Predicting Outcome of Dota 2 Match

DIAS SANIYAZOV, 57478958
MUSTAFA ABYLKHANULY, 57797030
YERNUR AIDARBEK, 57478958
NUR ALDEKEN, 56679433
BEKET YERMEKOV, 56679630

Data analysis stands as a cornerstone in the realm of Dota 2, a captivating game boasting a roster of 124 heroes, 150 purchasable items, and 58 neutral items. With an astounding player base of 490,000 engaged warriors per hour, the game generates an immense volume of data. Compounded by its ever-evolving nature, characterized by the continuous expansion of heroes and the periodic release of patches every one to three months, accurately predicting game outcomes based on individual player performance becomes a formidable challenge. The inherent inconsistency of human players and the unpredictable battlefield scenarios within Dota 2 further compound this complexity. Nonetheless, a glimmer of hope emerges in the ability to forecast match results based on the initial hero selection. Thus, this project embarks on a rigorous exploration, meticulously analyzing a vast dataset comprising 150,000 Dota 2 professional matches. This project aims to unveil the intricate patterns that leads to lose or win in this game.

Additional Key Words and Phrases: datasets, machine learning, dota 2, art and science of data

ACM Reference Format:
Dias Saniyazov, Mustafa Abylkhanuly, Yernur Aidarbek, Nur Aldeken, and Beket Yermekov. 2024. Predicting Outcome of Dota 2 Match. *J. ACM* 37, 4, Article 111 (April 2024), 13 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The idea to train a model and conduct data analysis in the Dota 2 sphere was inspired by Team Spirit's victory at The International 2021. This win was unexpected, as the team had a low rank compared to others at that time and it was their first appearance in a top-tier Dota tournament. Their victory sparked discussions about fake matches and pre-written scripts. However, Team Spirit went on to win The International 2023, Riyadh Masters 2023, and other tournaments, earning a total of 27,404,259 dollars and becoming one of the most renowned teams in Dota 2 history[1].

Following their first win at The International 2021, Valve released a True Sight video that lasted one hour and thirty-seven minutes. The video provided a detailed look at the last five matches between Team Spirit and PSG.LGD, including team communication before, during, and after the matches, as well as the preparation process[2]. While most spectators focused on the games themselves, this True Sight video shed light on the heroes that were picked and banned, and the order in which they were chosen. It highlighted the consequences of these decisions and how the

Authors' addresses: Dias Saniyazov, dsaniyazo2-c@my.cityu.edu.hk, 57478958; Mustafa Abylkhanuly, mabylkhan2-c@my.cityu.edu.hk, 57797030; Yernur Aidarbek, yaidarbek2-c@my.cityu.edu.hk, 57478958; Nur Aldeken, naldeken2-c@my.cityu.edu.hk, 56679433; Beket Yermekov, byermekov2-c@my.cityu.edu.hk, 56679630.

0004-5411/2024/4-ART111 $15.00
https://doi.org/XXXXXXX.XXXXXXX

planning during the draft stage affected the game's outcome. The True Sight video revealed that the seemingly inconspicuous draft stage had a significant impact on the match result.

Despite the existence of websites like DOTABUFF.com and Stratz.com, which provide information on win possibilities based on drafts, hero win rates against each other, and the frequency of purchasing specific items on each hero, this project aims to analyze the data from a different perspective. It primarily focuses on identifying the features that have the greatest impact on match outcomes, based on data from completed games. The project takes into account the player's position and hero, and aims to train a model that can predict the game outcome based solely on the ten heroes selected during the draft stage.

2 BACKGROUND

Dota 2 is a complex and competitive multiplayer online battle arena (MOBA) game that involves two teams, Radiant and Dire, battling against each other on a symmetrical map. Here are some key aspects of Dota 2:

– `Heroes`: Dota 2 features a diverse roster of 124 unique heroes, each with their own abilities, strengths, and roles. Heroes are the playable characters that players control during the game. They can be categorized into various roles/positions such as carry, support, initiator, and more, depending on their playstyle and abilities[3].
– `Strategy and Teamwork`: Dota 2 is heavily focused on strategic decision-making and teamwork. Players must work together to coordinate their actions, communicate effectively, and devise strategies to outsmart the opposing team. Team composition, item builds, and map control are crucial elements that contribute to team success.
– `Laning and Farming`: The game typically begins with a laning phase, where players distribute themselves across different lanes on the map. Each lane consists of three towers for each team, and players must last-hit enemy creeps to earn gold and experience. Efficient farming and denying enemy creeps are essential for gaining an advantage in the early game.
– `Items and Economy`: Dota 2 has a vast array of purchasable items that enhance heroes' abilities and attributes. Players earn gold by killing enemy units, heroes, or structures, which they can then use to purchase items from the shop. Itemization is crucial for heroes to become stronger and adapt to different situations during the game[4].
– `Map and Objectives`: The Dota 2 map consists of three main lanes (top, middle, and bottom) connected by a jungle area. Each team's objective is to destroy the enemy's Ancient, a heavily fortified structure located in their base. Along the way, teams can also aim to take down towers, secure Roshan (a powerful neutral boss), and claim other objectives to gain an advantage.
– `Patches and Meta`: Dota 2 receives regular updates in the form of patches, which introduce changes to heroes, items, and gameplay mechanics. These updates often result in shifts in the meta, influencing the preferred hero picks, strategies, and playstyles adopted by professional players and the wider player community.
– `Draft Stage and Captains Mode`: In competitive play, before the actual match begins, teams engage in a draft stage where they strategically select their heroes. Captains Mode is the most commonly used draft mode in professional Dota 2 matches. It involves a series of alternating picks and bans, where team captains take turns selecting heroes to form their team composition while also denying certain heroes from the opponent. The draft stage plays a crucial role in setting the foundation for team strategies, synergies, and counterpicks, as the chosen heroes greatly influence the dynamics and potential outcomes of the game.

3 DATASET

In this study, we utilized a comprehensive dataset taken from Kaggle comprising over 150,000 professional matches from the popular online multiplayer game Dota 2. The dataset covers matches that have taken place since June 19, 2011, enabling us to analyze a substantial period of competitive Dota 2 gameplay.

Our selection of this dataset was deliberate, as it allowed us to concentrate on matches featuring skilled and experienced players. By specifically focusing on professional matches, we aimed to capture the highest level of gameplay and strategic decision-making within the Dota 2 community. This emphasis on skilled players ensures that the dataset reflects the nuances and intricacies of competitive play, providing valuable insights into advanced strategies, hero picks, team compositions, and match outcomes.

3.1 Missing Values Analysis

During our analysis of the Dota 2 Matches (Pro Leagues) dataset, we conducted a missing values analysis. As part of this analysis, we excluded matches that did not have a recorded winner value. These matches accounted for approximately 0.98% of the dataset. By excluding these matches, we ensured that our analysis focused solely on matches with a clear winner.

Furthermore, we observed that approximately 69% of the data frame contained one or more missing values. These missing values could be present in various fields within the dataset, such as player statistics, match details, or other relevant attributes. Taking a closer look at the missing values by years, it was evident that number of records with missing values decreased as the year was closer to present. This suggests that the data collection process or data management practices improved over time, leading to more complete and reliable data from that point onwards. So we decided to work on data from January 1, 2023, which is both more relevant and accurate.

3.2 Feature Transformation and Selection

Prior to making feature selection and analysis, we converted categorical variables representing hero roles, positions and match outcomes into numeric representations. By performing the feature transformation through mappings, the dataset is prepared for further analysis or modeling tasks that may require numeric inputs.

For features selection we employed two techniques: correlation analysis and information gain. Considering correlation analysis, we determined the level of correlation using a threshold of 0.85 at which features would be considered redundant. Surprisingly, after careful evaluation, we found that all of the features in our dataset met the threshold and therefore no features were dropped. This suggests that all features in our dataset were relatively independent and made unique contributions to the analysis.

In addition to correlation analysis, we also employed information gain to assess relevance of features. As a result, we have chosen the top 25 features out of 52 overall, namely dire kills and radiant kills are found to be most informative features for our analysis. After selecting the top features, we further reduced dimensionality using PCA. After using PCA, 75% of variance was explained by 5 selected components.

For a given dataset X, we want the K largest directions of variance that are all mutually orthogonal. For finding K-th largest direction of variance, we have the following optimization problem:

$$\begin{aligned} \min_{v} \quad & v^T \Sigma v \\ \text{s.t.} \quad & \|v\|_2^2 = 1 \\ & v^T v_i = 0, for\ i = 1, 2, ..., K-1 \end{aligned} \tag{1}$$

Where $\Sigma = \frac{1}{M}X^TX$. As a result, this problem can be solved into this: $\Sigma v_K = \lambda v_K$. This means that the top K directions of variance $v_1, ..., v_K$ are given by the K eigenvectors corresponding to the K largest eigenvalues of $\frac{1}{M}X^TX$ [5]

3.3 Dropping Unnecessary Values

To ensure the quality and relevance of our analysis, we performed a thorough examination of the dataset and identified unnecessary values that were not pertinent to our research objectives. So we dropped continuous categorical features such as match_id, match duration and league_id.

4 METHODOLOGIES

In this sections, methods that we used upon the development of our model will be described. We used K-means clustering to cluster heroes for different positions as well as association analysis and correlation between various fields.

4.1 Correlation

For correlation the main field for studying will be the player net worth. First, we selected matches within first 4 months of 2024, because dataset is large and the results will be overly comprehensive. We will find correlation between net worth of a player and match duration in minutes. In Dota 2, net worth as a main index for player's farming outcome is the sum of their gold and all of the items they own. Larger net worth implies bigger chance for creating impact on a game. As game moves on, all the players on a map receives gold, however some heroes have abilities that allow them to gain net worth faster than others. This means that net worth of a player grows with every minute of the game. Furthermore, winning team in majority of circumstances have net worth than the team that is losing. We have created a graph that illustrates how a player's net worth varies with match duration and divided players into two categories - win and lose.

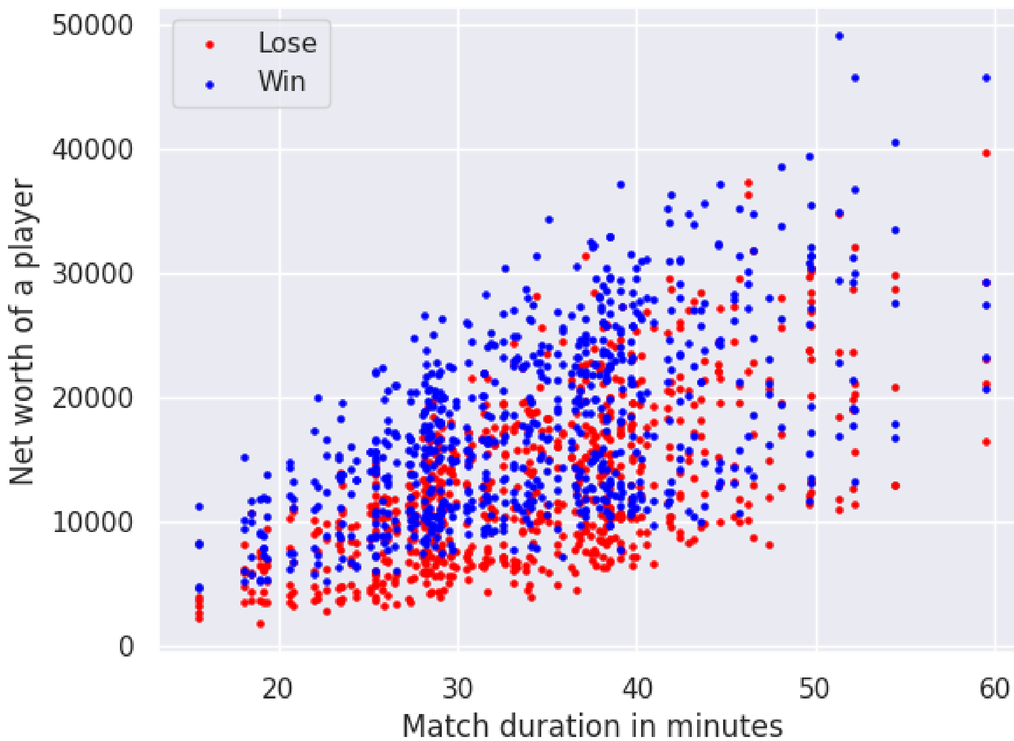

Fig. 1. Correlation between net worth of player and match duration depending on game outcome.pl

The observation from figure above supports the statement we claimed before. Now we want to depict the correlation between same fields, but for different positions. As we know, there are 5 positions in Dota 2 which indicate the farming and gold receiving priority of a player within a team. Then, we need to calculate the Pearson's correlation coefficient for 5 correlations. The results are illustrated below.

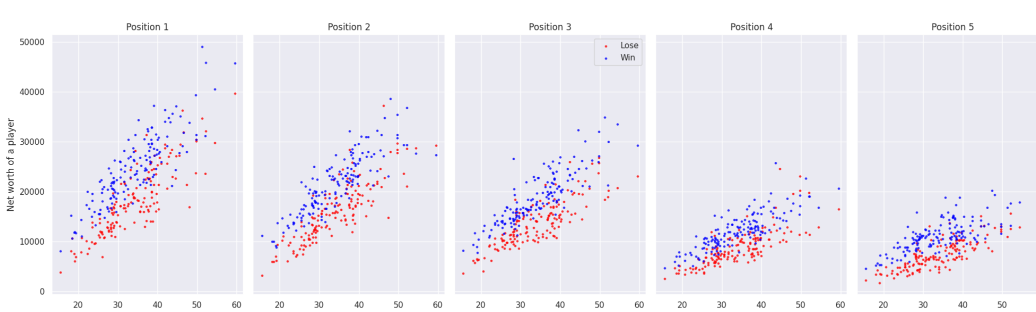

Fig. 2. Correlation for different positions.

The graph above shows that Position 1 players have higher net worth than Position 2 players and so on. Calculated Pearson's Correlation coefficient values are: 0.779 for position 1, 0.774 for position 2, 0.750 for position 3, 0.725 for position 4, and 0.662 for position 5. Those values and graph tells us that correlation becomes smaller with higher position.

4.2 Association Analysis

For Association Analysis we need to find which pair of heroes occurs more often than other using "`support`" metric. First, we discovered which heroes are the most chosen. We filtered matches so only matches played in 2023 are counted. After that we created a bar chart:

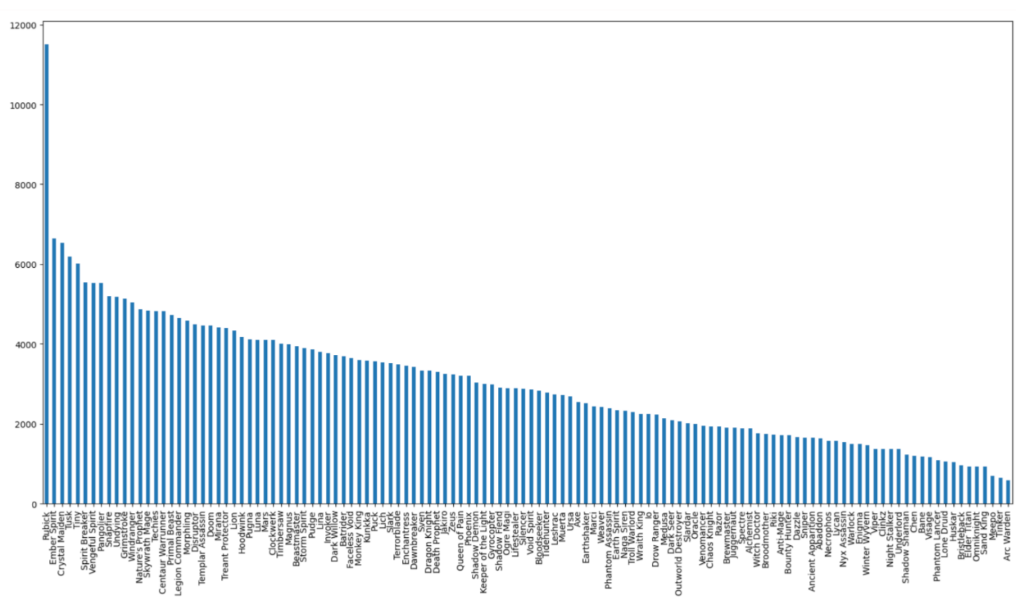

Fig. 3. Hero pick frequencies

From bar chart above we see, that most picked hero is Rubick with more than 11,000 mathces played. The difference between the second most picked hero Ember Spirit is enormous and the least picked hero is Arc Warden with less than 1,000 matches played. After that, we used Apriori algorithm to identify which pair of heroes are most frequent and to build association rules by using measures such as confidence and lift. First we selected minimal support to be 0.005 and used only rules with lift at least 1.5. The reason that we use lift, is because if we are to rely on confidence only, most of the association rules will be with Rubick, because it is the most picked hero. Lift allows us to see how much better a rule is at predicting the result than just assuming the result in the first place. The results are shown in a table below:

Table 1. Apriori Association Rules

antecedents	consequents	confidence	lift
Centaur Warrunner	Grimstroke	0.110	1.578
Grimstroke	Centaur Warrunner	0.105	1.578
Clockwerk	Snapfire	0.108	1.605
Snapfire	Clockwerk	0.085	1.605
Crystal Maiden	Pudge	0.087	1.648
Pudge	Crystal Maiden	0.146	1.648
Death Prophet	Tiny	0.166	2.098
Tiny	Death Prophet	0.090	2.098
Death Prophet	Tusk	0.143	1.763
Tusk	Death Prophet	0.075	1.763
Mirana	Ember Spirit	0.132	1.531
Ember Spirit	Mirana	0.088	1.531
Faceless Void	Snapfire	0.103	1.525
Snapfire	Faceless Void	0.075	1.525
Legion Commander	Skywrath Mage	0.183	2.871
Skywrath Mage	Legion Commander	0.178	2.871
Magnus	Techies	0.099	1.534
Techies	Magnus	0.082	1.534
Storm Spirit	Pugna	0.111	1.960
Pugna	Storm Spirit	0.100	1.960

We have 20 association rules and there is only 10 pairs of heroes which have high confidence and lift, meaning those 10 pairs of heroes are frequently picked together and have a higher chances of winning. The pair with highest lift is Tiny and Death Prophet, meaning these heroes have abilities that interact with each other and can deal more impact in synergy with each other.

4.3 K-means Clustering

In our report, we use K-means clustering to classify heroes into two main roles - cores and supports. Core heroes are the ones that receive most gold and experience, purchase damage dealing and durability items. In Dota 2, cores are players of Postion 1, 2, and 3. While supports are heroes that do not require huge amount of gold and have abilities that allow to help other teammates. Usually positions 4 and 5 are marked as supports. For our clustering we will analyze fields such as player's kills, deaths, assists, and net worth. Our main objective is to group players by their played heroes and calculate average indexes for the attributes for a corresponding hero. For simpler visualization we combined attributes Kills, Deaths, Assists into one new attribute KDA (Kills Deaths Assists):

$$KDA = \begin{cases} Kills + Assists, & \text{if } Deaths = 0 \\ \frac{Kills+Assists}{Deaths}, & \text{otherwise} \end{cases} \tag{2}$$

To visualize this K-means clustering we use two dimensions - Average Net Worth and Average KDA for each hero. In order to do that, we created a new data frame to store calculated average indexes and illustrate it using plot:

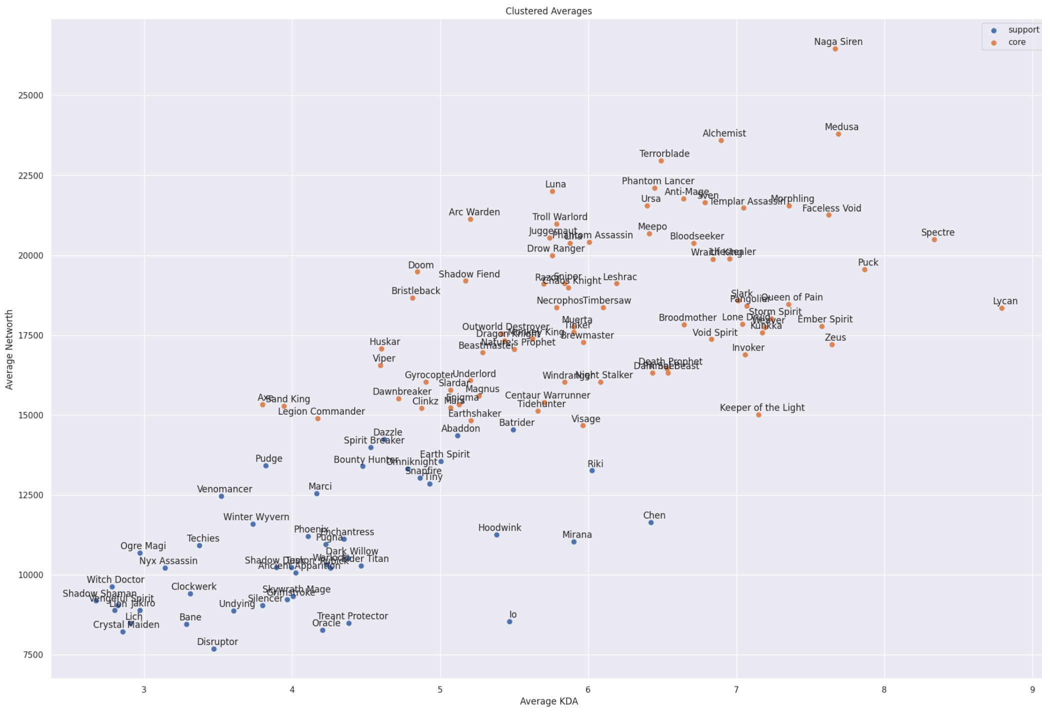

Fig. 4. Correlation for different positions.

As a result, we have two clusters of cores and supports. Overall, there are 46 support characters and 78 core characters. This clustering means one or the other hero is usually picked as core or support, thus the player should have an appropriate farming priority for specific hero.

5 RESULTS

In this section we evaluate the results of our model. How it predicts the data from our training and testing sets, and how the model performs for hero drafts in typical matches outside of dataset.

5.1 Model Comparison

We successfully developed a model and divided our chosen dataset into training and testing sets. 75% of dataset is used for training and 25% of dataset is used for testing. Since we want to find probability of win we use both generative and discriminative classifiers such as Naive Bayes Classifier, Linear Discriminative Analysis (LDA), and Logistic Regression.

Gaussian Naive Bayes and Linear Discriminant Analysis share the same approach of approximating Optimal Bayes Classifier: given an input vector x = $[x_1, x_2, \ldots, x_N]$, $x_i = [x_i^1, x_i^2, \ldots, x_i^M]^T$, pick the class c $c \in \{1, 2, \ldots, C\}$ with the largest posterior probability $p(y = c|x) = \frac{p(x|y=c)p(y=c)}{p(x)}$ (in our case - $c \in \{1, 2\}$ with 1 being Radiant Win and 2 is Radiant Loss). In order to achieve this, we need to be able to calculate $p(x|y = c)$.

Gaussian Naive Bayes approximates it as following:

$$p(x_i|y = c) = \frac{1}{\sqrt{2\pi\sigma_c^2}}\exp(-\frac{(x_i - \mu_c)^2}{2\sigma_c^2}) \tag{3}$$

Where μ_c is the average of i-th feature and σ_c^2 is the variance of i-th feature.

And here is how Linear Discriminant Analysis does so:

$$p(x|y=c) = \frac{1}{((2\pi)^N \Sigma)^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu_c)^T \Sigma^{-1}(x-\mu_c)) \tag{4}$$

Where $\mu_c = [\mu_1, \ldots, \mu_N]$ is the averages of features for the given class c and Σ is the shared covariance matrix of features.

We can clearly see that Gaussian Naive Bayes isolates each feature to calculate the likelihood while LDA considers features altogether to make a decision.

Logistic regression can classify based on the given input x = $[x_1, x_2, \ldots, x_N]$, such that:

$$Prediction = \begin{cases} +1, & \frac{p(y=+1|x=[x_1,x_2,\ldots,x_N])}{p(y=-1|x=[x_1,x_2,\ldots,x_N])} > 1 \\ -1, & \text{otherwise} \end{cases} \tag{5}$$

Where $+1 \iff$ Radiant win, $-1 \iff$ Radiant loss, $p(y=+1|x=[x_1,x_2,\ldots,x_N]) = \sigma(f(x))$ and $p(y=-1|x=[x_1,x_2,\ldots,x_N]) = 1-\sigma(f(x)), \sigma(z) = \frac{1}{1+e^{-z}}$ (Sigmoid function) and$f(x) = w^T x + b$ (Linear function). [5]

As we can observe, these models operate within a probabilistic framework. They model the conditional probability of each class given the input features to make predictions. This concludes the reason for having probabilistic prediction outcome of the match, rather than discretized prediction.

Training of the aforementioned models is straightforward : given the list of hyperparameters, tune the model to maximize the accuracy. We received valuable insights into the strengths and weaknesses of each model and compared them in the table below:

Table 2. Model evaluation & comparison

	Gaussian Naive Bayes	LDA	Logistic Regression
Accuracy	0.939	0.949	0.946
Precision	0.930	0.949	0.946
Recall	0.951	0.949	0.947
F1-Score	0.940	0.949	0.946
ROC AUC Score	0.982	0.949	0.987

5.2 Prediction of outcomes of matches

The observation from Model Comparison is that our model is dealing with the matches from dataset almost perfectly. Given that features are accurately chosen and correctly transformed, model performs best. However we need to predict outcome of the matches before the game starts, meaning without provided features. That is why we need to generalize data by working with averages on each hero and on each position. For that, we firstly created a new data frames for each divided by time frames of match durations. The created data frames are for matches of duration between 0 and 20 minutes, 20 and 30 minutes, 30 and 40 minutes, 40 and 50 minutes, 50 and 60 minutes, and for 60 minutes onwards. Then, we select the average indexes for each hero and on each position for every created data frame. As a result, model accepts hero drafts of 10 heroes for 2 teams as an input, and outputs winning probabilities of 2 teams for each divided time frame.

Table 3. Prediction Accuracies

Match Duration interval (min)	0-20	20-30	30-40	40-50	50-60	60+
Accuracy	0.717	0.582	0.538	0.553	0.566	0.500

After an evaluation of outcome prediction at different times, our model performed with an overall accuracy of 0.56, recall of 0.52, precision of 0.51, and F1-Score of 0.52. There are a lot of reasons for model to perform this way which will be discussed further.

6 DISCUSSION

6.1 Limitations

- **Dota 2 changes**: Dota 2 is characterized by its dynamic nature, as it undergoes frequent changes facilitated by Valve through the release of patches and updates. These updates are released at varying intervals, typically ranging from 10 to 200 days. Each update brings about alterations in the power levels of heroes, item effectiveness, and average game duration. Furthermore, Valve introduces new heroes and items into the game through these patches. The aforementioned changes have a substantial impact on the game, occasionally rendering certain heroes ineffective while bolstering the strength of others, thereby disrupting the balance. Consequently, relying on outdated data is both impractical and erroneous, as its relevance diminishes over time. Consequently, utilizing solely top-tier professional games restricts the dataset, resulting in a limited pool of training data. Conversely, lower-tier professional games and average non-professional matches fail to meet the requirements for comprehensive analysis and evaluation.
- **Human factor**: A significant factor to consider is the inherent human element in Dota 2 gameplay, which poses a challenge for predictive models. Our model assumes that gamers play ideally, without making mistakes. However, even at the highest level of professional play, players are susceptible to making errors that can potentially lead to losing matches. From an analytical perspective, it is often impossible to account for and predict these specific mistakes, as they can arise from split-second decisions, miscommunication, or lapses in judgment. Consequently, the assumption of flawless gameplay overlooks the unpredictable nature of human performance and the potential impact of mistakes on match outcomes.Another important consideration is the variability in individual player performance throughout a tournament, which can be influenced by factors such as sleep, health issues, or other personal problems. While professional gamers strive for consistency, external factors can affect their individual performance levels from game to game. These fluctuations in performance introduce additional complexity when analyzing and predicting match outcomes.
- **Assumption to use averages as an input alone for prediction model is sufficient**: While averages can offer valuable insights into general trends and patterns, they may not capture the full complexity and nuances of the game. This model does not take into account counterpicks and synergies between different heroes, and it does not fully understand the intricate interactions and strategic decision-making that occur during drafts and gameplay.
- **Lack of data after filtering missed values**: A potential challenge arises from the lack of data resulting from the filtering of missed values in Dota 2 analysis. Given the vast amount of available data, it is common practice to exclude matches or instances with missing or incomplete information from the dataset. While this filtering process improves data quality, it can inadvertently reduce the overall data volume, potentially leading to a narrower and less representative sample. Consequently, the exclusion of matches with missing values may

result in a limited dataset, which can impact the reliability of analysis and predictions. Careful consideration and appropriate techniques for handling missing data are necessary to mitigate this limitation and ensure the integrity of the analysis.

6.2 Anomaly detection

We also deeply studied the dataset to find any anomalies. When we analyzed Match Duration Distributions we found out that the graph have a normal distribution, where majority of games are finished in between 30 and 40 minutes. After that, we illustrated plots for matches with duration of over 90 minutes and for matches with duration of under 10 minutes. The second one showed us an interesting anomaly:

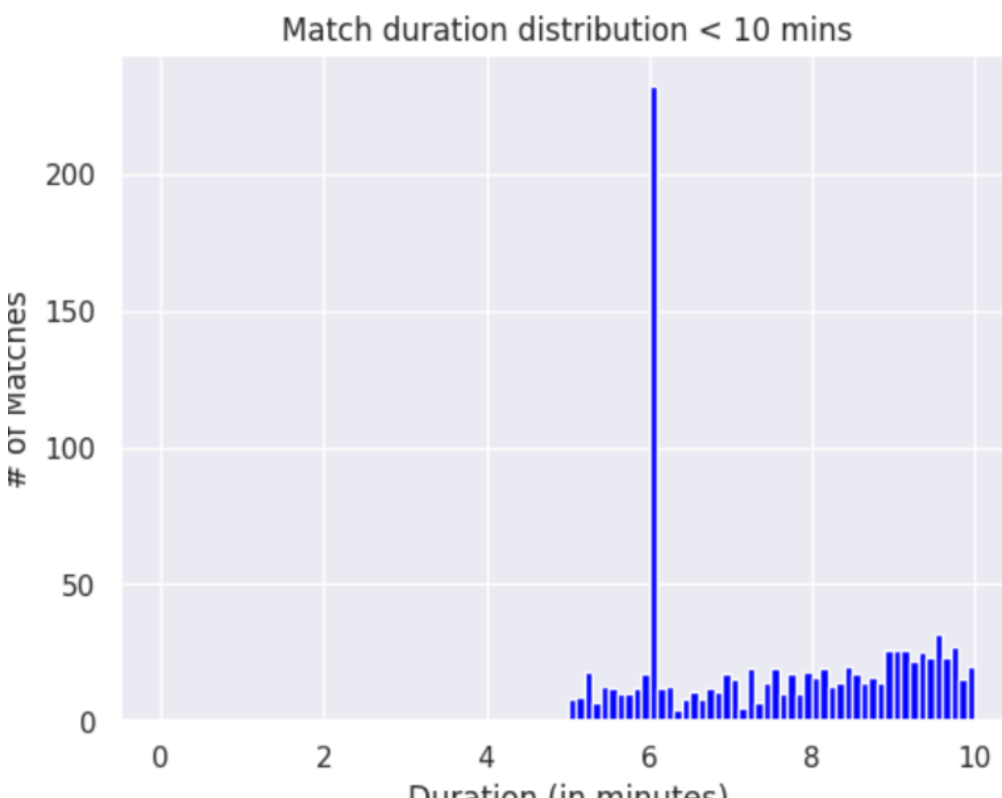

Fig. 5. Match Duration Graph

From figure above we can see that there are more than 200 games with exactly 6 minutes in duration, which is odd comparing with other values that do not even reach 50.

6.3 Future Improvements

- **Taking into account counterpicks and synergies between heroes**: Enhancing the prediction model by incorporating knowledge of hero counterpicks and synergies can significantly improve its accuracy. Understanding the interactions between heroes and their strategic implications during the draft phase would provide valuable insights into match outcomes.

- **Analyzing usual matchmaking of higher-ranked players**: Expanding the dataset to include matches from higher-ranked players in the usual matchmaking system can enhance the model's training data. This would provide a more diverse range of gameplay scenarios and help capture a broader spectrum of strategies and patterns.
- **Utilizing deep learning models**: Utilizing deep learning models: Deep learning models, such as neural networks, have proven effective in capturing complex relationships in various domains. Applying deep learning techniques to Dota 2 prediction models can help uncover intricate patterns and dependencies within the game's dynamics, leading to more accurate predictions.
- **Exploring alternative methods of approximation**: Instead of relying solely on average values, exploring alternative methods of approximating and aggregating data can improve the prediction model's effectiveness. Techniques such as weighted averages, Bayesian statistics, or regression models could be considered to better capture the nuances.
- **Connecting to the Dota 2 API**: Integrating the model with the Dota 2 API to access real-time data and predict matches based on the current patch would enhance the model's relevance. By incorporating up-to-date information on hero balance changes, item updates, and gameplay adjustments, the model can provide more accurate predictions aligned with the current state of the game.

7 CONCLUSION

Despite the inherent difficulty in predicting match outcomes based solely on the draft phase of Dota 2, our analysis has aligned with our understanding of the game. By studying and analyzing all professional games over the past years, we have gained valuable insights into the factors that have the most significant impact on match outcomes. As results show, methods we used were not enough to accurately predict match outcomes, meaning there are still some other factors that affect match outcome apart from raw data. So there is still much to be explored in the area. Additionally, our observations have allowed us to uncover an anomaly, a deviation from the expected patterns, which has deepened our understanding of the game even further.

8 ACKNOWLEDGEMENTS

These are the contributions of each group members of group 9:

Table 4. Work Contributions

Name	EID	Contribution
SANIYAZOV Dias	dsaniyazo2	20%(Clustering, Report)
ABYLKHANULY Mustafa	mabylkhan2	20%(Association, Report)
AIDARBEK Yernur	yaidarbek2	20%(Presentation, Report)
ALDEKEN Nur	naldeken2	20%(Model development and deployment)
YERMEKOV Beket	byermekov2	20%(Data transformation, model development)

Our code can be accessed within a zip file for submission.

9 REFERENCES

[1] Team Spirit (n.d.). Retrieved April 20, 2024 https://liquipedia.net/dota2/Team_Spirit
[2] True Sight (2022, September 24), The International 2021 Finals https://www.youtube.com/watch?v=NFEwN1N3vvA
[3] Hawk (2024, February 19), How many heroes in Dota 2? And Who's Next? https://hawk.live/posts/how-many-heroes-dota-2
[4] Sengupta, S. (2020, October 13), Dota 2 Items Guide: Best items and how to get them. Red Bull. Retrieved from, https://www.redbull.com/in-en/dota-2-items-tips-guide
[5] Discriminative & Generative classifiers, PCA (2023), Lecture 1-2, 2-1, 2-2, 8-1, CS4487 Machine Learning