DOTA 2 HERO SELECTION ANALYSIS

by

ZHAN GONG

A master’s thesis submitted to the Graduate Faculty in Liberal Studies in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York

2021

© 2021

ZHAN GONG

All Rights Reserved

Dota 2 Hero Selection Analysis

by

Zhan Gong

This manuscript has been read and accepted for the Graduate Faculty in Liberal Studies in satisfaction of the thesis requirement for the degree of Master of Arts.

|  |  |  |
| --- | --- | --- |
| Date |  | [enter full name here without titles]  Thesis Advisor |
| Date |  | [enter full name here without titles]  Executive Officer |

THE CITY UNIVERSITY OF NEW YORK

# ABSTRACT

Dota 2 Hero Selection Analysis  
 by

Zhan Gong

Advisor: Matthew K. Gold

This research explores how to choose heroes for the popular online eSport game Dota 2. Hero selection, also known as drafting, is so crucial to the game that carefully designed hero choices can “implicitly give a team a large advantage before the match even begins” (Conley and Perry 2013).

Different combinations of heroes in a team will have different interactions, ultimately yielding a variety of match outcomes. By combining data visualization and neural network machine learning, this project seeks to help players choose heroes to maximize a team’s likelihood of victory. I will start by introducing some game-play terms. I then visualize key statistics of heroes and interpret how to use these visualizations to choose heroes. Next, I use a basic neural network machine learning model to predict a team’s match outcome (win/lose) based on their hero choices. Future research should use a larger dataset to maximize the model’s accuracy and incorporate each hero’s performance (GPM, XPM, DPM, BDPM, and win rate) into the machine learning model.

# ACKNOWLEDGMENTS

I owe my deepest gratitude to numerous individuals for the selfless support they have provided as this capstone project has grown and evolved. I would like to thank my advisor, Dr. Matthew K. Gold, for his constant encouragement and guidance throughout my educational journey at the City University of New York, the Graduate Center. Without his invaluable feedback, this capstone project could not have reached its present form. I would also like to thank Jason Nielsen for his patience in answering all my questions and reminding me of all the important dates related to this project. My sincere acknowledgment also goes to my friends Demetrios Lambropoulos and Xuemeng Li, who provided me with insightful suggestions when I was building my neural network machine model. Finally, to my wife and my family, thank you for providing me with unfailing support and continuous encouragement throughout my two years of study. Thank you.

TABLE OF CONTENTS

[ABSTRACT 4](#_Toc70884184)

[ACKNOWLEDGMENTS 5](#_Toc70884185)

[CHAPTER 1: INTRODUCTION 1](#_Toc70884186)

[Lanes 2](#_Toc70884187)

[Positions 4](#_Toc70884188)

[Pick or Ban 5](#_Toc70884189)

[Key Statistics of Heroes 6](#_Toc70884190)

[CHAPTER 2: DATA AND METHODOLOGY 8](#_Toc70884191)

[CHAPTER 3: DATA VISUALIZATION 10](#_Toc70884192)

[Heroes’ GPM and XPM 11](#_Toc70884193)

[Heroes’ DPM and BDPM 14](#_Toc70884194)

[Heroes’ Win Rate 16](#_Toc70884195)

[CHAPTER 4: MACHINE LEARNING 18](#_Toc70884196)

[CHAPTER 5: CONCLUTION AND EVALUATION 24](#_Toc70884197)

[REFERENCES 25](#_Toc70884198)

# CHAPTER 1: INTRODUCTION

This research explores how to choose heroes for the popular online eSport game Dota 2. As a sequel to Defense of the Ancients (Dota), Dota 2 is a multiplayer online battle arena (MOBA) game developed and published by Valve Corporation. The game has the highest prize pool among all types of eSports competitions. As of November 15, 2019, the total prize pool of Dota 2 International has reached $34,330,068, making it the most lucrative game in E-sport (Dota 2 Prize Pool Tracker 2019).

The game is played in matches between two teams of five players; one team will play as Radiant, and the other team will play as Dire. Each team occupies and defends its respective base on the map, and the team that destroys the other team’s ancient building (known as “Ancient,” located within each team’s base) wins the game. Each player controls one of the 117 characters (known as “heroes”). Once a player chooses a hero, other players cannot select that same hero for the same match. Heroes have their unique strengths and weaknesses based on their skills and characteristics. Different combinations of heroes in a team will have different interactions, ultimately yielding a variety of match outcomes.

Hero selection, also known as drafting, is so crucial to the game that carefully designed hero choices can “implicitly give a team a large advantage before the match even begins” (Conley and Perry 2013). As Dota 2 is a team game where every player has to play specific roles according to every situation, picking good heroes for every role while ensuring these heroes fit well together is essential to guarantee a team’s victory (Dota 2 Freaks 2021). The majority of professional players already have a predetermined game plan before entering the drafting phase; for game analysts and commentators, predicting drafting decisions is one of their main topics during the broadcast of a match’s drafting phase (Summerville, Cook, and Steenhuisen 2016).

By combining data visualization and neural network machine learning, this project seeks to help players choose heroes to maximize a team’s likelihood of victory. I will start by introducing some game-play terms. I then visualize key statistics of heroes and interpret how to use these visualizations to choose heroes. Next, I use a basic neural network machine learning model to predict a team’s match outcome (win/lose) based on their hero choices. Finally, I evaluate my project and discuss what needs to be done in the conclusion section.

## Lanes

When a player starts to choose a hero, the player first needs to decide which lane he/she is trying to play. A lane is one of three paths (the Top line, the Middle line, and the Bottom line) that leads from one team’s Ancient (base) to the other (see Figure 1). When the game starts, the lane creeps (non-player characters) will push along these lanes (Dota 2 Wiki 2018). Players need to kill these creeps to obtain gold[[1]](#footnote-1) and experience points and destroy defense towers while approaching the enemy base in these lanes (eSportsGuide 2021).

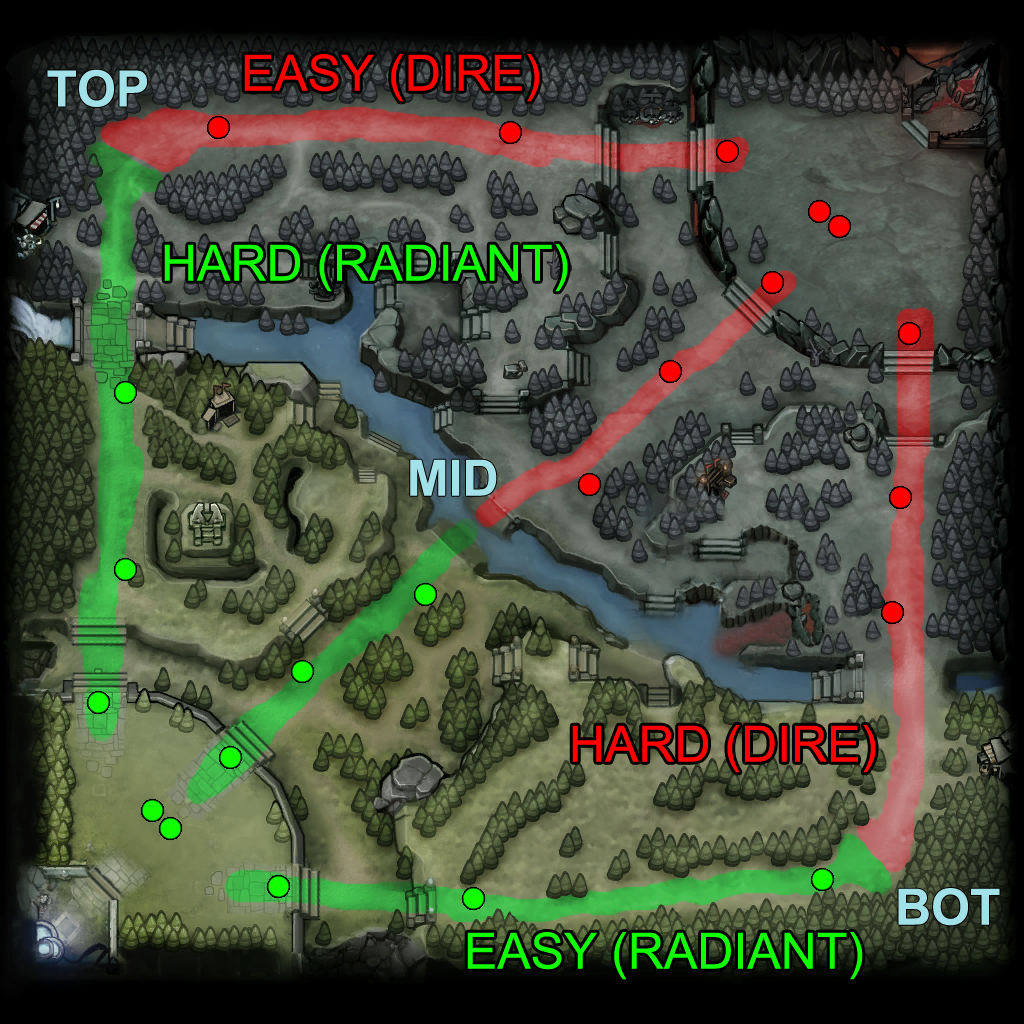


Figure 1: Lanes in Dota 2 (Dota 2 Wiki 2018)

Depending on the distance between the natural creep equilibrium (where the two creeps initially meet up) and the Tier 1 defense tower, these lanes are also known as the safe lane and the hard lane. The safe lane is the one with the shortest distance so that it is easy for a *laning[[2]](#footnote-2)* hero can retreat to their tower’s protection. By contrast, a team’s safe lane is the hard lane for the enemy team, as it is difficult for enemies to sneak through the jungle and *gank[[3]](#footnote-3)* the laning hero (Dota 2 Wiki 2018). For example, for team Radiant, the bottom line is their safe lane, whereas it is the hard lane for team Dire. In general, a team will choose a hero with better survival abilities for their hard lane and let their most important player play the safe lane.

## Positions

           There are five positions in each team, with 1 indicating that the player has the highest *farm*[[4]](#footnote-4) gold priority and 5 having the lowest. At the early stage of the game, a team usually has one player focusing on the safe lane, and this player is known as the Position 1 hero. Position 1 hero needs to have good farming abilities to gather enough gold quickly. With gold, Position 1 heroes can buy items that give the hero extra abilities. The hero will become the strongest hero later in the game if the team gains a significant gold advantage.

           To make sure the Position 1 hero farms safely and is not disturbed by the enemy, a team usually has a support hero following the Position 1 hero on the safe lane, whom we call the Position 5 hero. Because the amount of gold is finite, the support hero will not farm gold. Instead, the support hero usually picks more spells and uses their abilities to gain an advantage for the entire team.

           In the middle of the game, a team usually encounters a situation when the amount of gold is not enough to sustain its heroes. Thus, it is essential for the Position 4 hero to farm gold for the team when other heroes have no time to acquire gold in the mid-game. Position 4 heroes are also responsible for supporting heroes in other lanes and *ganking* the enemy.

           Position 2 heroes usually play the middle line. As the middle line is usually a 1v1 situation, Position 2 heroes need strong skills to dominate a 1v1 situation and can farm enough gold to upgrade skills quickly. Heroes in this position are chosen for their farming abilities and *ganking* abilities, which are reflected in their GPM (gold per minute) and XPM (experience per minute). Due to these reasons, there are not many options for Position 2 heroes.

           The last position is Position 3. Position 3 heroes usually play the hard lane. Sometimes, the Position 4 support hero will help the Position 3 hero in the hard lane; more often, the Position 3 hero is solo in the hard lane. Therefore, the Position 3 hero’s primary focus is to survive and farm gold when possible. Heroes in this role are required to have a set of survival skills and can be upgraded by the hero’s items.

## Pick or Ban

During a tournament match, the two teams alternate to Pick or Ban a hero in a precise predetermined order, and this process is known as drafting (Summerville, Cook, and Steenhuisen 2016). If a hero is banned by one team, this hero cannot be picked by either team. Similarly, if a hero is picked by one team, the other team cannot pick the same hero. The Pick or Ban order is shown in Figure 2.



Figure 2: The order of picks and bans in Captain’s Mode as of version 7.28c. The draft progresses from left to right. A column marked ‘B’ indicates a Ban, while ‘P’ indicates a Pick. Note that the team with second pick picks two heroes in a row, and also gets to pick last (Summerville, Cook, and Steenhuisen 2016).

## Key Statistics of Heroes

Hero statistics — heroes’ abilities, DPM and BDPM, GPM and XPM, and win rate — are key to help players understand Dota 2 and improve their game and will be explained below.

1) Heroes’ Abilities

From heroes’ abilities, we can have a general idea of each hero’s position in Dota 2. All five positions are indispensable during a game but picking a suitable hero for each position increases the team’s chance to win.

2) Heroes’ DPM and BDPM

Damage Per Minute (DPM) tells us how much damage a hero does to their enemy’s hero during the match, while Building Damage Per Minute (BDPM) tells us how much damage a hero does to their enemy’s tower. A hero usually plays the first or second position if the hero has a higher DPM. Also, DPM tells us how strong the hero’s spell could be. When picking heroes, if Position 1 and Position 2 heroes do not have enough damage, we usually choose a higher DPM hero for Position 4 or Position 5. On the contrary, if the Position 1 and Position 2 heroes have enough damage, we usually consider choosing a Position 4 or Position 5 hero with lower DPM but more disablers.

3) Heroes’ GPM and XPM

Gold Per Minute (GPM) represents the amount of gold a hero has farmed during the match. Experience Per minute (XPM) represents how much experience a hero has gained during the match. If a hero has the highest GPM and XPM, the hero usually plays Position 1 or Position 2. If a hero has the lowest GPM and XPM, the hero usually plays Position 4 or Position 5.

4) Heroes’ Win Rate

By visualizing heroes’ win rates, we can understand which hero is better for the newest game version. Dota 2 frequently updates or changes heroes’ skills and attributes to balance the game. Thus, understanding heroes’ performance in the newest game version is important. Choosing a hero who has a better win rate in most matches will also increase the likelihood of winning the game.

# CHAPTER 2: DATA AND METHODOLOGY

In this project, I first used Tableau to visualize heroes’ abilities, heroes’ DPM and BDPM, heroes’ GPM and XPM, and heroes’ Win Rate in Dota 2. I then used a neural network machine learning model to predict a team’s match outcome (win/lose) based on their hero choices. By combining data visualization and machine learning, this project hopes to provide a valuable resource for Dota 2 players when choosing heroes (please see more details in Chapter 3 and Chapter 4) .

My data was obtained from a website called dotamax.com. Dotamax gathers data from the Steam Web API and tracks every public game played in Dota 2, and all users can view this in-depth match data free of charge. Also, the website itself performs different visualizations for Dota 2 players to better understand the game. I extracted the data by using the power query function in Excel. I only used the tournament games data because I wanted to minimize players’ skill differences, and I assume that every hero has their best performance during the match. My data includes the following information for every hero:

* Name
* Ability/Attributes (Agility, Intelligence, and Strength)
* Gold Per Minute (GPM)
* Experience Per minute (XPM)
* Damage Per Minute (DPM)
* Building Damage Per Minute (BDPM)
* Win Rate
* Number of Matches the hero has been Played

I also collected data on 100 recent matches (2020/6/20-2021/1/1) played by a European tournament Dota 2 team named Team Secret. Team Secret is one of the most dominant teams in Dota 2 history. Specifically, the following information for each match is collected:

* Hero's name（I recoded the 117 heroes into numbers from 1-117）
* Choice Type (Pick/Ban) by Team Secret
* Choice Type (Pick/Ban) by the enemy team
* The Order of Picks and Bans
* Match outcome of Team Secret (0 = lost, 1 = won)

# CHAPTER 3: DATA VISUALIZATION

I used the software Tableau to perform data visualizations for this project. In Visual and Design Fundamental, a course I took in Fall 2019, I learned how to use Tableau to create interactive data visualizations. I chose Tableau because the software creates beautiful and easy-to-understand data visualizations, and these visualizations can be easily linked to HTML files. I wanted this project to look like a search engine: when a user clicks on a hero's profile, the hero's statistics (for example, a hero's XPM) will pop up on the right side of my visualization project. I mainly used bar charts to present these statistics. Bar charts are good for comparing categorical data, so it is easy for users to find out each hero's performance and compare between different heroes.

I created three sets of visualizations: Heroes' GPM and XPM, Heroes' DPM and BDPM, and Heroes' Win Rate; each will be discussed in detail below.

## Heroes’ GPM and XPM

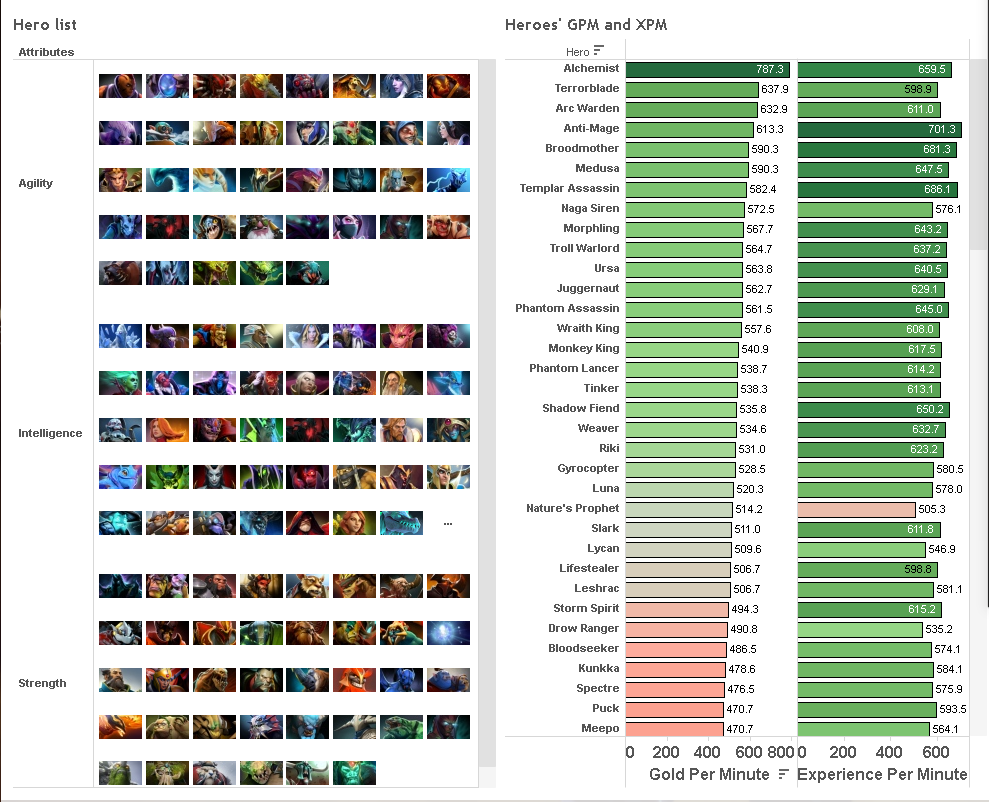


Figure 3: Heroes’ GPM and XPM Sorted by GPM

Figure 3 visualizes heroes’ GPM (Gold Per Minute) and XPM (Experience Per Minute) sorted by GPM, with green indicating higher values and red indicating lower values. In general, the higher GPM a hero gets, the better items this hero will be able to afford. Similarly, the higher XPM a hero gets, the stronger the hero’s skills will become.

Visualizing GPM and XPM can help us better decide the positions of heroes in the game. Heroes who have higher GPM and XPM are often placed in the 1st and 2nd positions for their farming abilities. By contrast, heroes with low GPM can never play the 1st or 2nd position. For example, the top 21 heroes (from Alchemist to Gyrocopter) in Figure 3 all have high GPM, making them ideal choices for the 1st and 2nd position, especially the 1st position, heroes.

When choosing heroes, a team needs to consider how to choose heroes to fill the five positions in the game, with each position having its unique focus. For example, as shown in Figure 3, the hero Alchemist has the highest GPM and XPM. Suppose our first four picks have already chosen 1st and 2nd position heroes. In that case, we cannot choose Alchemist as our 5th pick hero as this choice will make the 1st and 2nd position heroes have no gold to farm (the amount of gold in a game is finite, if Alchemist farms too much gold, the 1st position hero will have no gold to farm) and ultimately make the team lose this match.

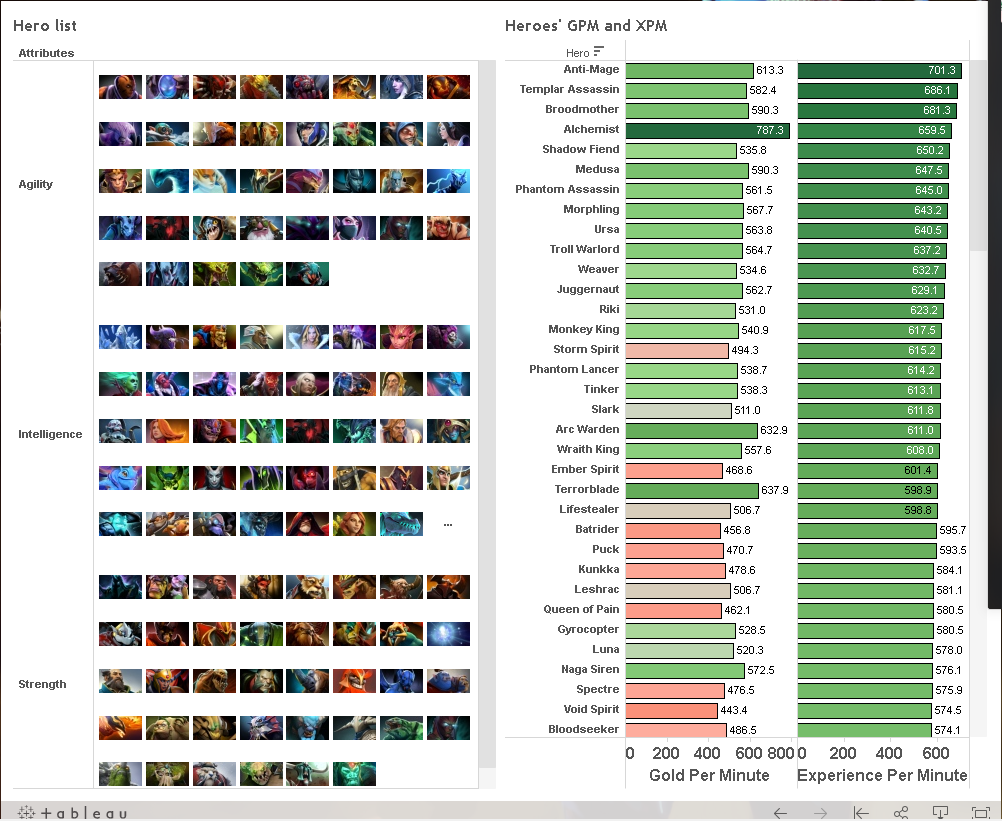


Figure 4: Heroes’ GPM and XPM Sorted by XPM

Figure 4 sorts heroes’ GPM (Gold Per Minute) and XPM (Experience Per Minute) by XPM, with green indicating higher values and red indicating lower values. From Figure 3, we can see that some heroes have good XPM performance but perform poorly in GPM (Storm Spirit, Ember Spirit, to name a few). In this case, heroes are best for playing the 2nd position, for their main role is *ganking* the enemy and participating in a team fight rather than farming gold (participating in team fights is the main method to earn XPM).

## Heroes’ DPM and BDPM

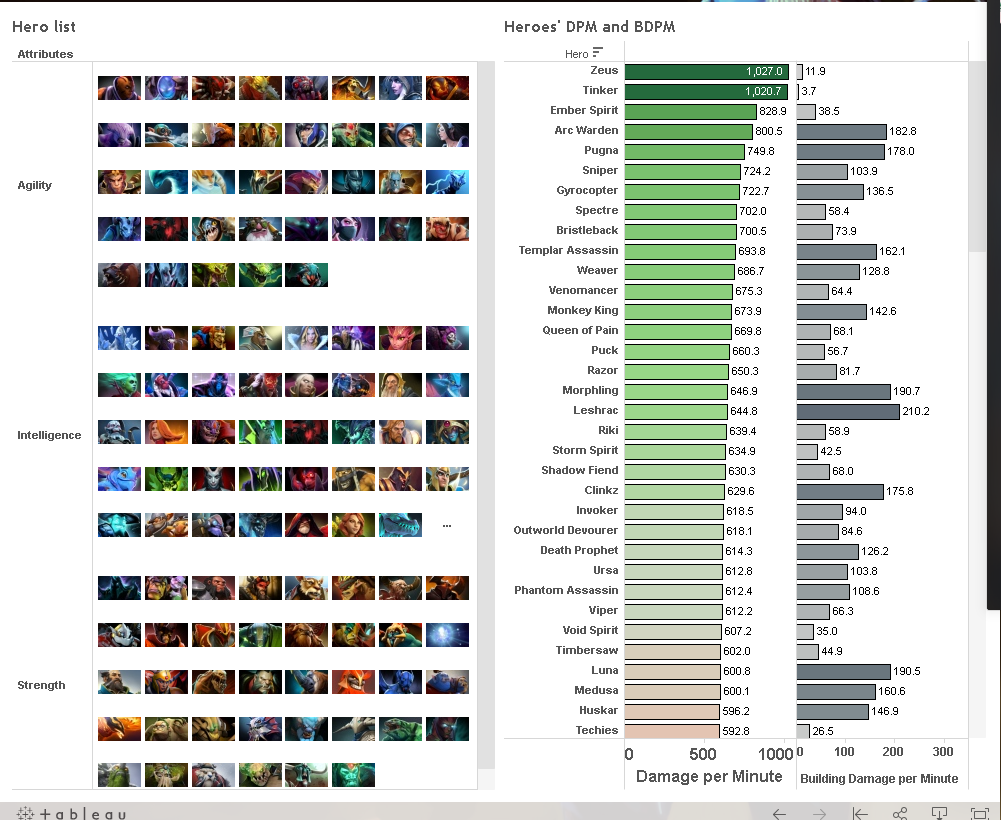


Figure 5: Heroes’ DPM and BDPM Sorted by DPM

Figure 5 shows how much damage a hero does to their enemy’s hero per minute (DPM) and how much damage a hero does to their enemy’s building per minute (BDPM), sorted by DPM. In general, the 1st and 2nd position heroes both have more DPM than other positions. Specifically, the 2nd position hero causes more damage than the 1st position hero, for the 2nd position hero participates more in team fights than the 1st position hero.

A team also needs to consider that some of the 1st or 2nd position heroes cannot cause damage to the buildings, such as Zeus and Tinker, who rank high in DPM but low in BDPM. In this case, we need to have a 3rd, 4th, or 5th position hero who can cause building damage to help other teammates destroy the enemy’s tower. This is why despite having the lowest DPM and very low GPM, Chen (not shown in Figure 5) is still a popular hero in the tournament matches for his relatively good performance in BDPM and has more than a 60 percent win rate.

To conclude, combining Figure 3, Figure 4, and Figure 5, we can easily find out which hero is suitable for which position. For example, a hero with three green bars for GPM, XPM, and DPM must be a Position 1 hero. A hero with just one or two green bars for GPM, XPM, and DPM could be a Position 2 hero. A hero with one or two light red bars for GPM, XPM, and DPM could be a Position 3 hero. A hero with all red bars for GPM, XPM, and DPM must be a Position 4 or Position 5 hero.

## Heroes’ Win Rate

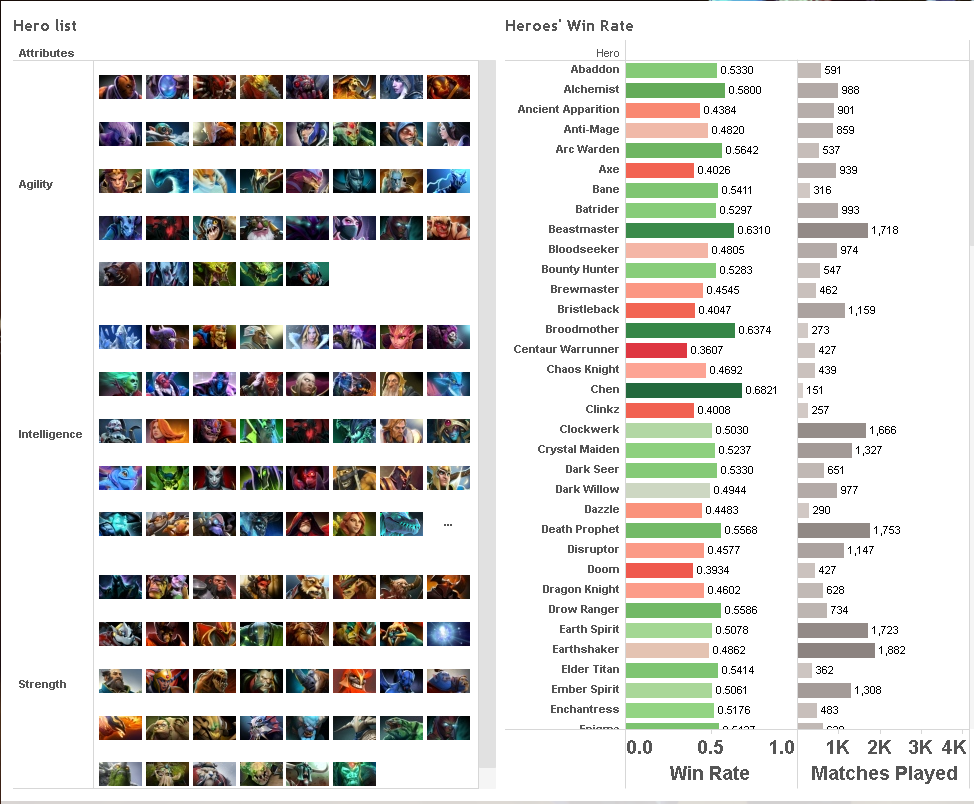


Figure 6: Heroes’ Win Rate

Figure 6 shows heroes’ win rate in tournament matches, and the game version is 7.28c. It should be noted that Valve updates Dota 2 frequently, bringing new content and changes to the game. For example, the game company would change heroes’ abilities or add new heroes to the game; some heroes may perform very well in the past version but perform poorly in the current or subsequent versions. Every single change gives the game environment a huge variation. Accordingly, heroes' win rate varies a lot in different versions.

Understanding a hero’s win rate in the current version is crucial, as it directly tells players how strong the hero is in the current game version. Players, in general, do not choose a hero with a low win rate. For instance, the hero Centaur Warrunner has a 36 percent win rate and has been played in 427 matches in version 7.28c. We assume that players have best performed this hero as all these matches are tournament matches. From Figure 2 and Figure 4, we can tell that Centaur Warrunner is a 3rd position hero. However, we also know that in the 7.28c version, picking this hero is not a good choice because somehow, this hero is not suitable for this version according to his low win rate. In short, the win rate is most useful when there are more than ten hero choices for a specific position. In this case, a higher win rate can provide a good reference.

# CHAPTER 4: MACHINE LEARNING

Apart from creating data visualizations, this project also used a basic neural network machine learning model to predict a team’s match outcome (win/lose) based on their hero choices. I learned the basic knowledge of machine learning from the course Advanced Data Analysis Methods. I realized that the neural network machine learning model was the proper method for this project for two reasons. First, I considered the entire draft, including picks and bans for both teams and the order of picks and bans, into consideration, and neural network machine learning could incorporate these factors into its learning process. Second, the purpose of this project is to predict if a team will or not based on some given match data, and neural network machine learning is a helpful tool to analyze the underlying patterns between the selection of heroes and the match outcome. A neural network is a series of algorithms that seeks to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates (Chen 2020). A simple neural network operates like this (Figure 7):

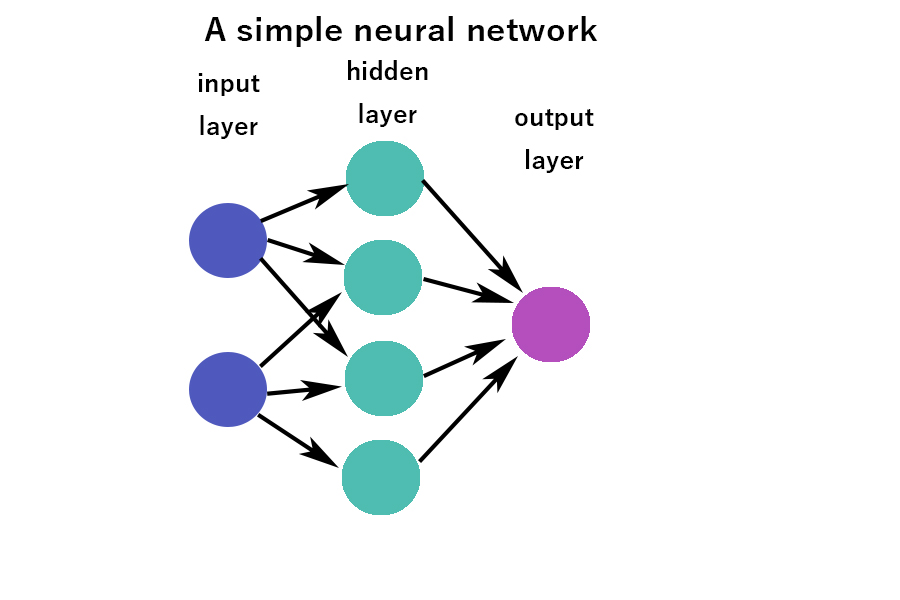


Figure 7: A Simple Neural Network

I used Pytorch API to build the neural network. Pytorch is a Python package that provides deep neural networks built on a tape-based auto grad system. Figure 8 presents how the past 100 matches played by Team Secret are coded. Columns 1 through 24 report the two teams’ pick and ban hero choices. I converted every hero’s name to a number as Pytorch API cannot deal with a string object. I encoded heroes from 1 to 117, and all the Nan values were coded as -1.

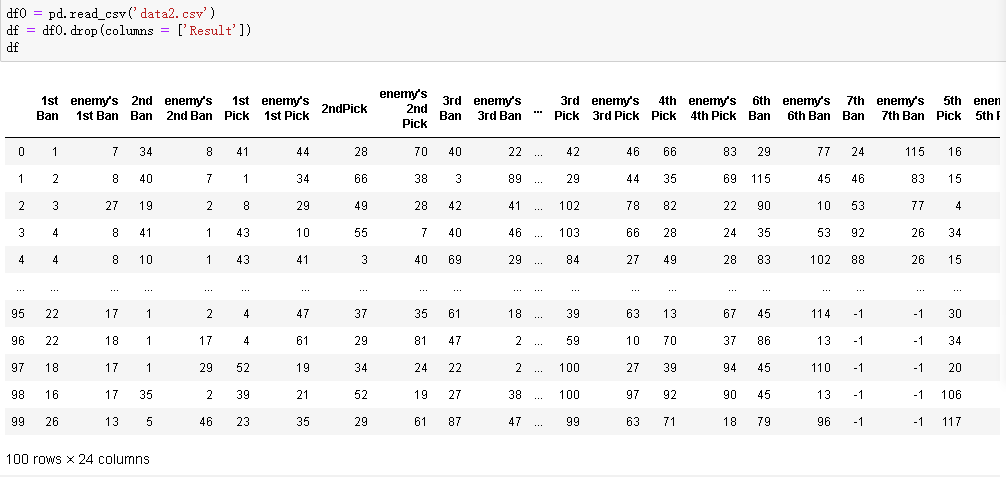


Figure 8: Data

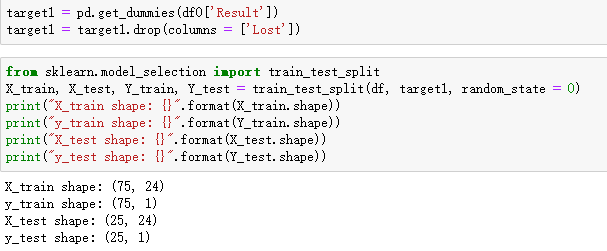


Figure 9: Split the dataset

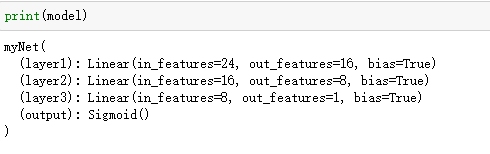


Figure 10: Net layers

Next, I split the data set into a training set and a test set (see Figure 9); because I have 100 past matches data, I got 75 training sets and 25 test sets. I created a network with 24 inputs, 16 first hidden layers, 8 second hidden layers, and 1 output (see Figure 10).

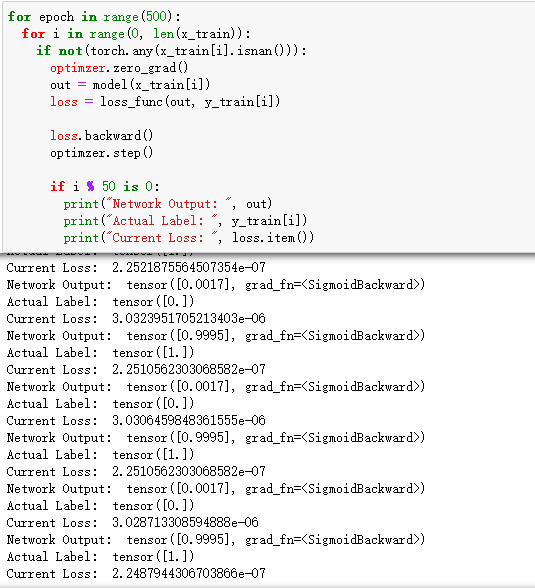


Figure 11: Training

Figure 11 shows the results of the neural network. We can see that the current loss[[5]](#footnote-5) is getting stable, suggesting the model has finished its training. Every output predicts a result of a win or a loss of a match played by Team Secret, and we can see if it is correct by comparing it with the actual game result.

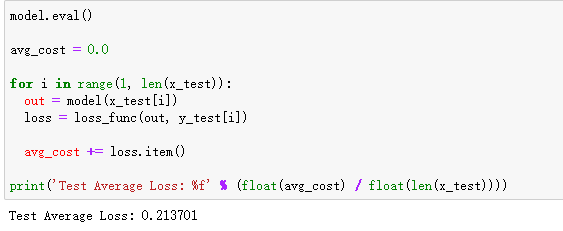


Figure 12: Test Average Loss

After training, I ran the training results on test sets and calculated the average loss using the average cost divided by the test set's size, and I got a high value of 0.21 (in general, the lower the loss, the better the model) (See Figure12). This is probably because my data set is very small; most of the heroes only showed up 1 or 2 times in the pick or ban sections.

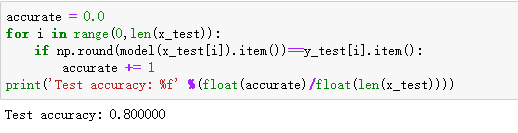


Figure 13: The Accuracy Test

I then tested the model accuracy using the number of matches that correctly predicted the match outcome divided by the test set's size (See Figure 13). The accuracy value is 0.80, which means the model has an 80 percent accuracy rate.

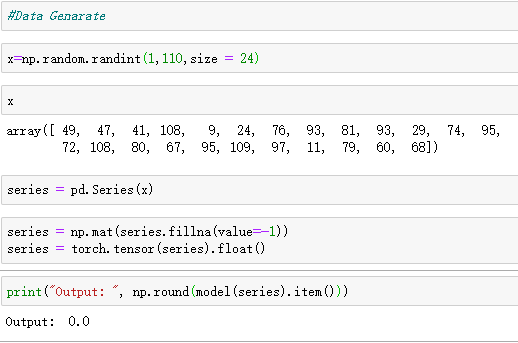


Figure 14: Testing

After building this model, I can test other matches by inputting their pick or ban choices into my model. The output of 0 means the model thinks Team 1 will lose the match; if the output is 1, it means the model thinks Team 1 will win the match.

# CHAPTER 5: CONCLUTION AND EVALUATION

This research explores how to choose heroes for the popular online eSport game Dota 2. First, I visualized heroes' key statistics, including GPM, XPM, DPM, BDPM, and win rate. Players can use these statistics as a reference when choosing heroes. For example, a hero with high values in GPM, XPM, and DPM could be a Position 1 hero. A hero with just one or two high values in GPM, XPM, and DPM could be a Position 2 hero. A hero with one or two low values for GPM, XPM, and DPM could be a Position 3 hero. A hero with all low values in GPM, XPM, and DPM could be a Position 4 or Position 5 support position hero. I then used a neural network machine learning model to predict if the selection of heroes will lead to winning or losing. Compared with previous research, which only accounted for one team's picking and banning choices, the model in this project also put the enemy team's picking and banning choices into consideration.

While the machine learning model's accuracy is high, one should note that the model is based on a relatively small dataset, which increases the margin of error. Future research should use a larger dataset to minimize the margin of error and maximize the model's accuracy. Moreover, future research should try to incorporate each hero's performance (GPM, XPM, DPM, BDPM, and win rate) into the machine learning model.

# REFERENCES

Chen, James. 2020. “Neural Network Definition.” *Investopedia.Com* 5. Retrieved April 30, 2021 (https://www.investopedia.com/terms/n/neuralnetwork.asp).

Conley, Kevin and Daniel Perry. 2013. “How Does He Saw Me ? A Recommendation Engine for Picking Heroes in Dota 2.” *Stanford University*. Retrieved April 30, 2021 (<http://cs229.stanford.edu/proj2013/PerryConley-HowDoesHeSawMeARecommendationEngineForPickingHeroesInDota2.pdf>).

Dota 2 Freaks. 2021. “Dota 2 Picking Guide: Draft Strategies for the Win! - Dota 2 Freaks.” Retrieved April 30, 2021 (https://dota2freaks.com/drafting/).

Dota 2 Prize Pool Tracker. 2019. “The International 2019 – Dota 2 Prize Pool Tracker.” Retrieved November 23, 2019 (<https://dota2.prizetrac.kr/international2019>).

Dota 2 Wiki. 2018. “Lane - Dota 2 Wiki.” Retrieved April 30, 2021 (<https://dota2.fandom.com/wiki/Lane>).

eSportsGuide. 2021. “Dota 2 Esports - Live Score, Streams, Schedule, News.” Retrieved April 30, 2021 (https://www.esportsguide.com/game/dota-2).

Summerville, Adam, Michael Cook, and Ben Steenhuisen. 2016. “Draft-Analysis of the Ancients: Predicting Draft Picks in DotA 2 Using Machine Learning.” *AAAI Workshop - Technical Report* WS-16-21-WS-16-23(Godec):100–106.

1. Gold is the currency used to buy items or instantly revive your hero in Dota 2. [↑](#footnote-ref-1)
2. *Laning* in Dota 2 means to continue to hold or push a lane so that you maintain or advance position on the map and gain experience. [↑](#footnote-ref-2)
3. *Gank* in Dota 2 refers to the action of leaving your lane during the laning phase and killing an enemy hero in another lane. [↑](#footnote-ref-3)
4. *Farm* in Dota 2 refers to the acquisition of gold. [↑](#footnote-ref-4)
5. Loss is a number indicating how bad the model's prediction is on a single example. If the model's prediction is perfect, the loss is zero. Otherwise, the loss is greater. [↑](#footnote-ref-5)