Data Mining in DOTA 2

# 1 Introduction

In this article I am going to explain how data mining and its main areas can help us to build much more reliable, enticing and balanced games. I am going to cover areas which mainly can be expressed by major areas in data mining such as classification, regression, clustering.

You might ask does it worth and the answer would be yes for sure. The stats in the next section will explain this idea to you why machine learning, in this article particularly data mining can save a lot of time/money/players for us which directly affects the budget for maintaining game and its artifacts.

In the next sections first, I will introduce Dota 2 and why this game is one of the greatest games ever existed, then we will go through its mechanics such as characters and as a follow up, we I will explain the matchmaking system (so many permanent bans here!) and finally we will finish our interesting thoughts by demonstrating servers.

# 2 Dota 2 Is Huge!

Dota is actually a map of the game called Warcraft by Blizzard back in 2002 which developed by a group of fans of the game where their leader is called IceFrog and still no one knows his real identity. Then this map got a lot of attention from users which basically, most of the people installed the game for this map. After a huge success, Valve, the current owner of Dota 2, acquired that team to create a standalone game based on the same mechanics/logics/map of original dota map which now it’s been called Dota 2 which has been announced in 2013 as a beta release (still in beta!!!).

In this section, I will not talk about complexity of the game as the reason that this game is big. The reason is money and player pool which by following stats you will be shocked. But to put complexity in another side, just remember that Dota 2 is the most complex game in the entire Esport scene in all genres and you might argue about StartCraft II but remember that in Dota 2, %99 of times, you only control one unit while in StarCraft II you need to control hundreds of different types in real time.

## 2.1 Player Pool

Dota 2 has about 900 thousand players on its peak and at least 300 thousand players in the hours with lowest traffic[1]. So, you cannot find a time in a day and in a month and in a year to see an empty player pool. Hence, more players, more data, more confliction, and of course what Valve cares most, more MONEY!



Figure 1 Dota 2 Player pool

Every month, on average, about 10 million new players join dota since it has been released as it is one of the most famous games in the entire Esport community.

## 2.2 Complexity

Many new users join but also many of them will leave game soon due to super complexity of game which literally they will be dead %90 of the time or get reported by other players which leads to account ban or restriction. Just I want to point that the main issue is that people do not study before doing something, as a programmer, I have seen people use a code without reading docs first, so obviously I will report them for toxic behavior!

Furthermore, OpenAI, one of the biggest AI research facilities in the world has created an AI for StarCraft II and Dota 2 which in the latter it is called OpenAI 5 using Deep Reinforcement Learning. [2]

As a summary for OpenAI timeline, it first created a single bot to fight 1v1 against best dota players in the world and it literally destroyed Dendi (the *face of Dota*) and other pro players 2-0 (best score).



Figure 2 Dota 2 1v1 (Dendi vs. OpenAI)

At the end of 2019, OpenAI 5 won 2 back to back games against OGEsposts, the “The international” winner of 2018 and 2019. You can watch matches here [3] and also, here is the stat shared by OpenArena, the AI team that is playing consistently against humans which demonstrates %99.4 win rate! [4]



Figure 3 OpenAI Five stats in pub games

But about power, I will say nothing, this text has been directly adopted from OpenAI Five blog:  
“OpenAI Five plays 180 years worth of games against itself every day, learning via self-play. It trains using a scaled-up version of [Proximal Policy Optimization](https://blog.openai.com/openai-baselines-ppo/) running on 256 GPUs and 128,000 CPU cores — a larger-scale version of the system we built to play the much-simpler [solo variant](https://blog.openai.com/dota-2/) of the game last year.” [5]

## 2. 3 Business

As I previously mentioned, more players more money, but how much?

Dota 2 has different type of competitive scenes such as amateur, professional and major which major games determine the teams who will compete in the “The International”, the biggest esport event in entire esport scene with millions of dollars prize pool.

The prize of TI9 (The International 2019) was $34,330,068 exactly but note that the base prize which Value put in was only $1,600,000. Yes, the other pool has been developed by players by buying battle pass which is full of sound/character/skin/image arts for the game and players by buying them, contribute %25 percent of the prize pool which means in less than few months by just using arts and business ideas, Valve obtained more than $100,000,000 money! Here are the stats of Dota 2 circuit tournaments [6]:

|  |  |  |  |
| --- | --- | --- | --- |
| Tournament | Base | Contributed | Total |
| TI9 | $1,600,000 | $32,730,068 | $34,330,068 |
| TI8 | $1,600,000 | $23,932,177 | $25,532,177 |
| TI7 | $1,600,000 | $23,187,916 | $24,787,916 |
| TI6 | $1,600,000 | $19,170,460 | $20,770,460 |
| TI5 | $1,600,000 | $16,829,613 | $18,429,613 |
| TI4 | $1,600,000 | $9,331,105 | $10,931,105 |

Figure 4 Dota 2 Pro TI Prize Pool

Just note that all other major tournaments have been omitted which contribute at least $5,000,000 each year.

# 3 Dota 2 Mechanics

But let’s talk about science and how the game really works. I will first define the core mechanic of game and its basic definitions then I will go through other components of game such as heroes and their respective variants.

## 3.1 Definitions

Dota 2 is an MMO game with top-down view which takes place against 2 team of 5 players which each one chooses a hero and only can control that hero and its summoned units. The map of the game is always same but opponent can choose Radiant (left corner) or Dire (right corner) which has their own benefits such as better control over particular towers or better vision near Roshan (an AI that holds precious items but really hard to kill)

In each lane, there are towers to give vision and defend you and the goal is to destroy ancient which is the primary building at the hear of the base of opponent.

In each lane, creeps (weak AI creatures) will be respawned which can be killed by heroes to get gold and experience to level up their abilities or buying items. So, more gold/XP leads to stronger hero which MAY lead to winning the game.

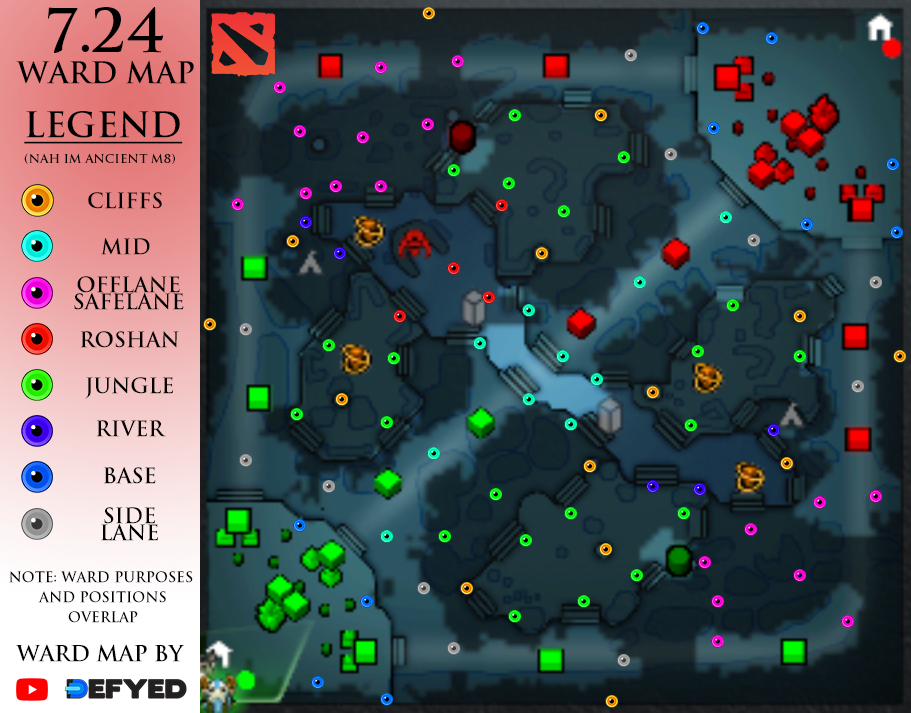


Figure 5 Dota 2 Map [7]

### 3.1.1 Concepts

As this game is super complex, I just mention few of primary stuff that we can propose some solutions of insights within this article.

First thing we need to know is that there are 3 type of damages, physical, magical and pure which can be absorbed by armor, magic resistance and nothing (pure pierces through everything) respectively which can be gained by leveling up abilities or buying items to counter. This is literally the whole idea of game, heroes have different type of powers and different type of defensive skills where you can gain from buying items or leveling up your character which can be combined to dodge enemy attacks and kill them.

### 3.1.2 Heroes

In this game exists about 100+ heroes which can be categorized based on 3 main attributes of each hero, agility (carry), intelligence (carry/supp) and strength (tank/supp) where they do high physical damage, magical damage or utility buffs/debuffs and defensive abilities or utility buffs/debuffs.

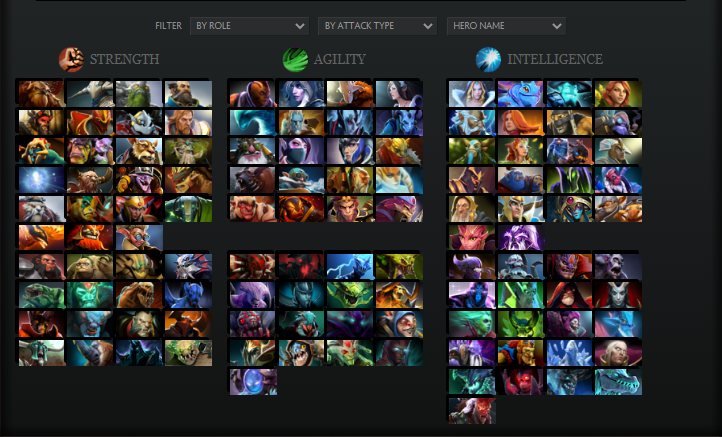


Figure 6 Dota 2 heroes [8]

Each hero has its own unique abilities (at least 4) which one of them is ultimate which has high cooldown/mana cost but high utility or damage output (usually).

### 3.1.3 Abilities

Each hero has its own abilities which correspond to the type of hero, for instance, agility heroes have some kind of abilities to do more physical damage or disable target or intelligence heroes can heal or disable target with low cooldown and high burst of magical damage.

Different abilities can help each other to make the effect stronger, increase the duration or vice versa against opponents. So, here is another challenge that what combination can be constructed without making unstoppable or weak structure.

To demonstrate how critical it is, in the last patch, many of abilities had changes about 0.02. This small amount in the many different stances depicts that drift concept can occur which means an error (we can assume that if we mistakenly update something bigger/smaller than what it should be) can accumulate through all abilities/items and make the game unstable and unbalanced. For instance, let’s say we used below ability (which one the simplest abilities in the game due to very simple effects) and also combined with other abilities to increase buff duration and decrease cooldown in the way minimize the distance between these two timers, then isn’t is nonsense to have cooldown lower than duration for such a key ability? The answer is yes and, in the game, there is at least 10 abilities/items that can affect it. Now assume complex abilities that almost entire hero/item pool can be effective.

Note that effects can be both additive or increase by multiplication.

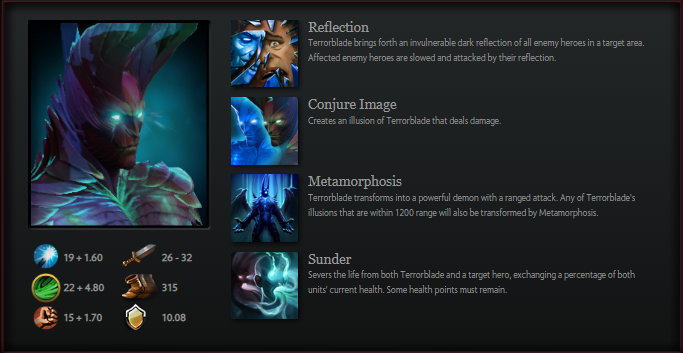
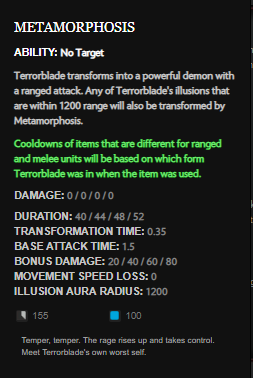


Figure 7 Terrorblade abilities [9]



As we can see, many factors affect the character, their teammates, their damages, attack speed, etc.

Now here comes the challenge. How can we determine these values for this hero so its combination with some particular heroes do not turn into unkillable set or even so weak that no one picks them?

Till today, this issue has been solved yet as we call nerf/buff in every patch because every time a group of heroes become stronger as previously, they were weak or vice versa.

Note that in some patches, the cooldown/animation delay has been changed in term of milliseconds. So, timing matters so much and makes million-dollar differences there!

There is saying: “6 Million dollar Echo slam” which changed the final match in less than 0.5 second.

Figure 8 Terrorblade's Metamorphosis

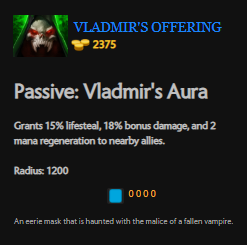
### 3.1.4 Items

The logic is almost same as abilities but main difference is that items are common between all heroes in the game and anyone with enough amount of gold can afford and use it. And again, items can be used to buff other abilities/items or nerf them.



Figure Dota 2 items [10]

I will demonstrate the effect of items again by using two simple items that can be combined to buff the used ability in the previous section.



Our hero does not have lifesteal (a percentage of damage as hit point/health) so to get sustainability, we buy a cheap item which gives %15 lifesteal and %18 damage as additive buff. It also has 1200 radius effect, you will now about its usage in next section.

Ok, now we have something on top of all our abilities.

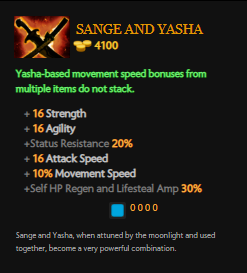
Here is the question: We used an item to buff our hero. Can we use another item on top of this to buff this item too?

YES!



This item gives us %24 lifesteal amplification which means our previous item which gave us %15 lifesteal now has much more effect but this effect is multiplied (not additive). Also, it has %16 status resistance which gives more duration to Metamorphosis.

Now I want to buff this item too! But I do not want to use another slot (I have only 6 free in total) so I upgrade this to get more buffs.



If we get Yasha and combine with Sange, we amplify everything in both items.

And of course, there are many items to counter this situation but due to super complexity, we will not go through this.

### 3.1.5 Teamwork

Here is the idea, the only thing we talked in every section was that we can combine everything. So, do we have to do it all on our own? No, our teammates can come to help and here is the game gets super complex! Timing of items/abilities can disable your own teammate or buff so much that a carry solely handles the entire enemy team and many million dollars just went from one team to another only because of the correct timing and correct choice of items/abilities.

We all know the logic here, so, let’s assume we are Terrorblade and playing carry, then let’s our support/tank by a cheap item such as Vladimir’s offering to give everyone lifesteal and we only pay for Sange and Yasha.

Why would be stop there? Let’s get more buffs from our teammates.



Again this item is cheap to but offers very good stats to our teammates which buying them on our own for every character costs more than 10000 golds.

That is the idea, support items can be shared by all heroes and this way each hero can focus on their roles.

Our Terrorblade has a lot of damage, has lifesteal, has sustain, has status resistance, but wouldn’t it be great to give movement speed to a high damage carry which is also ranged? Yes, sure. So, here comes another combination that can affect entire game.

Now we now enough about challenges, let’s provide some ideas that might improve game experience.

## 3.2 Balancing

The main challenge in this type of games are the balancing between characters, abilities and items and as you have seen, a small percentage of buff/nerf can make a hero everyone’s favorite or a dead hero!

We try to use some of our knowledge in data mining to provide at least another helpful hand in determining these values.

### 3.2.1 Heroes

Here are some ideas we can use:

* Let’s create characters by our sense, then roll out the patch and analysis the post-game stats. We can determine that pro players are spamming particular build of items and combination of heroes and by calculating the corresponding distribution, we can understand that we need buff/nerf or not.  
  To determine item spamming, we can use ***association rules*** to rank most common used patterns as spamming act and provide buff/nerf for highest/lowest ranks. In this task features such as win rate, KDA, and net worth all can be used.  
  Above aforementioned features can be used for a clustering algorithm to put spammed heroes in different clusters and we rank them and same updating idea can be used.

### 3.2.2 Abilities

### 3.2.3 Items

### 3.2.4 Teamwork

# 4 Dota 2 Ranking

## 4.1 Definitions

### 4.1.1 Medals

### 4.1.2 MMR

### 4.1.3 Smurfs

### 4.1.4 Reports

### 4.1.5 Overall Conduct

### 4.1.6 Teamwork

## 4.2 Balancing

### 4.2.1 Medals

### 4.2.2 MMR

### 4.2.3 Smurfs

### 4.1.4 Reports

### 4.1.5 Overall Conduct

### 4.1.6 Teamwork

# 5 Servers

## 5.1 Locations

## 5.2 Distributions

# References

[1] <https://steamcharts.com/app/570>

[2] <https://openai.com/projects/five/>

[3] <https://www.twitch.tv/videos/410533063?t=44m53s>

[4] <https://arena.openai.com/#/results>

[5] <https://openai.com/blog/openai-five/>

[6] <https://dota2.prizetrac.kr/>

[7] <https://www.reddit.com/r/DotA2/comments/f44mx8/724_ward_map_infographic_updated_common_warding>

[8] <http://www.dota2.com/heroes/>

[9] <http://www.dota2.com/hero/terrorblade/>

[10] <http://www.dota2.com/items/>

[11]

[12]

[13]

[14]

[15]