Football Result Prediction by Deep Learning Algorithms

Thesis · May 2019

DOI: 10.13140/RG.2.2.25014.45122

1 author:

Stefan Samba
Tilburg University

Football Result Prediction by Deep Learning Algorithms

Stefan Samba
STUDENT NUMBER: u980019

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN
COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee:

Supervisor: prof. dr. E.O. Postma
Second Reader: dr. G. Chrupala

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
May 2019

Preface

Before you lies my master thesis “Football Result Prediction by Deep Learning Algorithms”, It has been written to fulfill the graduation requirements of the Cognitive Science & Artificial Intelligence track at the Tilburg University in May 2019.

It is my background in sports management and my own experience as a former professional athlete that has driven this research. In the ideal case a model is able to predict the performance based on a set of features. If the model is accurate, changing parameters allows you to predict the future and design your own and best performance.

I would like to thank my supervisor prof. dr. E.O. Postma for his guidance and support during the process. Also, I wish to thank myself for my structured approach and for always believing in myself.

Enjoy your reading.

Stefan Samba

Tilburg, May, 2019

Abstract

In this work the performance of deep learning algorithms for predicting football results is explored. To date, there are only few studies that have investigated to what extent a neural network is able to predict the outcome of football matches. To explore this matter, a dataset has been created that contains matches from the Premier League, Championship, League 1 and League 2, resulting in 20.000 matches. The ncorporated feature vector contains team dependent features for the home team, team dependent features for the away team and team independent features. The team dependent features contain team strength, cumulative sum for different variables during the season, form and distance between matches. Team independent features comprehends the referee, bookmakers odds, season and division. Corresponding labels are a home win, away win or draw. To test to what extent deep learning algorithms are able to predict the outcome of football matches, different multilayer perceptrons have been created that differ in deepness, number of neurons per layer and output. The main finding of this study is that a neural network with 3 output neurons is better in predicting the outcome of football matches than a neural network that consist of 1 output neuron with respectively 48% and 43% accuracy. As prior studies regarding football predicting created networks that consist of 1 output neuron, this leaves abundant room for future research.

[Keywords] Football Result Prediction • Sport Prediction Model • Deep Learning • Multilayer Perceptron • Artificial Neural Network • Artificial Intelligence • Premier League • Championship • League 1 • League 2

Contents

1. Introduction

Deep learning has shown promising results in classification and prediction of complex cases. During this study, football or soccer, hereafter named football, will be the sport representing this complex area. Football is a decent representation of a complex area because of its dependency on many features, such as team strategy, player morale, team morale, quality of offense, quality of defence, possession and fatigue and luck. For laws of the game, Fifa (2018) can be consulted. Large and detailed collections of historical data, per season, per league and per match offers great possibilities to investigate the modelling of football results. The remainder of this section contains an introduction to football result prediction, the research question and the outline of this paper.

Football result prediction shows it's relevance in a practical and scientific way. From a practical point of view it is interesting for clubs and nations to improve performance, for bookmakers to increase finances and in other ways for different football related parties. From a scientific point of view, it is interesting because the outcome of matches is difficult due to its dependency of many features (Aslan & Inceoglu, 2007; Bunker & Thabtah, 2019; Langaroudi & Yamaghani, 2019) and luck (Aoki et al., 2017).

The freely available football data per match contains different features and offers possibilities to use deep learning for football result predictions. If something as "unpredictable" as football can be predicted to a certain extend this might open open up possibilities to predict other complex scenarios or could indicate that data should be gathered for other domains to make more accurate predictions.

Deep learning for football result prediction has not yet been thoroughly explored. Aslan and Inceoglu (2007) supports the idea that neural network approaches can be used for predicting the outcome of football matches by using a set of 4 features. Arabzad et al. (2014) also used a neural network to predict matches and indicated at least two limitations of their study. One indicated limitation is that Arabzad et al. (2014) did not incorporate the distance between matches. A second indicated limitation is that the referee during the match was not incorporated in their model. Besides these possible features, Giuliodori (2017) and Tax and Joustra (2015) both indicate that bookmakers' odds can be an indicator for result prediction.

1.1 Research Question

The lack of scientific papers towards football prediction models by using deep learning algorithms in combination with limitations of existing studies results in the following research question:

To what extent can a deep learning algorithm predict the outcome of football matches?

Obtained data from Football-Data (2019) includes match details per season for the four leagues in England, the Premier League, Championship, League 1 and League 2 and offer possibilities to train a deep learning algorithm and predict football results.

1.2 Outline

The following section, 2. Related Work begins by laying out the theoretical dimensions of the research. The third section, 3. Method, is concerned with the methodology to answer the research question above. Chapter 4. Results will describe the main findings and post-hoc analysis and in the remainder, the discussion will be described and the paper is finalized by a conclusion and future work section.

2. Related Work

This chapter will cover 4 sections. The first section will briefly introduce the multilayer perceptron and its architecture. Second, scientific papers related to football prediction by deep learning algorithms will be discussed. The third and fourth section will give insights in the deepness and the number of neurons per layer for football result prediction. Other deep learning techniques such as convolutional neural networks or recurrent neural networks will not be discussed as they are not relevant for this paper.

2.1 Multilayer Perceptron & Architectures

The multilayer perceptron (MLP) is a deep learning model that is able to separate non-linear data. As indicated by its name, the MLP consists of perceptrons within multiple layers. More information about the perceptron and its components can be found in Minsky and Papert (1988).

The concept of a MLP finds its roots in the biological brain where each neuron can receive, process and transmit incoming signals. Combining the strength of the different neurons together in an artificial settings leads to perceptron with multiple layers. These layers and how the neurons within the layers connect to each other is considered the architecture of the network (Goodfellow et al., 2016).

An artificial neural network architecture consists of at least 3 layers as visualized in Figure 1, which can be found on the next page. The first layer is the input layer and contains the values of the input features. The second layer is considered the hidden layer. More hidden layers can be added to extent the deepness of the model. E.g. Figure 1 shows a MLP with 1 hidden layer of 3 neurons and Figure 2. pictures a MLP consisting of 2 hidden layers with respectively 4 and 3 neurons per layer. Last named MLP architecture consists of more layers, this model can therefore be seen as deeper than the proposed MLP architecture of Figure 3. The third and final layer is the output layer which will translate the signal into a predicted label or class. Figure 3 visualizes a MLP architecture with 1 class as output while Figure 3 pictures a MLP architecture with 3 classes within the output layer. Goodfellow et al (2016) and/or Reed and Marks (1999) can be consulted for a more extensive explanation of the MLP and its components.

Figure 1
A neural network architecture with 1 hidden layer of 3 neurons and 1 output layer of 1 neuron.

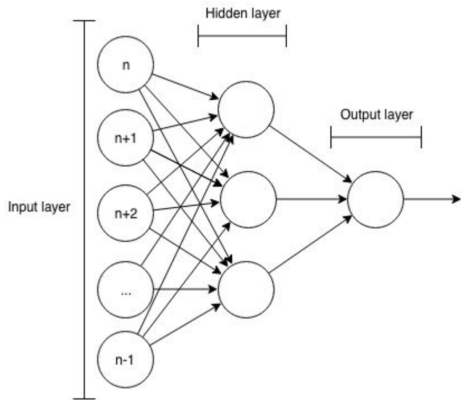
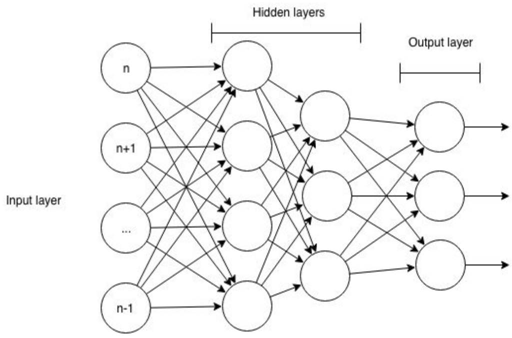


Figure 2
A neural network architecture with 2 hidden layers of respectively 4 and 3 neurons per layer and 1 output layer of 3 neurons.



While there is a growing body of literature that recognises the effectiveness of neural networks, to date there has been little agreement about the number of layers and number of neurons per layer for an optimal performance of a neural network and specifically for football result prediction.

2.2 MLP for Football Result Prediction

To date, there are few studies that have investigated the ability of deep learning algorithms to predict football results. The existing body of literature towards football result prediction by deep learning algorithms, is diverse in various ways such as the incorporated league(s), input features, deep learning architectures and accuracy. Therefore, in this study, 5 prior studies have been examined that all use deep learning techniques for football result prediction. These studies are difficult to compare as they are so diverse but they do contribute to the body of knowledge towards football result prediction.

The first study by Arabzad et al. (2014) incorporates the team, form of the team in the entire league till match date, form of the team in 4 most recent matches, quality of the last opponents and the league and the week. This study is different from other studies as the predicted output is the number of goals for the home team and the number of goals for the away team. Nevertheless the study by Arabzad et al. (2014) is still relevant because of their MLP architecture which will be discussed further in section 2.3 Layers for football result prediction and section 2.4 Neurons for football result prediction, and because of at least two limitations that might attribute to a more accurate model such as the club investment and weather.

The second study, by McCabe and Trevathan (2008), tries to predict the outcome of football matches as a home win, away win or draw. The label is predicted based on a richer set of features than the study by Arabzad et al. (2014) and incorporates the goals for, goals against, current win-loss record, the general home and away performance, the performance of the team in previous 4 matches, the current position on the ranking, location and player availability. With this set of features they found an accuracy of 54% and they indicate that a richer feature vector might enhance the accuracy of the model.

A third incorporated study, is the study by Huang and Chang (2010). They created a MLP to predict the outcome of football matches of stages after the group stage of the World Cup 2006, with the matches of the group stage as input. After the group stages there is no possibility for a draw as one of the two teams has to proceed to the next round. Huang and Chang (2010) incorporated a set of feature that includes match details such as the scored goals, shots, shots on goal, corner kicks, free kicks, ball possession and fouls. Even though this model was trained on only matches of the group stages of the World Cup 2006, they found an accuracy of 77% for the matches after the group stages.

Tax and Joustra (2015), is the fourth investigated study and is the only one that incorporated bookmakers' odds within their model. Besides these odds they also incorporated the goals for, goals against, results in previous matches, top scorers, days since previous match and the performance (win, draw, lose) of the team in percentages. With this feature vector they were able to predict football matches of the Dutch Eredivisie with 55%. To improve their model their feature set can be enriched and tested for other leagues.

The fifth and final incorporated study is the study by Aslan and Inceoglu (2007). While focussing on the Italian Serie A they conducted two experiments. The feature vector of the first experiment consists of four features. A rating for home matches and a rating for away matches for the home team and for the away team. This model is described as LVQ-A and results in an accuracy of 51%. Their LVQ-B model, that only incorporate 2 features, the home rating for the home team and the away rating for the away team, results in an accuracy of 53%. This study indicates that relative simple model can still perform well.

An overview of the introduced studies corresponding league, features, future work and/or limitation(s) and the accuracy can be found in Table 1. Besides the division and features that are described in Table 1, the network architecture might also affect the accuracy of a model. Therefore the different MLP architectures will be discussed in the following section.

Table 1
An overview of the reviewed studies.

Study	Division	Features	Future Work / Limitation	Accuracy
Arabzad et al., 2014	Iran Pro League	Teams, form of teams in last matches & league, quality of last opponents	Distance between matches, Club investment, Weather	N.A.
McCabe and Trevathan, 2008	Premier League	Points for and against, win-loss record, home and away Performance, performance in previous 4 games, ranking, location, player availability	Richer feature sets	54%
Huang and Chang, 2010	World Cup 2006	Goals for, Shots, Shots on Goal, Corner Kicks, Free kicks, Ball possession & Fouls	Limited training data	77%
Tax and Joustra, 2015	Dutch Eredivisie	Goals for, goals against, result previous matched, top scorers, days since previous match, win/draw/lose percentage, Odds	Expand to other Leagues	55%

Aslan and Inceoglu, 2007	Italian Serie A	Home rating & Away rating for home team and away team	Different leagues, structures and input features	51%
Aslan and Inceoglu, 2007	Italian Serie A	Home Rating home team, Away Rating away team	Different leagues, structures and input features	53%

2.3 Layers for Football Result Prediction

Hidden layers are the layers between the input layer and the output layer as visualized in Figure 1 and Figure 2. These hidden layers allow the model to separate non-linear data, e.g. the XOR problem (Goodfellow et al. 2016). Related to the hidden layers, Heaton (2009) describes that there is no reason to create more than two hidden layers while Goodfellow et al. (2016), on the other hand, indicates that deeper models seem to outperform shallower models and that the ideal network architecture can only be found by experimenting and monitoring the validation error.

From the 5 reviewed studies, Arabzad et al. (2014) is the only one using 2 hidden layers while Aslan and Inceoglu (2007), Huang and Chang (2010), McCabe and Trevathan (2008), were all using 1 hidden layer. The number of layers that is used by Tax and Joustra (2013) remains unknown. The findings from these studies are summarized in Table 2 and suggest that shallower models are more popular for Football Result Prediction and are in line with Heaton (2005). Besides the differences in layers between the reviewed studies, the studies also differs in leagues, features and neurons per layer. Therefore the effect of the deepness of the model on the accuracy remains unclear.

Table 2
An overview of network architectures of the reviewed studies.

Study	Total Layers	Hidden Layers	Neurons
Arabzad et al., 2014	4	2	10-20-20-2
McCabe and Trevathan, 2008	3	1	20-10-1
Huang and Chang, 2010	3	1	8-11-1
Tax and Joustra, 2015	N.A.	N.A.	N.A.
Aslan and Inceoglu, 2007	3	1	4-125-1
Aslan and Inceoglu, 2007	3	1	2-25-1

2.4 Neurons for Football Result Prediction

The neurons that receive, process and transmit incoming signals can be found within the hidden layers as visualized by Figure 1 and Figure 2. When looking at the optimal number of neurons per layer, this is considered to be between the size of the input layer and the size of the output layer(s) (Heaton, 2009).

In imitation of earlier findings, the reviewed studies also differ in the number of neurons per layer within the hidden layer(s). While the number of neurons used by Tax and Joustra (2013) is unknown, the study of McCabe and Trevathan (2008) is the only one that created a neural network where the number of neurons per layer is between the size of the input and size of the output, which is in line with Heaton (2009). Arabzad et al. (2014), Aslan and Inceoglu (2007) and Huang and Chang (2010), all created a network where the number of neurons for each of the incorporated hidden layers is larger than the size of the input and output layer. The number of neurons per layer are indicated in the column Neurons in Table 2. The number of neurons within the output layer is similar for all the reviewed studies and is shaped by 1 output neuron. Equivalent to the optimal number of layers, the optimal number of neurons per layer for football result prediction remains unclear.

3. Method

This section describes the dataset, the software that is used, the features and corresponding labels, the deep learning architectures or experimental set up and the evaluation of the model.

3.1 Football dataset

For this paper, a total of 48 files from 12 seasons (2006-2007 until 2017-2018) have been downloaded from Football-Data (2019). Each downloaded .csv file, represents one season of one of the four football leagues in England. Incorporated leagues are the Premier League, Championship, League 1 and League 2 with respectively 4.175, 6.617, 6.618 and 6.619 matches. Each .csv file contains the matches of that specific season and league as rows and different features as columns. This results in a total of 24.029 matches.

3.2 Software

The software that is used to conduct this experiment is Python within Jupyter Notebooks in Anaconda Navigator. Data manipulation and modelling is done through the packages Numpy, Pandas, Plotly, Sklearn and Keras with Tensorflow as backend. Code can be accessed via Appendix C.

3.3 Features & Labels

The Football-Data (2019) dataset is downloaded and a distinction is made between team dependent features for both the home team and the away team that are all engineered, and team independent features that are all extracted from the available data. The following sections will describe the team dependent and team independent features in more detail. The combination of team dependent and team independent features results in a total of 41 features. Corresponding labels are a home win (H), away win (A) or draw (D). Both categories of features are explained in more detail in the following sections.

From the complete dataset, features that contain 5.000 or more NaN's are removed from the dataframe. Non-numerical values within the complete dataset are converted into numerical values before proceeding with the model and data is normalized to values between 0 and 1. The dataset is shuffled and samples have been randomly selected. The data is then converted into an Numpy array and split into a train, val, and test set with respectively 60%, 20% and 20% of the dataset.

3.4 Feature Extraction - Team Independent

The Team Independent features can all be extracted from the data. This category consists of the referee, odds for a home win, draw or away win by 6 bookmakers, season and division and results in a vector of 21 features.

3.4.1 Referee

During a football match, the referee is there to make sure everyone complies with the rules. Even though the rules are equal it is indicated that referee bias contributes to home advantage in the English Premier League (Boyko et al.. 2007). One referee might be more susceptible to this phenomenon than the other. Therefore the referee incorporated in the feature vector and can be extracted from the dataset by using Pandas. Table 3 indicates the referee that was involved during the first 5 home matches of the Premier League club Man United in the season 2006-2007.

Table 3
Referee during the first 5 matches of Man United during the Premier League 2006-2007.

Home Match	Man United	Man United	Man United	Man United	Man United
Referee	A Marriner	M Riley	G Poll	M Dean	G Poll

3.4.2 Bookmakers

One way to bet on football matches is through 1x2 type betting. Bookmakers calculate the odds for a home win (1), odds for a draw (x) and the odds for an away win (2). As each bookmaker has its own methodology to calculate these odds, the odds may differ per bookmaker. E.g. Bookmaker A indicates the following odds for Manchester United vs. West Ham United 1.40, 4.65, 6.65. With these odds a €100,- bet on Manchester United will result in a pay out of €140,- when Manchester United wins the match.

Tax and Joustra (2015) did include bookmakers' odds and found that these odds enhance the performance of the model. Within the dataset the 1X2 odds are incorporate for 6 bookmakers, Bet365, Bet&Win, Interwetten, Ladbrokes, William Hill and VCBet. The by the bookmakers assigned odds are extracted by using Pandas. Table 4 shows the odds by Bet365 for a home win (B365H), draw (B365D) and an away win (B365A) for the first five matches of Man United in the the Premier League 2006-2007.

Table 4
The 1X2 odds by Bet365 for the first 5 home matches of Man United during the Premier League 2006-2007.

Home Team	Man United	Man United	Man United	Man United	Man United
Away Team	Fulham	Tottenham	Arsenal	Newcastle	Liverpool
B365H	1.22	1.50	1.90	1.30	1.90
B365D	5.50	3.75	3.25	4.33	3.25
B365A	13.00	7.00	4.00	13.00	4.00

3.4.3 Season

After a complete football season follows a period of rest, transfers, cards are dismissed and the ranking starts over. It seems logical that this might affect the outcome of football matches. As the MLP able to find optimal weights for features and therefore the importance of the season, this features is incorporated. The feature season will be constant for 1 entire season as indicated in Table 5.

Table 5
Referee during the first 5 matches of Man United during the Premier League 2006-2007.

Home Match	Man United	Man United	Man United	Man United	Man United
Season	Season 2006 2007 Premier League	Season 2006 2007 Premier League	Season 2006 2007 Premier League	Season 2006 2007 Premier League	Season 2006 2007 Premier League

3.4.4 Division

The English football scene is shaped by 4 leagues. The highest attainable competition is the Premier League, followed by the Championship, League 1 and League 2. As the MLP able to find optimal weights for features and therefore the importance of the division, this features is incorporated. The feature division will remain similar for 1 entire season as indicated in Table 6 and might change in the following season, depending on the performance of the team.

Table 6
Referee during the first 5 matches of Man United during the Premier League 2006-2007.

Home Match	Man United	Man United	Man United	Man United	Man United
Division	Premier League	Premier League	Premier League	Premier League	Premier League

3.5 Feature Engineering - Team dependent

Team dependent features are all engineered from the data and are engineered for the home team and for the away team. Engineered features can be distinguished in four categories: team strength, cumulative sum, form and days between matches. The last 3 mentioned categories are calculated separately for the same team for home and away matches. In total, the team dependent feature vector is shaped by 20 features.

3.5.1 Team Strength

To feed the model with information about the strength of the home team and the away team the team strength feature is engineered. There are a number of instruments available for measuring the team strength. In this study the Soccer Power Index (SPI) has been adopted because it covers the teams in the English Leagues, market value of teams and it is within available resources.

The Soccer Power Index is originally created by FiveThirtyEight to indicate the team strength of international football teams and has later been expanded to clubs. The

Preseason SPI is created by combining the SPI rating at the end of previous season with the Market Value of the team (Boice, 2018)

The SPI rating for football clubs is downloaded from Fivethirtyeight (2018) and the obtained .csv file is processed in Pandas to extract teams that are active in the English leagues. The remaining teams are compared to the dataframe from Football-Data. Team names that did not match, have been adjusted to the correct names so that they match with the existing dataset and the spi can be added as a feature to the dataset as shown in Table 7.

Table 7
The spi score of Man United.

Home Team	Man United	Man United	Man United	Man United	Man United
Away Team	Fulham	Tottenham	Arsenal	Newcastle	Liverpool
spi	78.8	78.8	78.8	78.8	78.8

3.5.2 Cumulative Sum

The cumulative sum for different variables in the current season might also be a predictor for the outcome of football matches as indicated by Hirotsu and Wright (2003), Huang and Chang (2010), McCabe and Trevathan (2008), Tax and Joustra (2015) and Vecer et al. (2009). The variables that were calculated by using the cumulative sum per season per club are illustrated in Table 8. The first column indicates the variable, the second column shows the variable name for the home team and the third column pictures the variable name for the away team. Fifa (2018) can be consulted for an explanation of the incorporated variables.

Table 8
An overview of the incorporated features that are engineered by a cumulative sum function.

Variable Name	Home Team	Away Team
Corners	C_HC	C_AC
Fouls	C_HF	C_AF
Red Cards	C_HR	C_AR
Yellow Cards	C_HY	C_AY
Shots	C_HS	C_AS
Shots on Target	C_HST	C_AST
Goals	C_FTHG	C_FTAG
Points	C_FTHP	C_FTAP

As the goal of this study is to predict the outcome of matches before the start of the match it is important to exclude data from the current match in the feature vector for both teams. To illustrate, Table 9 provides the values of the cumulative sum of shots, shots on target and yellow cards for first 5 home matches of Man United in the Premier League 2006-2007.

Table 9
The features cumulative values for shots, shots on target and yellow for the first 5 home matches of Man United during the Premier League 2006-2007.

Home Team	Man United	Man United	Man United	Man United	Man United
Away Team	Fullham	Tottenham	Arsenal	Newcastle	Liverpool
C_HS	0	15	25	42	66
C_HST	0	7	12	21	32
C_HY	0	1	3	5	6

3.5.3 Form

The debate about the role of the form of a team as a predictor for football result is inconcise. Within these discussions the effect of a losing streak and the effect of a winning streak are discussed. Goddard (2006) indicates that a losing streak increases the probability of winning and that a winning streak decreases the probability of winning while Heuer and Rubner (2009) found the opposite. Even though the role of the form towards football result prediction is still under discussion, the form can be calculated automatically and the MLP should be able to find out if it is relevant for football result prediction.

To add the feature form to the feature vector, the form is calculated separately for home and away matches for the home team and for the away team. Ulmer and Fernandez (2013) found that the optimal number of matches to include in the form is between 4-7 for different machine learning techniques. As a neural network was not part of the Ulmer and Fernandez (2013) study, in this study, 4 matches are included. To engineer the form a rolling sum function is used in Pandas that calculates the sum of the obtained points during the last 5 matches minus the points obtained in current match. When the season starts, the first 4 matches will be scored with 0 as there are no 4 previous matches in that season. This feature is visualized for the first 5 matches of Man United in the Premier League 2006-2007 in Table 10.

Table 10
The form of Man United during the first 5 home matches of the Premier League 2006-2007.

Home Team	Man United	Man United	Man United	Man United	Man United
Away Team	Fullham	Tottenham	Arsenal	Newcastle	Liverpool
F_HT	0	0	0	0	9

3.5.4 Days between matches

Fatigue can also affect the team performance as indicated by Constantinou et al. (2012). Constantinou et al. (2012) describe that fatigue is shaped by different variables such as toughness of the previous match, days since last match and more. Due to available resources in this study only the days between matches is incorporated.

The days between matches are engineered for home matches and away matches separately by calculating the difference in days between the match day and the previous match day. Table 11 shows the distance since last home match where Man United plays as the home team.

Table 11
The distance between the first 5 home matches of Man United during the Premier League 2006-2007.

Home Team	Man United	Man United	Man United	Man United	Man United
Away Team	Fulham	Tottenham	Arsenal	Newcastle	Liverpool
D_HM	0	20	8	14	21

3.6 Deep Learning Architectures

In this experiment there are 12 multilayer perceptron (MLP) created that differ in architecture. The architectures differ in three aspects. The first aspect is the number of hidden layers and varies between 1 and 3. The second aspect is number of neurons per layer which differs so, that it is larger or smaller than the number of input neurons. The final aspects that is experimented with, is the output. The output neurons consist of either one output neuron with a relu activation, or three output neurons with a sigmoid activation function. Table 10 gives an overview of the incorporated MLP architectures.

Table 12
The different network architectures for football result prediction.

Total Layers	Hidden Layers	Neurons
3	1	41-75-1
3	1	41-10-1
4	2	41-75-75-1
4	2	41-10-10-1
5	3	41-75-75-75-1
5	3	41-10-10-10-1
3	1	41-75-3
3	1	41-10-3
4	2	41-75-75-3
4	2	41-10-10-3
5	3	41-75-75-75-3
5	3	41-10-10-10-3

Within the neural networks that contain 1 output node, Stochastic Gradient Descent (SGD) with a 0.01 learning rate is used as optimizer with a Relu activation function, mean squared error as loss and accuracy as metric. For the neural networks that contain 3 output nodes, Stochastic Gradient Descent (SGD) with a 0.1 learning rate is used as optimizer with a Sigmoid activation function, categorical cross entropy as loss and accuracy as metric.

3.7 Evaluation

The model is trained on a 60% train set, validated on a 20% validation set and tested on a 20% test set. Accuracy is the evaluation metric and is calculated for the different architectures. To prevent overfitting the loss is monitored and the number of epochs is empirically determined.

4. Results

In this section, the performance of different neural networks as described in section 3 will be presented in two parts. The first parts contains the main findings and the second part is shaped by post hoc analysis.

4.1 Main findings

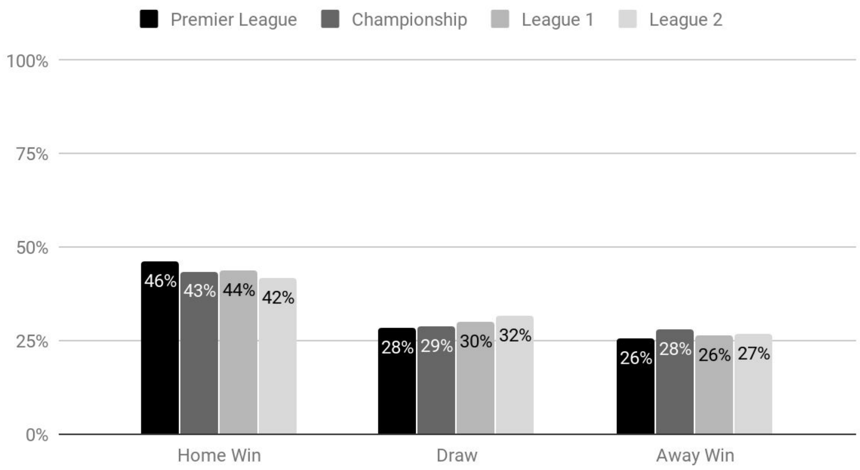
To test till what extent a neural network is able to predict the outcome of football matches, different neural networks have been created. The accuracy and loss of the train and validation set are visualized in the figures of Appendix A and Appendix B. The performance of the different models are noted in Table 13 and show that the different neural networks with 1 output neuron are able to reach a 43% accuracy and that the incorporated neural networks with 3 output nodes are able to reach 48% accuracy.

Table 13
The different network architectures and corresponding accuracy.

Total Layers	Hidden Layers	Neurons	Epochs	Accuracy
3	1	41-75-1	50	43%
3	1	41-10-1	50	42%
4	2	41-75-75-1	50	43%
4	2	41-10-10-1	50	43%
5	3	41-75-75-75-1	50	43%
5	3	41-10-10-10-1	50	43%
3	1	41-75-3	200	48%
3	1	41-10-3	200	**48%**
4	2	41-75-75-3	200	48%
4	2	41-10-10-3	200	48%
5	3	41-75-75-75-3	200	48%
5	3	41-10-10-10-3	200	43%

Further exploration of the data revealed that the data, which is labeled as a home win, draw or away win, was not normally distributed. A home win occurs most often, followed by a draw and an away win as indicated by Figure 3.

Figure 3
Data distribution for the Premier League, Championship, League 1 and League 2.

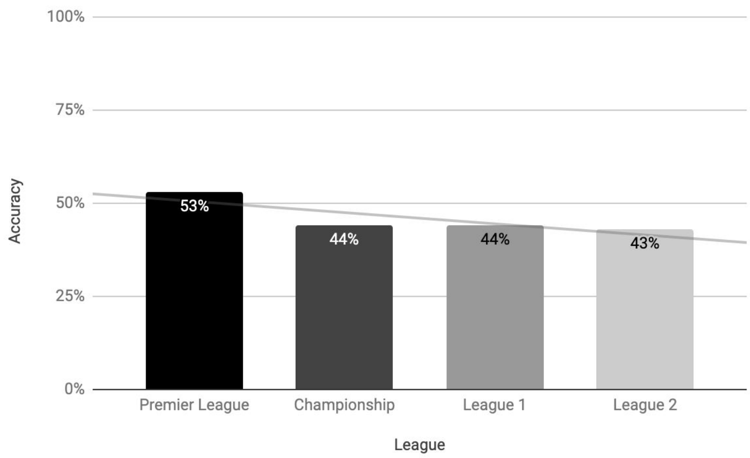


4.2 Post-hoc analysis

The remainder of this section covers the post-hoc analysis and is shaped by a more in depth analysis of first, the performance of the model for different leagues, and second, the performance of the model for different clubs. For this analysis the 41-10-3 model is saved and used as it is one of the faster and best performing models.

The first part of the post-hoc analysis investigates the accuracy for different leagues with the saved, 41-10-3 model. Results indicate that matches of the Premier League are, with 53%, most predictable, followed by Championship and League 1, with both 44%. Results indicate that the League 2 is least predictable with an accuracy of 43%. The accuracy for the different leagues are visualized in Figure 4. From the related work, as mentioned in section 2, only McCabe and Trevathan (2008) predicted matches of the Premier league and found an accuracy of 54% and Tax and Joustra (2015) was the only one that incorporated the bookmakers odds and found an accuracy of 55%, which is similar to the performance of the proposed 41-10-3 MLP that found an accuracy of 53%.

Figure 4
Accuracy for the Premier League, Championship, League 1 and League 2



Exploring the differences between leagues in more depth by comparing season 2016-2017 and 2017-2018 results in Table 14. Table 14 indicates that the largest difference in accuracy between leagues in 2016-2017 versus 2017-2018 is 6% in the Premier League and 4% in League 2 while the the Championship and League 1 remain constant.

Table 14
Accuracy per league foe season 2016-2017 and 2017-2018.

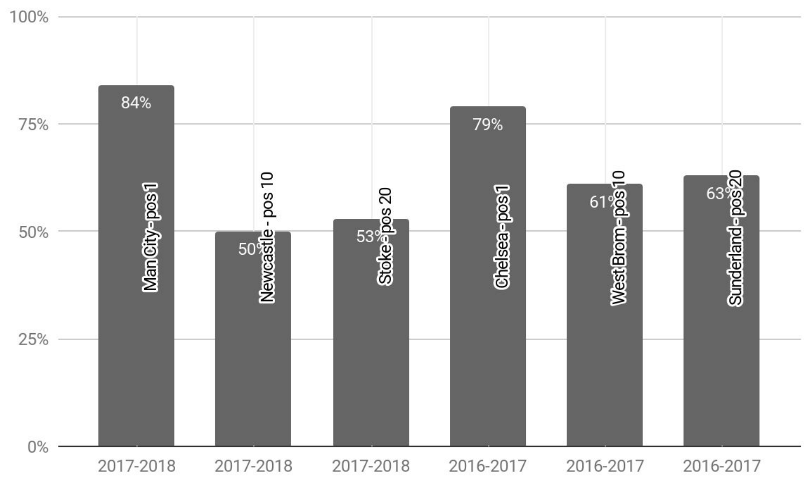
Season	League	Model	Accuracy
2017-2018	Premier League	45-10-3	54%
2017-2018	Championship	45-10-3	47%
2017-2018	League 1	45-10-3	43%
2017-2018	League 2	45-10-3	47%
2016-2017	Premier League	45-10-3	**60%**
2016-2017	Championship	45-10-3	47%
2016-2017	League 1	45-10-3	44%
2016-2017	League 2	45-10-3	43%

The second part of the post-hoc analysis investigates the accuracy for different teams by using the same saved 45-10-3 model. Teams have been selected based on rank in the Premier League. The 3 incorporated teams per season represent a team at the top of the ranking, a team at the center of the ranking and a team at the bottom of the ranking in the Premier League. Figure 5 illustrates the accuracy for predicting matches of Man City, Newcastle and Stoke for the season 2017-2018 and Chelsea, West Brom and Sunderland during the 2016-2017 season.

Further analysis shows that within the 2016-2017 and 2017-2018 season, matches of the top teams, Chelsea and Man City, are most predictable, followed by the teams in bottom position, Sunderland and Stoke. Teams in a middle position, West Brom and Newcastle, are predicted least accurate. Figure 5 shows the accuracy for these teams with the corresponding season.

Figure 5

Accuracy for the top team, middle team and bottom team of the Premier League for season 2016-2017 and 2017-2018



5. Discussion

The main goal of this study is to investigate to what extent a neural network is able to predict the outcome of football matches as a home win, away win or a draw. To investigate this matter different neural networks have been created that differ in deepness, number of neurons per layer and output.

An interesting main finding that emerges from the analysis, is that the proposed networks with 3 output neurons are able to predict football matches with an accuracy of 48% while the proposed models with 1 output neuron achieves a 43% accuracy for predicting matches of the Premier League, Championship, League 1 and League 2. These results suggest that neural networks with 3 output neurons are more appropriate for football result prediction than neural networks consisting of 1 output neuron. This suggestion is somewhat unexpected as prior studies only created networks that consist of 1 output neuron. Results also suggest that for football prediction models 1 hidden layer is sufficient which is in line with the network architectures of Aslan and Inceoglu (2007), Huang and Chang (2010) and McCabe and Trevathan (2008), and that the number of neurons per layer is sufficient when it is between the size of the input neurons and output neurons which is in line with Heaton (2009) and McCabe and Trevathan (2008).

At first it seemed somewhat surprising that the different networks did not outperform one of the models by Arabzad et al. (2014), Aslan and Inceoglu (2007), Huang and Chang (2010), McCabe and Trevathan (2008) and Tax and Joustra (2015). but post-hoc analysis revealed a similar performance of 53% accuracy when the model is used to predict only Premier League matches. This performance is similar to the baseline performance of McCabe and Trevathan (2008) and Tax and Joustra (2015). with respectively 54% and 55% accuracy.

Post-hoc analysis revealed that the different football leagues in England differ in predictability and suggest that the Premier League is the most predictable competition, followed by the Championship, League 1 and League 2. The reason for this is not within the scope of this study but it seems logical that this is due to larger variances in features between teams of the Premier League and the other leagues. This idea could also be the reason for the differences in predictability between a club at the top of the ranking and a club at the center of the ranking.

A first limitation of this study may be the incorporated features and creation of these features and consist of at least 5 points of discussion. The first one is that the team dependent features are calculated for home and away teams separately due to limited resources. As this might be beneficial for some features, it seems logical that this is not the case for other features e.g. distance since last match. Secondly, Constantinou et al. (2012), indicate that the distance since the last match, is only 1 out of 4 components that attribute to the total fatigue of a team. A more sophisticated construction could attribute to a more accurate prediction. The third point of discussion comprehends the team strength. Even though the current team strength feature, spi, does incorporate the clubs investment, the value is a constant. As team performance and club investment changes per season this can be seen as a limitation of the current study. A fourth point of discussion is that current study did not incorporate player availability. McCabe and Trevathan (2008) incorporated this feature and indicates that the outcome of a football match is affected by player availability. The fifth and final point of discussion related to the features is the absence of attendance. The attendance or crowd is a features that affects the results of a football match at least in two ways. The first one is that crowd support attributes to home advantage by encouraging players' performance (Ponzo & Scoppa, 2016). The second way that the crowd affects football results is by influencing referee decisions in favor of the home team (Goumas, 2012; Ponzo & Scoppa, 2016). As the attendance was available for only a limited amount of matches, this feature has been removed from the dataset.

A second limitation might be the simplicity of the proposed neural networks There are different optimizers and hyperparameters that can be applied to a neural network. As hyperparameter tuning was not within the scope of this study it is most likely that the performance of the model can be enhanced by tuning different hyper parameters.

This study shows that there is abundant room for further progress in determining to what extent a neural network is able to predict the outcome of football matches in one of the following, or a combination of 2 ways. First, to investigate football result prediction by using a neural network with 3 output neurons as the existing body of literature predicts the outcome of football matches with models consisting of 1 output neuron. Second to enhance incorporated feature vector by adding more features and by replacing current features with more accurate features.

6. Conclusion

The present study is designed to investigate to what extent a neural network is able to predict the outcome of football matches. The lack scientific papers towards this matter and the limitations of existing studies offers possibilities to investigate football result prediction by deep learning algorithms.

To investigate the predictability of football matches a dataset consisting of 24.029 matches and 41 features has been created. Incorporated feature vector comprehends team dependent features and team independent features. The team dependent features contain team strength, cumulative sum for different variables during the season, form and distance between matches for the home team and for the away team. Team independent features consist of the referee, bookmakers odds, season and division. Data is shuffled and randomly split into a 60% train, 20% validation and 20% test set. Different neural networks have been created that differ in deepness, neurons per layer and output neurons. To prevent overfitting, loss is monitored and the number of epochs has been empirically established.

This study suggests that neural networks are able to predict the outcome for football matches of the Premier League, Championship, League 1 and League up to 48%. The main finding is that neural networks that contain 3 output neurons are better able to predict the outcome of football matches than neural networks with 1 output neuron with respectively 48% and 43% accuracy. Results also indicate that deeper models do not enhance performance and dat the number of neurons per layer is sufficient if the number of neurons are between the number of input features and the number of output features. Post hoc analysis revealed that the predictability of football matches may differ per league, per season and per club and that the proposed 41-10-3 MLP architecture achieves similar performance as the baseline performance of Tax and Joustra (2015) that also included bookmakers odds and McCabe and Trevathan (2008) that also predicted matches of the Premier league.

7. Future work

The main implication, that is supported by the results of this study, is that a neural network that consists of 3 output neurons is better in predicting football matches than a neural network that consists of 1 output neuron. As prior research investigated football result prediction with neural networks shaped out of 1 output neuron, this result is somewhat unexpected and offers, in combination with the limitation of current features and simplicity of the proposed models, abundant room for future research.

References

Arabzad, S., Soheil, S., Araghi, M., & Ghofrani, N. (2014). Football Match Results Prediction Using Artificial Neural Networks; The Case of Iran Pro League. *International Journal of Applied Research on Industrial Engineering,1*(3), 159-179.

Aslan, B., & Inceoglu, M. (2007). A Comparative Study on Neural Network Based Soccer Result Prediction. *Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)*. doi:10.1109/isda.2007.12

Boice, J. (2018, August 10). How Our Club Soccer Predictions Work. Retrieved April 15, 2019, from https://fivethirtyeight.com/methodology/how-our-club-soccer-predictions-work/

Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007). Referee bias contributes to home advantage in English Premiership football. *Journal of Sports Sciences,25*(11), 1185-1194. doi:10.1080/02640410601038576

Constantinou, A. C., Fenton, N. E., & Neil, M. (2012). Pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems,36*, 322-339. doi:10.1016/j.knosys.2012.07.008

FIFA. (2018). Laws of the Game. Retrieved from https://www.fifa.com/development/education-and-technical/referees/laws-of-the-game.html

Fivethirtyeight. (2018). SPI Ratings. Retrieved from https://github.com/fivethirtyeight/data/tree/master/soccer-spi

Football-Data. (2019). Data Files: England. Retrieved from http://www.football-data.co.uk/englandm.php

Goddard, J. (2006). Who wins the football? *Significance,3*(1), 16-19. doi:10.1111/j.1740-9713.2006.00145.x

Goumas, C. (2012). Home advantage and referee bias in European football. *European Journal of Sport Science,14*(Sup1). doi:10.1080/17461391.2012.686062

Heaton, J. (2009). *Introduction to neural networks with Java*. Chesterfield (MO, USA): Heaton Research.

Heuer, A., & Rubner, O. (2009). Fitness, chance, and myths: An objective view on soccer results. *The European Physical Journal B,67*(3), 445-458. doi:10.1140/epjb/e2009-00024-8

Hirotsu, N., & Wright, M. (2003). An evaluation of characteristics of teams in association football by using a Markov process model. *The Statistician,* 52: 4, 591-602.

Huang, K., & Chang, W. (2010). A neural network method for prediction of 2006 World Cup Football Game. *The 2010 International Joint Conference on Neural Networks (IJCNN)*. doi:10.1109/ijcnn.2010.5596458

Mccabe, A., & Trevathan, J. (2008). Artificial Intelligence in Sports Prediction. *Fifth International Conference on Information Technology: New Generations (itng 2008)*. doi:10.1109/itng.2008.203

Minsky, M. L., & Popeit, S. (1988). *Perceptrons*. MIT Press. Retrieved April 11, 2019, from https://ia801601.us.archive.org/1/items/Perceptrons/Perceptrons.pdf.

Odachowski, K., & Grekow, J. (2013). Using Bookmaker Odds to Predict the Final Result of Football Matches. *Lecture Notes in Computer Science Knowledge Engineering, Machine Learning and Lattice Computing with Applications,*196-205. doi:10.1007/978-3-642-37343-5_20

Ponzo, M., & Scoppa, V. (2016). Does the Home Advantage Depend on Crowd Support? Evidence From Same-Stadium Derbies. *Journal of Sports Economics,19*(4), 562-582. doi:10.1177/1527002516665794

Reed, R. D., & Marks, R. J. (1999). *Neural smithing: Supervised learning in feedforward artificial neural networks*. Cambridge (Massachusetts): MIT.

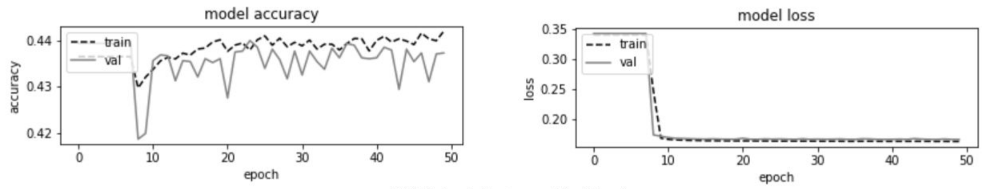
Tax, N., Joustra, Y.P., Predicting the Dutch football competition using public data: A machine learning approach, Trans. Knowl. Data Eng. 10 (10) (2015) 1– 13.

Ulmer, B., & Fernandez, M. (2013). Predicting Soccer Match Results in the English Premier League. Retrieved from http://cs229.stanford.edu/proj2014/ Ben Ulmer, Matt Fernandez, Predicting Soccer Results in the English Premier League.pdf

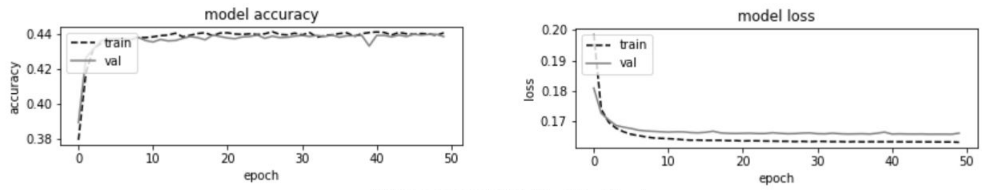
Vecer, J., Kopriva, F., & Ichiba, T. (2008). *Estimating the Effect of the Red Card in Soccer*. Retrieved from http://www.stat.columbia.edu/~vecer/redcard.pdf

Appendix A: Accuracy & Loss of MLP 1 output neuron

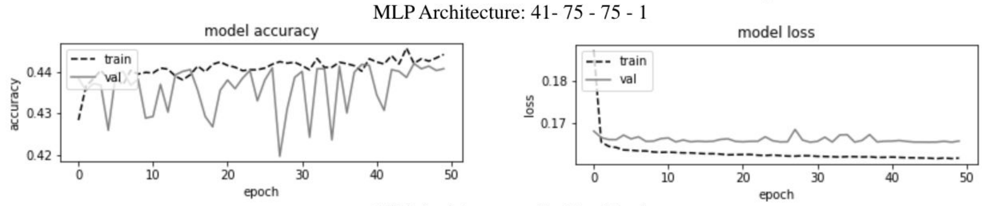
MLP Architecture: 41- 75 - 1



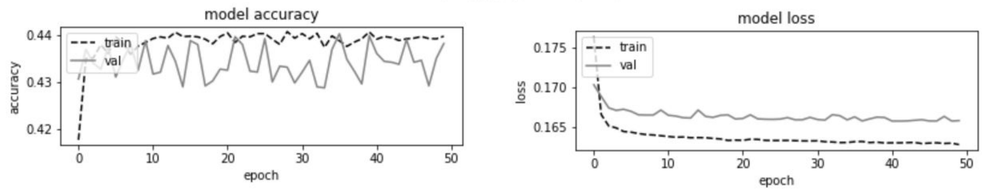
MLP Architecture: 41- 10 - 1



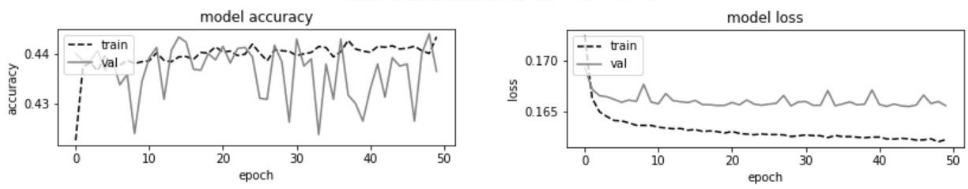
MLP Architecture: 41- 75 - 75 - 1



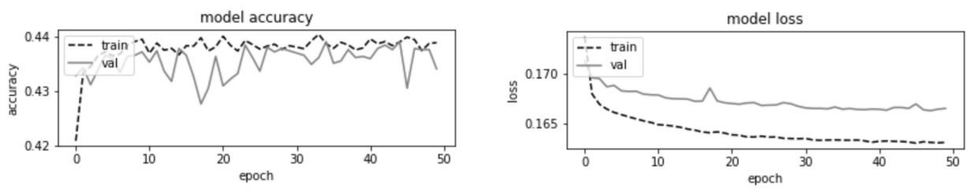
MLP Architecture: 41- 10 - 10 - 1



MLP Architecture: 41- 75 - 75 - 75 - 1

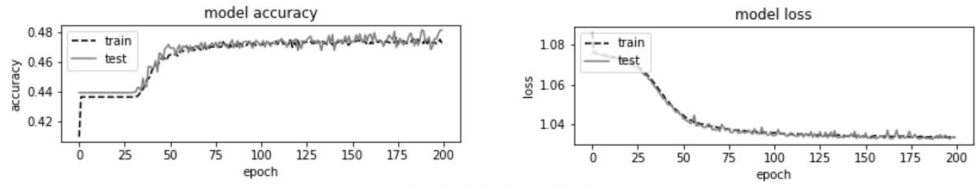


MLP Architecture: 41- 10 - 10 - 10 - 1

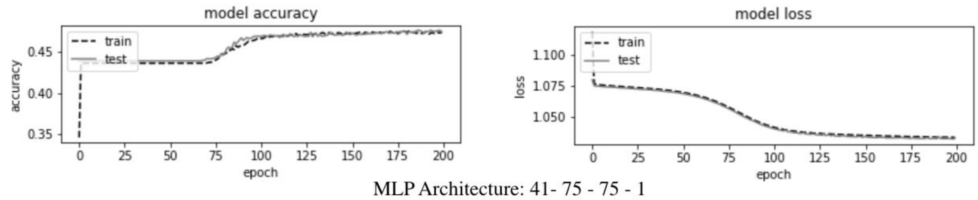


Appendix B: Accuracy & Loss of MLP 3 output neuron

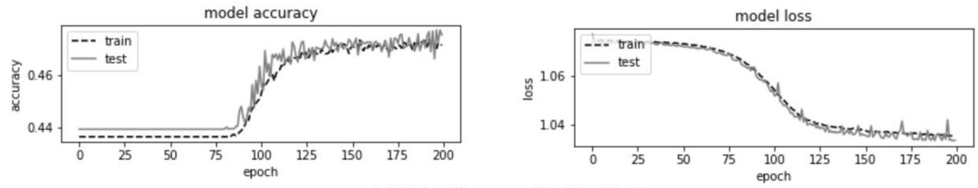
MLP Architecture: 41- 75 - 1



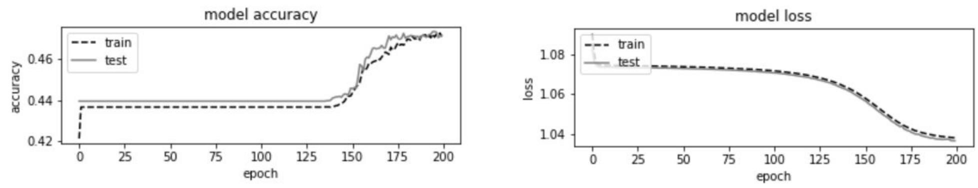
MLP Architecture: 41- 10 - 1



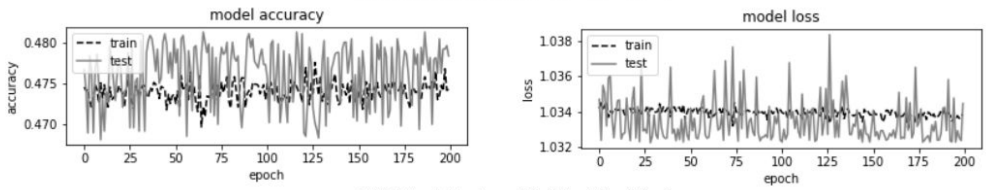
MLP Architecture: 41- 75 - 75 - 1



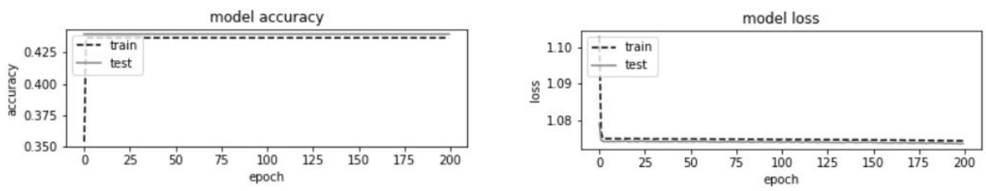
MLP Architecture: 41- 10 - 10 - 1



MLP Architecture: 41- 75 - 75 - 75 - 1



MLP Architecture: 41- 10 - 10 - 10 - 1



Appendix C: Python code

https://github.com/StefanSamba/CSAI_Thesis/