# Proposal: Using Machine learning to predict professional Dota 2 game results

Valeriia Chirkova
Central Supelec
Russia
valeriia.chircova@student-cs.fr

Julien Crabie
Central Supelec
France
julien.crabie@student-cs.fr

Victor Sanchez
Central Supelec
Peru
victor.sanchez@student-cs.fr

Arnaud Cluzel
Central Supelec
France
arnaud.cluzel@student-cs.fr

**ACM Reference Format:**
Valeriia Chirkova, Julien Crabie, Victor Sanchez, and Arnaud Cluzel. 2018. Proposal: Using Machine learning to predict professional Dota 2 game results. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 MOTIVATION

Today, the Defense of the Ancients 2 (Dota 2) game is one of the most played game in history. It had between 10 and 11 million unique players in the last month and considers approximately 800 thousand players, per day, connected at the same time all around the world. It's so famous that they are championships all around the world, these are later considered as qualifiers for the biggest event of the year (during August): The International, with prices around of 10 million dollars for the 1st place. The best teams around the world participate in this world cup of Dota 2.

According to the paper done by Kevin Conley, Daniel Perry. *How Does He Saw Me? A Recommendation Engine for Picking Heroes in Dota 2, 2013*[6], it is possible to predict a match outcome looking solely at the hero composition of a team and whether or not that team was victorious can provide a useful model for match outcome prediction and as a result, hero recommendation. Despite the high accuracy, they didnâĂŹt consider the changes in the game patches (during several months, something that can change the outcome) and the players. Later on, this article was taken in consideration by *Song, Kuangyan, Tianyi Zhang, Chao Ma "Predicting the winning side of DotA2"*[2]. CS229 2015 and they started working with the players information, but there were some big assumptions:

(1) The players in one game have a similar simulation level.
(2) The players could play every hero equally well.

It is clear that these assumptions could affect the outcome, but they were necessary to just work with the problem of Hero Selection.


*Conference'17, July 2017, Washington, DC, USA*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Finally, articles were taken in consideration *by Petra Grutzik, Joe Higgins and Long Tran: "Predicting outcomes of professional DotA 2 matches"*[3] and include the work from the perspective of sports analytics seeing it in parallel: Sabermetrics and rigorous statistical analysis ïňĄrst transformed baseball management during the 1930s-1940s, which then led to a revolution in statistical applications in other sports and opened new ïňĄelds of research in statistics itself. Progress has been made on predicting the outcomes of two player games. Nevertheless, games with multiple teammates on each team prove more difiňĄcult to predict. There were good outcomes from the research but there is still missing the good relation between the information of heroes and the information of players. What we will focus on is to combine the factors of:

- The information of the heroes
- The information of the players
- The information regarding the team of the players.
- The information of the role of the player in the team.

This should increase our accuracy and will give us a good outcome for the prediction.

## 2 METHODOLOGY

The problem is stated as follows: we want to predict the outcome of a match with a winner team and a loser team, which is defined as a classfication problem. The project includes the following steps:

- mining and preparing the dataset
- applying machine learning algorythms
- evaluating the accuracy of the model

Our data set will be extracted from Valve and Gosu.net API. We have not yet extracted our dataset, as we need to be more familiar with API calls, but a lot of help is provided online on how to extract Dota 2 game data.

There has been a lot of research done on this topic. In this research, different models have been applied to find the win probability prediction of a Dota 2 profressional match based on team drafts. Most of them applied logistic regression, either alone or in a combination k-nearest-neighours (K-NN), principal component analysis (PCA) or random forest. More sophisticated models are based on support vector machine (SVM) and neural networks.

Based on these researches, we suggest first to apply a logistic regression to our dataset. With this model we will be able to understand the impact of the different team features on the winning chances of the two squads.

After analysing the results and depending on the performance of this model, we will look for possible improvements:

- test other algorityhms (like the ones listed above)
- reduce dimensionality of team features
- try new combination of variables

## 3 EVALUATION

To evaluate our Data we plan on following the methodology used by *by Petra Grutzik, Joe Higgins and Long Tran: âĂIJPredicting outcomes of professional DotA 2 matchesâĂİ*[3]. We will separate our data into a training set, testing set and validation set. As we have not yet extracted the data set from Valve and Gosu.net API, we don't have specific numbers, but given the large literature on this data set and it's expected size (16,800+ match)[3], we believe that a 90%/10% division of our data set between the training set and test set is appropriate. We will then evaluate success depending on the success rate of our model on our testing sample.

If our model is better than luck, then we will consider our model to be successful. Given that there can be only two outcomes (team 1 wins or team 2 wins), the expected value of prediction through luck is 50%, thus a prediction success above 50% for our model would indicate success. However given the previous work on the subject, we will consider our work truly successful if we obtain a success rate higher than Grutzik, Higgins, and Tran[3].

## REFERENCES

[1] K. Conley and D. Perry, How Does He Saw Me? A Recommendation Engine for Picking Heroes in Dota 2, tech. rep., 2013.
[2] Song, Kuangyan, Tianyi Zhang, Chao Ma Predicting the winning side of DotA2. CS229 2015.
[3] Petra Grutzik, Joe Higgins, Long Tran, Predicting outcomes of professional DotA 2 matches. CSS229 2017
[4] Zhengyao Li, Dingyue Cui, and Chen Li, Dota2 Outcome Prediction.
[5] Aleksandr Semenov, Peter Romov, Kirill Neklyudov, Daniil Yashkov,and Daniil Kireev, Applications of Machine Learning in Dota 2: Literature Review and Practical Knowledge Sharing
[6] Kevin Conley, Daniel Perry. How Does He Saw Me? A Recommendation Engine for Picking Heroes in Dota 2, 2013.

Received October 2018