

“

*Data Science Project*

# What is Dota 2?


## And why prediction of hero to choose matters?



# Scraping and getting the right data:

1. got all heroes inside the game and their abilities.
2. use of regular expressions to extract changes made to heroes in different versions.
3. put all changes to one data frame.


## General [\[edit\]](#)

- Neutral Ancients gold bounties reduced by 10%.
-  Arcane Rune mana cost reduction reduced from 40% to 30% (same as the cooldown reduction).

## Heroes [\[edit\]](#)





### Alchemist

-  Unstable Concoction damage increased from 150/220/290/360 to 160/240/320/400.



### Bane

-  Enfeeble duration reduced from 20 to 14/16/18/20.
-  Brain Sap damage rescaled from 90/160/230/300 to 75/150/225/300.


















### Brewmaster

## Version 7.07d

(Full list)

### • Nerfed:

-  Bane
-  Brewmaster
-  Broodmother
-  Earth Spirit
-  Medusa
-  Morphling
-  Night Stalker
-  Omniknight
-  Pudge
-  Razor
-  Sand King
-  Shadow Demon
-  Tinker
-  Tiny
-  Winter Wyvern

### • Buffed:

-  Alchemist
-  Chen
-  Pudge

```
pattern = r'[A a]gility'
in_pattern_ability = r"[ *]"
in_pattern_stats = r"[ *]Strength|Agility|Intelligence|Ba
in_pattern_plus = r"[ *]increased[ *]"
in_pattern_minus = r"[ *]reduced[ *]"
```

```
#universal pattern
```

```
#ability pattern
```

```
for i in right_hero_abilities_lst:
```

```
    in_pattern_ability += (i[:-1] + "|")
```

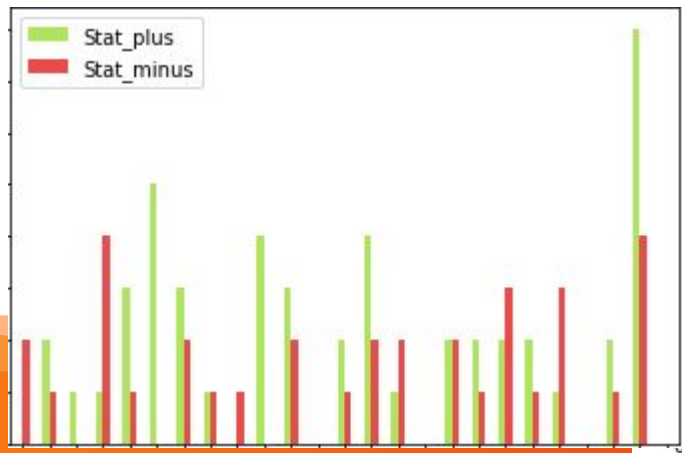
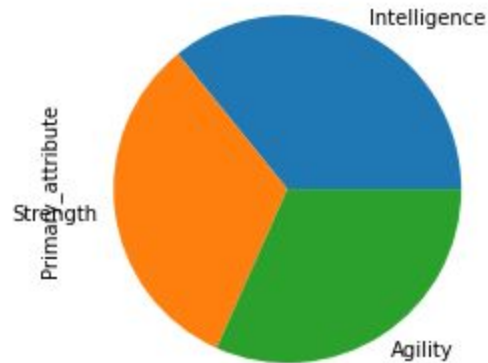
```
in_pattern_ability = (in_pattern_ability[:-1] + "[ *]")
```

```
#
```

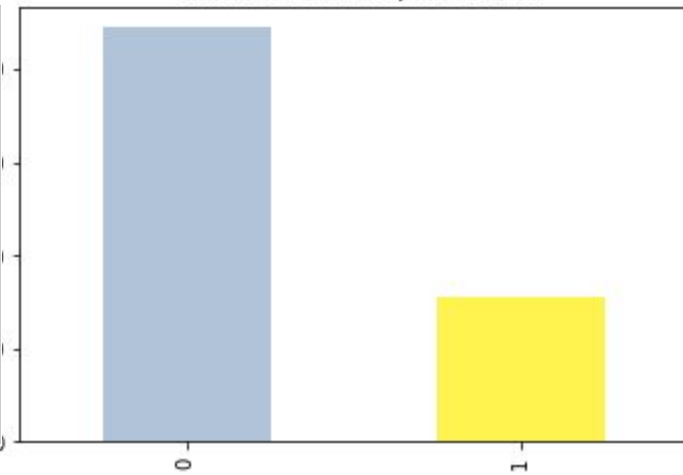
	Heroes	Abilities
0	Abaddon	[Mist Coil , Aphotic Shield , Curse of Aver, B...
1	Alchemist	[Acid Spray , Unstable Concoction , Unstable C...
2	Ancient Apparition	[Cold Feet , Ice Vortex , Chilling Touch , Ice...
3	Anti-Mage	[Mana Break , Blink , Counterspell , Blink Fra...
4	Arc Warden	[Flux , Magnetic Field , Spark Wraith , Tempes...

# EDA and data analysis:

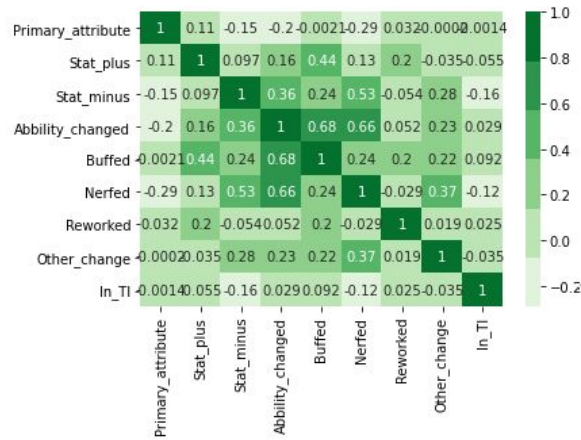
- The heroes spread by attributes is balanced.
- All the heroes got changes to abilities in some sort.
- Some heroes are buffed with stats while other nerfed.
- There are cases where heroes for example, treant protector. That even to it seems that they are buffed and nerfed well, get a lot more reduction in stats.
- Around quarter of the heroes were picked in the last The International match.
- Even heroes that compared to other buffed a lot, they still were not picked.



Number of heroes picked in TI



- Some heroes were deleted from the data set due to the fact that even though they got changes because of the tournament rules could not be picked.
- The correlation of different features was very weak, only features that got closely correlated were buffed/nerfed and ability/stat change. Which is expected as those changes what makes hero stronger or weaker.
- Names were dropped of the data set and the primary attribute of the heroes changed to numerical, but still categorical for. For example Strength changed to 1.



# ML and Conclusion:

- For machine learning algorithm logistic regression were chosen. As fitted for the 1/0 categorial problem, and supposed to return a good prediction.
- Reasons for failure are many, such as scaling down the problem and the structure of the data set. Maybe even the known tools and approaches are not the right ones for this sort of problem.

## First model Evaluation:

accuracy is: 0.6666666666666666  
precision is: 0.3333333333333333  
recall is: 0.14285714285714285  
f1 is: 0.2

## Second model Evaluation:

accuracy is: 0.6666666666666666  
precision is: 0.0  
recall is: 0.0  
f1 is: 0.0