total Assists, Total last hits, Total last hits average, Most GPM in a match, GPM average
- Team Wise : Most Different Heroes Picked, Fewest Different Heroes Picked, Longest Average Match Duration, Shortest Average Match Duration, Most Deaths, Fewest Deaths, Highest Kill Average, Lowest Kill Average, Most Kills, Least Kills, Highest Kill Death Ratio, Lowest Kill Death Ratio, Highest Win Rate, Lowest Win Rate.
- Across Tournament : Average Kills Per Minute, Average Match Duration, Average Kills Per Match, Average Gold Per Minute, Average Last Hits, Total Denies, Average Assists, Average First Blood Time,
- Across Tournament for heroes : Total no of wins, Total no picks, Total no bans.

3 Results

3.1 Analysis

3.1.1 Tournament Wise

For each International Tournament from year 2016 to 2021, except for year 2020 when tournament was cancelled due to COVID-19 pandemic. Following results were observed

- Unique Heroes Picked

  - 2016 : 104
  - 2017 : 107
  - 2018 : 110

  - 2019 : 114
  - 2021 : 113

- Unique Heroes Banned

  - 2016 : 82
  - 2017 : 97
  - 2018 : 96
  - 2019 : 102
  - 2021 : 105

- Most Contested Hero

  - 2016 : Dragon Knight 136 times
  - 2017 : Rattle Trap 158 times
  - 2018 : Enchantress 184 times
  - 2019 : Alchemist 172 times
  - 2021 : Monkey King 188 times

- Most Kills in a Match

  - 2016 : 115
  - 2017 : 104
  - 2018 : 108
  - 2019 : 96
  - 2021 : 89

- Longest Match of Tournament

  - 2016 : 1 H 16 M 57 S
  - 2017 : 2 H 8 M 8 S
  - 2018 : 1 H 21 M 52 S
  - 2019 : 1 H 26 M 29 S
  - 2021 : 1 H 4 M 42 S

- Shortest Match of Tournament

  - 2016 : 0 H 15 M 28 S
  - 2017 : 0 H 15 M 12 S
  - 2018 : 0 H 16 M 59 S
  - 2019 : 0 H 15 M 8 S
  - 2021 : 0 H 17 M 5 S

- Most Kills by a Hero in a Match

  - 2016 : By Invoker 23 Kills
  - 2017 : By Queen of Pain 25 Kills
  - 2018 : By Tiny 31 Kills
  - 2019 : By Storm Spirit 27 Kills
  - 2021 : By Dawn Breaker 27 Kills

- Most Deaths of a Hero in a Match
  - 2016 : Of Keeper of the Light 17 Deaths
  - 2017 : Of Pugna 20 Deaths
  - 2018 : Of Jakiro 18 Deaths
  - 2019 : Of Crystal Maiden 17 Deaths
  - 2021 : Of Ogre Magi 17 Deaths
- Highest Gold Per Minute by a Hero in a Match
  - 2016 : Of Keeper of the Light 17 Deaths
  - 2017 : Of Pugna 20 Deaths
  - 2018 : Of Jakiro 18 Deaths
  - 2019 : Of Crystal Maiden 17 Deaths
  - 2021 : By Luna 973 Gold Per Minute

3.1.2 Player Wise

For each International Tournament from year 2016 to 2021, except for year 2020 when tournament was cancelled due to COVID-19 pandemic. Following results were observed

- Most Kills by a Player
  - 2016 : 232 by 86725175
  - 2017 : 321 by 105248644
  - 2018 : 295 by 106863163
  - 2019 : 265 by 86700461
  - 2021 : 309 by 321580662
- Lowest Death Average of Player
  - 2016 : 2.818 by 86725175
  - 2017 : 2.346 by 177416702
  - 2018 : 3.045 by 86725175
  - 2019 : 2.115 by 125581247
  - 2021 : 1.857 by 898754153
- Most Assists by a Player
  - 2016 : 409 by 87382579
  - 2017 : 474 by 101356886
  - 2018 : 526 by 101695162
  - 2019 : 483 by 82262664
  - 2021 : 472 by 113331514
- Most Last Hits by a Player
  - 2016 : 10815 by 86725175
  - 2017 : 9983 by 105248644
  - 2018 : 9761 by 125581247
  - 2019 : 11980 by 105248644

- 2021 : 14758 by 321580662

- Highest Gold Per Minute by a Player
  - 2016 : By 86725175 19525 Gold Per Minute
  - 2017 : By 105248644 20834 Gold Per Minute
  - 2018 : By 125581247 18166 Gold Per Minute
  - 2019 : By 105248644 20274 Gold Per Minute
  - 2021 : By 321580662 25050 Gold Per Minute

3.1.3 Hero Wise

For each International Tournament from year 2016 to 2021, except for year 2020 when tournament was cancelled due to COVID-19 pandemic. Following results were observed

- Most Picked Hero
  - 2016 : 81 Mirana
  - 2017 : 70 Earth Shaker
  - 2018 : 89 Vengeful Spirit
  - 2019 : 75 Elder Titan
  - 2021 : 73 Elder Titan
- Most Banned Hero
  - 2016 : 95 Wisp
  - 2017 : 107 by Night Stalker
  - 2018 : 150 Enchantress
  - 2019 : 139 Alchemist
  - 2021 : 140 Monkey King
- Most Contested Hero
  - 2016 : 136 Elder titan
  - 2017 : 158 Night Stalker
  - 2018 : 184 Enchantress
  - 2019 : 172 Alchemist
  - 2021 : 188 Monkey King
- Highest Win Rate of Hero
  - 2016 : 1 Obsidian Destroyer, Spectre, Nevermore, Viper, Zuus, Chaos Knight, Centaur,Abaddon
  - 2017 : 1 Huskar , Ogre Magi , Abaddon
  - 2018 : 1 Queen of Pain , Meepo , Tide Hunter , Beast Master , Spirit Breaker , Naga Siren
  - 2019 : 1 Winter Wyvern , Night Stalker
  - 2021 : 1 Chen, Arc Warden , Visage , Riki
- Highest Kill Average of a Hero
  - 2016 : 14 Obsidian Destroyer

- 2017 : 12.429 Phantom Assassin
- 2018 : 14 Chaos Knight
- 2019 : 12.333 Monkey King
- 2021 : 12 Arc Warden

- Highest Experience Per Minute Average of a Hero
  - 2016 : 598.57 Ember Spirit
  - 2017 : 808 Meepo
  - 2018 : 843.667 Meepo
  - 2019 : 782.846 Brood Mother
  - 2021 : 834 Arc Warden
- Highest Last Hit Average of a Hero
  - 2016 : 452 Nevermore
  - 2017 : 551.313 Antimage
  - 2018 : 590.667 Tinker
  - 2019 : 584 Luna
  - 2021 : 598 Alchemist

3.1.4 Team Wise

For each International Tournament from year 2016 to 2021, except for year 2020 when tournament was cancelled due to COVID-19 pandemic. Following results were observed

- Most Different Heroes Picked
  - 2016 : The Wings Gaming 58
  - 2017 : Newbee 54
  - 2018 : Team Secret 58
  - 2019 : Team Liquid 62
  - 2021 : Team Spirit 6
- Fewest Different Heroes Picked
  - 2016 : EHOME 30
  - 2017 : 123 32
  - 2018 : TNC Predator 34
  - 2019 : Mineski 37
  - 2021 : Evil Geniuses 38
- Longest Average Match Duration
  - 2016 : PSG.LGD 0 H 45 M 31 S
  - 2017 : Team Empire 0 H 47 M 16 S
  - 2018 : TNC Predator 0 H 43 M 15 S
  - 2019 : TNC Predator 0 H 50 M 58 S
  - 2021 : Fnatic 0 H 42 M 26 S
- Shortest Average Match Duration

  - 2016 : Natus Vincere 0 H 33 M 47 S
  - 2017 : Fnatic 0 H 31 M 14 S
  - 2018 : Team Serenity 0 H 34 M 36 S
  - 2019 : OG 0 H 34 M 13 S
  - 2021 : Evil Geniuses 0 H 36 M 26 S

- Most Deaths

  - 2016 : 123 678
  - 2017 : Team Liquid 887
  - 2018 : OG 1024
  - 2019 : Team Liquid 802
  - 2021 : Team Spirit 708

- Fewest Death

  - 2016 : Team Secret 281
  - 2017 : Evil Geniuses 398
  - 2018 : Fnatic 443
  - 2019 : Alliance 407
  - 2021 : Evil Geniuses 379

- Highest Kill Average

  - 2016 : Evil Geniuses 27.391
  - 2017 : LGD.Forever Young 28.423
  - 2018 : Evil Geniuses 34.815
  - 2019 : OG 30.5
  - 2021 : INVICTUS GAMING 28.8

- Lowest Kill Average

  - 2016 : Escape Gaming 12.8
  - 2017 : Hell Raisers 16.75
  - 2018 : Winstrike 23.191
  - 2019 : Chaos EC 17.25
  - 2021 : Thunder Predator 17.063

- Most Kills

  - 2016 : 123 750
  - 2017 : Team Liquid 911
  - 2018 : PSG.LGD 1027
  - 2019 : OG 854
  - 2021 : Team Spirit 851

- Least Kills

  - 2016 : Escape Gaming 192
  - 2017 : Hell Raisers 268
  - 2018 : paiN Gaming 402

  - 2019 : Chaos EC 276
  - 2021 : Thunder Predator 273

- Highest Kill Death Ratio

  - 2016 : EHOME 1.328
  - 2017 : LGD.Forever Young 1.543
  - 2018 : Virtus.pro 1.313
  - 2019 : PSG.LGD 1.457
  - 2021 : PSG.LGD 1.523

- Lowest Kill Death Ratio

  - 2016 : Escape Gaming 0.563
  - 2017 : Fnatic 0.522
  - 2018 : paiN Gaming 0.667
  - 2019 : Chaos EC 0.559
  - 2021 : Thunder Predator 0.543

- Highest Win Rate

  - 2016 : EHOME 0.7
  - 2017 : Team Liquid 0.771
  - 2018 : Evil Geniuses 0.741
  - 2019 : OG 0.821
  - 2021 : PSG.LGD 0.821

- Lowest Win Rate

  - 2016 : Escape Gaming 0.133
  - 2017 : Hell Raisers 0.063
  - 2018 : INVICTUS GAMING 0.222
  - 2019 : Chaos EC 0.188
  - 2021 : Thunder Predator 0

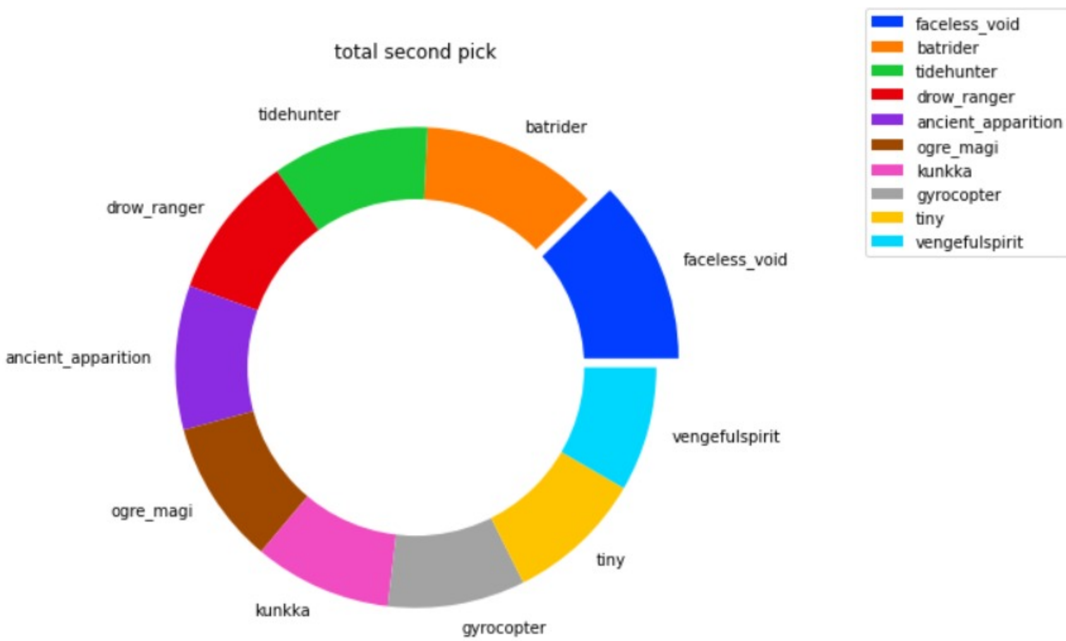

Figure 10: Top 10 Second Picked Heroes

For more detailed results of each of aforementioned analysis, please refer generated CSV's and python notebook present in the project to get visual representation of above mentioned analysis and for various other analyses.

3.2 Prediction

Despite using items and backpacks in addition to heroes the model is still not that accurate. So it is better to predict the chances of a team winning instead of just mentioning who will win. These chances were calculated for 3 random matches. First of all the probability of a team winning before the game starts is calculated. For this purpose the first model used the hero ids of players and predicted the winning team along with the probability of winning. Similarly the mid-game(after choosing items) chances were calculated . For this the second model used items, backpacks and hero ids for prediction. These chances were shown as percentages for each match as shown in the Figure 8.

```
For match id: 561
Before the game begins:
Dire has 59.7% chance of winning.
After choosing items:
Radiant has 54.82% chance of winning.


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.
```

Figure 11: Prediction of matches

4 Future Work

- Every few months, the game developers release a new patch to balance the game like adding new heroes and changing the power,health and abilities of some old heroes. So players need to adapt their play style to these changes. Currently our prediction engine does not take into account a hero's abilities and therefore is most accurate for the current patch. Therefore to build an engine such that it takes the hero's ability too into account during prediction, therefore making it useful across different patches.
- To build a recommendation engine which reads the opponent team's choice of heroes and outputs one or more hero combinations with winning odds for the user and also suggests which item(s) to build for gaining early advantage in a match. This was not done at present as it requires a huge dataset of matches to analyse winning chances of a combination against the opponent's team and also lack of method to verify the recommendation given.
- 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

5 Conclusion

In this project, The real-time data of 26000 matches was collected for prediction, which took around five days using Steam API and scraping the Dotabuff website[3]. The data for analysis was collected from scraping liquipedia website[4] and Steam API[5].

Most of the analysis is done seperately for each International Tournament. Each year's analysis was done in four categories heroes, players, teams and matches.

The prediction model predicts which team is likely to win using the inputted data. We see that choosing more attributes gives us more confidence in predicting the outcome of the match. The model tries to find the relationship between heroes and items. Our results show that the model accuracy can be increased by giving a large amount of data, and the model can make predictions better.

References

[1] K. Kalyanaraman, “To win or not to win ? a prediction model to determine the outcome of a dota 2 match,” 2015.

[2] L. T. Petra Grutzik, Joe Higgins, “Predicting outcomes of professional dota 2 matches,” 2017.

[3] “Dota buff- https://www.dotabuff.com/players/132909157,”

[4] “Liquipedia- https://liquipedia.net/dota2,”

[5] “Steam api-http://wiki.teamfortress.com/wiki/webapi,”