Introduction

In the world of Pokémon there is always only one who can win. Pokémon are creatures of all shapes and sizes and are trained to battle against each other. A Pokémon has a type and different stats, such as speed and attack. Every Pokémon type comes with both advantages as well as disadvantages when battling different Pokémon types.

The aim of this project is to predict the outcome of Pokémon battle using various algorithms of Machine Learning. During the project we want to compare different algorithms to investigate which algorithm can best predict the Pokémon that win. We applied the models Naïve Bayes, Logistic Regression, Random Forest, Decision Tree and Support Vector Machine. Therefore the main research question discussed in this paper will be:

*“Which of the five proposed models predicts the outcome of a Pokémon battle best?”*

The data we use is from kaggle.com and contains results of 50.000 battles, which contains the id of the two Pokémon’s and which one won. Another dataset has the stats of all Pokémon types. A final dataset has information about the power of the move of a Pokémon type against any other possible type.

There have been several approaches to predicting outcomes of a game. We looked at different research of predicting the outcomes of sports matches, such as cricket. When looking how previous researches tried to predict the outcomes of game can see that there is a wide variety in using techniques of Machine Learning.

*S. Kampakis and W. Thomas (2015)* investigate to what degree it is possible to predict the outcome of cricket matches. They used models of different combined team and player features. On these models various algorithms are done to predict the outcomes of the cricket game.

*Raza Ul Mustafa (2017)* used social media to predict the outcome of a cricket match. For prediction they used three method*s* that depend on the total number of tweets before the game for each team, fans sentiments toward each team and fans score predictions on Twitter. In this research the Support Vector Machine has shown advantage over other classifiers.

*A. Bandulasiri (2008)* looked at the different factors in a cricket match. It was investigated whether and how much influence certain factors have on winning a cricket match, like winning the toss, game plan (batting first or fielding first) and match type (day or day & night).

The study of *Haghighat, Rastegari & Nourafza (2013)* reviews previous research on data mining systems to predict sports results and evaluates the advantages and disadvantages of each system. A conclusion of the research is that including different features such as player performance will contribute to more accurate predictions.

The results of games are also predicted in other sports. In football are 3 possible outcomes variables win, lose or draw. *Joseph, Fenton, & Neil, (2006)* using Bayesian nets and other machine learning techniques to predict the football results and *Pettersson & Nyquist (2017)* using deep learning.

The content of this paper is structured as follows. The first part will exist of data-analysis. Hereby different variables are investigated. Also a lot of the data has to be preprocessed. Many variables must be made numeric before we can use them in the models. Hereafter, various algorithms are applied. Finally, the outcomes are evaluated.

References

[1]Lee, S., & Togelius, J. (2017). *Showdown AI Competition.* http://game.engineering.nyu.edu/wp-content/uploads/2017/02/CIG\_2017\_paper\_87-1.pdf

[2] Raza Ul Mustafa, M. S. (2017). *PREDICTING THE CRICKET MATCH OUTCOME USING CROWD OPINIONS ON SOCIAL.*

[3] Bandulasiri, A. (2008). *Predicting the Winner in One Day International.*

[4] Kampakis, S., & Thomas, W. (2015). *Using Machine Learning to Predict the Outcome of English County twenty over Cricket Matches.* https://arxiv.org/ftp/arxiv/papers/1511/1511.05837.pdf

[5] Joseph, A., Fenton, N., & Neil, M. (2006). *Predicting football results using Bayesian nets and other machine learning techniques.* https://pdfs.semanticscholar.org/7272/1c599a2213753f7396a744800891fac90622.pdf

[6] Pettersson, D., & Nyquist, R. (2017). *Football Match Prediction using Deep Learning.* http://publications.lib.chalmers.se/records/fulltext/250411/250411.pdf

[7] Haghighat, M., Rastegari, H., & Nourafza, N. (2013). *A Review of Data Mining Techniques for Result Prediction in sports.*

[8] Hsu, C., Chang, C., & Lin, C. (2010). *A Practical Guide to Support Vector Classification*

[9] Bradley, A. (1997). *The use of the area under the ROC curve in the evaluation of machine learning algorithms.*

30%

70%

cross-validation

Training algorithms

Testing algorithms

Test set

Training set

Running algorithms

ROC Curve

Data-analysis

Multiplier data

Combats data

Pokémon data

Combined dataset

Normalized dataset