

DWIT COLLEGE
DEERWALK INSTITUTE OF TECHNOLOGY

(Affiliated to Tribhuvan University)



SIFAL, Kathmandu

Final Year Project Report

On

Premier League Game Result Prediction

For the partial fulfillment of the requirement for the degree of Bachelor of Science in Computer
Science and Information Technology awarded by Tribhuvan University

Under the supervision of

Asst. Prof. Sarbin Sayami

IOST, TU

Submitted by

Sumit Shrestha (1817/069)

Submitted to

IOST, TU

August, 2016

Student's Declaration

I hereby declare that I am the only author of this work and that no sources other than the listed here have been used in this work.

... ..

Sumit Shrestha

Date:

Supervisor's Recommendation

I hereby recommend that this Project prepared under my supervision by Sumit Shrestha entitled "Premier League Result Prediction" in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Information Technology be processed for the evaluation.

... ..

Asst.Prof. Sarbin Sayami
IOST, TU

Date:

Certificate of Approval

The undersigned certify that he has read and recommended to the Department of Computer Science for acceptance, a project report entitled Premier League Game Result Prediction submitted by Sumit Shrestha in partial fulfillment for the degree of Bachelor of Science in Computer Science & Information Technology.

.....

Asst. Prof. Sarbin Sayami

Supervisor

IOST, TU

.....

External Examiner

.....

Mr. Hitesh Karki

Chief Academic Officer

Deerwalk Institute of Technology

Acknowledgements

I would like to give my thanks to our highly respected guide Asst. Prof. Sarbin Sayami for his valuable guidance who gave me great encouragement for this work and his suggestions were very much helpful. I would like to express my sincere thanks to Mr. Hitesh Karki who helped me directly or indirectly during this project work.

Sumit Shrestha

TU Exam Roll no: 1817/069

Abstract

Prediction is a statement of an uncertain event which uses past data, analyze it and predicts the future. This application focuses on predicting game result of the premier league on the basis of Back Propagation Algorithm . The input parameters chosen were Home Teams, Away Teams, Home Team Goals, Away Team Goals and Goal Differences. The accuracy of the system was found to be 47 percent.

Keywords: *Prediction, Back propagation, Premier League*

LIST OF FIGURES

Figure 2.1 : Use Case Diagram Of Premier League Game Result Prediction	20
Figure 2.2 : Activity Network Diagram of Premier League Game Result Prediction.....	21
Figure 3.1: Method for Prediction	23
Figure 3.2: Neural Network Of Premier League Result Prediction	28
Figure 3.3: Back Propagation Algorithm	30
Figure 3.4 : Sequence Diagram of Premier League Game Result Prediction.	30
Figure 3.5 : Event Diagram Of Premier League Game Result Prediction	31
Figure 3.6 : Class Diagram Of Premier League Game Result Prediction.....	32

LIST OF TABLES

Table 2.1: Functional and Non-Functional Requirement	19
Table 3.1: Original Data Taken From Website	24
Table 3.2: Selected Data	25
Table 3.3: Normalized Data	25
Table 3.4: Sample Training Data.....	26
Table 3.5: Trained Result Output	26
Table 3.6: Sample Testing Data	27
Table 3.7: Test Output Result	27
Table 4.1: Trained Data Used For Implementation	33
Table 4.2: Trained Data Result Used For Implementation	34
Table 4.3: Test Data Used For Implementation	34
Table 4.4: Test Data Result Used For Implementation	35
Table 4.5: Unit Testing of Premier League Game Result Prediction.....	39

TABLE OF CONTENTS

Chapter 1: Introduction.....	12-13
1.1 Background.....	12
1.2 Problem Statement.....	13
1.3 Objectives	13
1.4 Scope.....	13
1.5 Limitation	13
1.6 Outline of Document	13
Chapter 2: Requirement Analysis and Feasibility.....	15-21
2.1 Literature Review	15
2.3 Requirement Analysis.....	19
2.3.1 Functional Requirement	19
2.3.2 Non Functional Requirement.....	19
2.4 Feasibility Analysis	20
2.4.1 Technical Feasibility.....	20
2.4.2 Operational Feasibility	20
2.4.3 Schedule Feasibility.....	21
Chapter 3: System Design.....	22-32
3.1 Methodology.....	22
3.1.1 Data Collection and Normalization	23
3.1.2 Training	25
3.1.3 Testing	26
3.1.4 Prediction.....	28
3.1.5 Algorithm.....	29
3.2 System Design	30

3.2.1 Sequence Diagram	30
3.2.2 Event Diagram	31
3.2.3 Class Diagram	32
Chapter 4: Implementation and Testing.....	33-39
4.1 Implementation	33
4.1.1 Tools Used	35
4.1.2 Description / Listing of major classes / methods / any other modules	37
4.2 Testing	39
4.2.1 Unit Testing	39
Chapter 5: Maintenance and Support.....	41
5.1 Adaptive Maintenance	41
5.2 Perfective Maintenance	41
Chapter 6: Conclusion and Recommendation.....	42
6.1 Conclusion	42
6.2 Recommendation	42
References.....	43
Appendix.....	45-47
Appendix - 1: Home Page	45
Appendix - 2: Selecting the teams.....	46
Appendix - 3: Result.....	47

ABBREVIATIONS AND ACRONYMS

PLS : Pseudo-likelihood Statistic

KNN: K Nearest Neighbors

ANN: Artificial Neural Network

LR: Logistic Regression

MLP: Multilayer Perceptron

JSP: Java Server Page

IDE: Integrated Development Environment

CSS: Cascading Style Sheet

HTML: Hyper Text Markup Language

Chapter 1: Introduction

1.1 Background

Football is one of the most famous games for the people worldwide. It is a game played by two teams, who needs to hit the ball on its opponent team's goal post. The way the football is getting worldwide and the fans are also crazy about it. Audience wants to know who wins the game and who loses. Prediction of such football game is hence popular. Also many people do the betting for the favourite team. They simply see their luck keeping the money on the team and they actually don't know who will win and who will lose. But, it's actually difficult to predict the result of the game. Many of the football experts also finds difficult to predict the score line or result of the game. The league that is famous among all the football lovers is English Premier League where 20 English teams compete with each other to win the league. (Reilly, Thomas; Gilbourne, 2003)

The craziness is increasing day by day and each time the teams wants to focus on the league. The points of the game is simple whichever team wins get 3 points and the team that loses get 0 point and if the game is draw then each team will share one point each. The team with highest point at the last wins the league. Each time different winner is seen in every year, but the most dominant team is Manchester United with 13 Premier League's trophies. The bottom three teams are relegated and replaced by other teams from lower leagues who perform better. Each team plays every other team twice, once at home and once away. Premier League is one of the most-watched football (soccer) league today within the world which is the league between the states of England clubs. It is the league between 20 teams where bottom 3 teams gets relegated to another sub division league of England. Prediction makes the game more really exciting and focused to the supporting team for the wait of the chances of the team to win the game. The prediction will be focused within the teams of English Premier League only. Premier League Game Result Prediction predicts the games that happens in every weekends. This makes the result of the game of within two teams as win, lose or draw. (F. Halicioglu, 2005) predicted Euro 2000 winner and conformed that the prediction is possible and hence the prediction was started.

1.2 Problem Statement

Premier League fans are more excited to know the results of the team. They can be more enthusiastic by knowing the result before the game. However, the main problem is to make the accurate result of the prediction between the teams.

1.3 Objectives

The main objective of the Premier League Result Prediction are:

- To predict the result of each game of the team of the premier league.
- To implement back propagation algorithm for training and testing the model.

1.4 Scope

This prediction will be related to English Premier League only.

1.5 Limitation

The limitation of the Premier League Result Prediction are:

- It can't predict the result of leagues and cups matches like Spanish League, French League, German League, World Cup, Euro Cup, etc.
- It can't predict the score line.
- It doesn't contain parameters like individual players details, key players, player transfer details.

1.6 Outline of Document

The remaining part of the document consists of following:

Chapter 2 describes the factors and the result of the prediction that are done using neural network and back propagation algorithm and also it describes the feasibility of this application.

Premier League Result Prediction

Chapter 3 describes the methodology that needs to be done while implementing the application and also the system design like class diagram, sequence diagram and event diagram is described.

Chapter 4 describes the real implementation, tools that are being used, major classes and testing like system testing, unit testing and integration testing of the application.

Chapter 5 describes the changes that need to be done for the maintenance of the application

Chapter 6 describes the result and the accuracy that the application gives while predicting the result of the game. Similarly, the chapter ends with the recommendation for improving the accuracy of the application.

Chapter 2: Requirement Analysis and Feasibility

2.1 Literature Review

(Aditya Srinivas Timmaraju, Aditya Palnitkar, VikeshKhanna, 2014) estimated that 4.7 billion people watched 2010-11 season. (FarzinOwramipur, ParinazEskandarain, and Faezeh Sadat Mozneb, 2015) used Bayesian Network to predict the score of Spanish Team, Barcelona. They found the model probability and relationship among the given domain. They collected information from valid websites that offered the statistics of the football. They also used two factors for the match prediction psychological and non-psychological factors through which they can predict the final result.

Logistic Regression Concepts

(Igiri, Chinwe Peace, Nwachukwu, Enoch Okechukwu, 2014) used Artificial Neural Network (ANN) and logistic regression (LR) techniques with Rapid Miner as data mining tool which result of 85% and 93% prediction accuracy respectively. They compared the existing system of prediction with their system and found that their system were twice more accurate than the current existing system. They took certain factors that affects the result of the match like Home advantage effect on team's performance, the effect of injuries of key players on team performance, effect of external cup on league performance. They have seen these factors and related that the effect of the certain factors can also be seen in the team. (YueWengMak, 2013) used a three multilayer perceptron concept and took factors like last 5 matches of each team and their last 3 encounters between the team and added home advantage and ranking inputs to predict the result which predicted that the more the consideration of the factors the more likely the prediction will be correct.

Bayesian Network Concepts

(A. Joseph, N.E. Fenton, M. Neil, 2006) used Bayesian network approach to predict the result of Spurs only. They predicted containing certain factors of Bayesian network like MC4 Learner, Naïve Bayesian Network, Hugin Bayesian Network, Expert Bayesian network and KNN (K Nearest Neighbour) algorithm. They used the process of machine learning with two

Premier League Result Prediction

tangible benefits, understanding and prediction. The MC4 learner identifies those attributes, which have the largest effect on the outcome of the game. It shows their relationships to each other in terms of their effect on the outcome of the game. This is a very simplified model of the game itself. The naïve Bayesian learner doesn't construct a model as such its model is predefined. The learning process for the naïve Bayesian learner is then simply one of discovering the relative strength, and polarity, of the effect of each attribute with respect to the result. KNN does not construct a model as such it simply uses the existing data and provides a likeness comparison with any test data. Thus KNN doesn't significantly enhance our understanding. The expert constructed Bayesian Network represents the knowledge of the expert, that is, it is a model is the expert's belief of the interrelationships between the attributes and their relative importance.

(Jeffrey Alan Logan Snyder, 2013) reviewed past and current research efforts in soccer prediction, categorizing their approaches and conclusion. It treats a match as a result-producing black box, ignoring the noisy, but beautifully complex processes that contribute to each shot and goal. As XY data makes its way into the hands of researches, more detailed models of in-game processes may be built. This will in turn open up new problems and avenues for analysis, hopefully leading to a deeper understanding of the game itself. It also has begun the process of determining the relative importance of these types of data in prediction and analysis. They also have presented an approximately optimal betting strategy for use in betting simultaneously on multiple games with mutually-exclusive outcomes, which performs substantially better than other strategies used in academic betting simulations.

Probabilistic Concepts

(Aditya Srinivas Timmaraju, Aditya Palnitkar, VikeshKhanna, 2014)selected the important characteristics for a feature they are Incorporate a notion of the competing nature of the problem, Be reflective of the recent form of a particular team and Manifest the home advantage factor. They took three approaches and they are:-

Premier League Result Prediction

a) Approach 1:-The first approach they took was using Multinomial Logistic Regression. They considered the performance metrics derived from the current match, rather than taking the average over the last “k” matches. During testing, they predicted the match outcome of team A vs team B, where they arrived at the feature vector using KPP.

b) Approach 2 :-In second approach, they trained the same way and more precisely, in the training phase too, instead of using the feature vector as the performance metric vector corresponding to the current match, they used KPP. This meant that the trained parameters now inform their beliefs about the result of a match based on the performance in last “k” matches. In this approach, they also used TGKPP instead of KPP.

c) Approach 3 :-All approaches they tried will find a global set of parameters, which were independent of the competing teams. So, the given past “k” performances of the team playing home and the team playing away, our model was agnostic to the identity of the actual teams playing. They felt they were missing some team-specific trends using this method. So, in this approach, they trained different models for the different teams. However, this approach placed a limitation on the data they could use to test/train our model. They could no longer combine data from two different seasons due to the form of the team varying between seasons, and major players being traded between teams. So, due to the limited data, and the increased noise induced by increasing the granularity in the model, they ended up getting a lower accuracy (an average of 47%).

With these approaches, they proposed a metric, which is computed as the geometric mean of the predicted probabilities of actual outcomes, which is called PLS. They obtain a PLS of 0.357 for EPL which had the best value of PLS as 0.36007. (AlbinaYezus, 2014) divided the work in the steps as choosing match set that is to be analyzed, deciding on key features, data extraction, testing various machine learning algorithms and improving the implemented algorithms. He found out that it is possible to find a classifier that predicts the outcome of soccer matches with the precision of more than 60%.

Markov chain and Monte Carlo Concept

(Havard Rue and Oyvind Salvesen, 1999) used MCMC (Markov Chain and Monte Carlo Methods) in which they produced as an irreducible aperiodic transition kernel. They presented that this approach seems superior to the earlier attempts to model the football games as it allows for coherent inference of the properties between the teams easily account for the joint uncertainty in the variables which is important in prediction allows for doing various interesting retrospective analysis of a season and finally provides a framework, where it is easy to change parts or parameterization in the model. They also have improved on the data, parameter estimation, goal model and home field advantage, which make that the point of the prediction highly rises than before with these parameters. (Douwe Buursma, 2013) found out the selection of the feature set led to the important insights and the performance of the classifier kept improving the bigger set of recent matches which showed the optimal number of matches to lie around 10, 20 and performance keeps improving. The history of opponents of home teams does not seem to play as important a role as the history of opponents of the away team. He trained a number of data, which made the system fare, which makes that the analysis can use more data. The feature set like Match History, Classifiers helps to get the correct model and it also helps to make the data more useful for the training set. The Classifiers classifies all matches as home wins, draws or away wins, depending on the features belonging to that match. He used classifiers like ClassificationViaRegression, MultiClassClassifier, Rotation Forest, LogitBoost, BayesNet, NaiveBayes and Home Wins. He concluded that the football matches would always be very hard to predict. (Francisco Louzadal, Adriano K. Suzuki and Luis E.B. Salazar, 2014) & (Brijesh Kumar Bhardwaj, Saurabh Pal, 2011) used the data mining concept for the prediction over the improvement using classification. In data mining process, they first prepared the data over the size of 300 and they selected only the field that was necessary for data mining. The predictor and variables were derived from the database and they used Bayesian Classification for the implementation of the mining model. (Gianluca Baio, Marta A. Blangiardo, 2010) used Bayesian network and poisson distribution for the analysis of the model where the parameters defined was home advantage, the scoring intensity. They have showed that the home team have greater potential to win and the team with the low points are

to concede more goals at home and lose. The estimation of the team was poor when bivariate poisson was used and the models parameters perform better than bivariate poisson.

2.3 Requirement Analysis

Table 2.1: Functional and Non-Functional Requirement

2.3.1 Functional Requirement	2.3.2 Non Functional Requirement
Predicts the result of the game between two teams.	Prediction is based on the basis of team name.
Compares the result of the data that are trained.	A result will be shown on which team will win, lose or draw the game.

The basic functionality of this application is that it predicts the result of the game between two teams. The results are then compared with the data that are being trained which is shown in Table 2.1. The non-functional requirement includes displaying of the input field and a predict result button, then a result will be shown comparing the data with the trained data which is shown in Table 2.1.

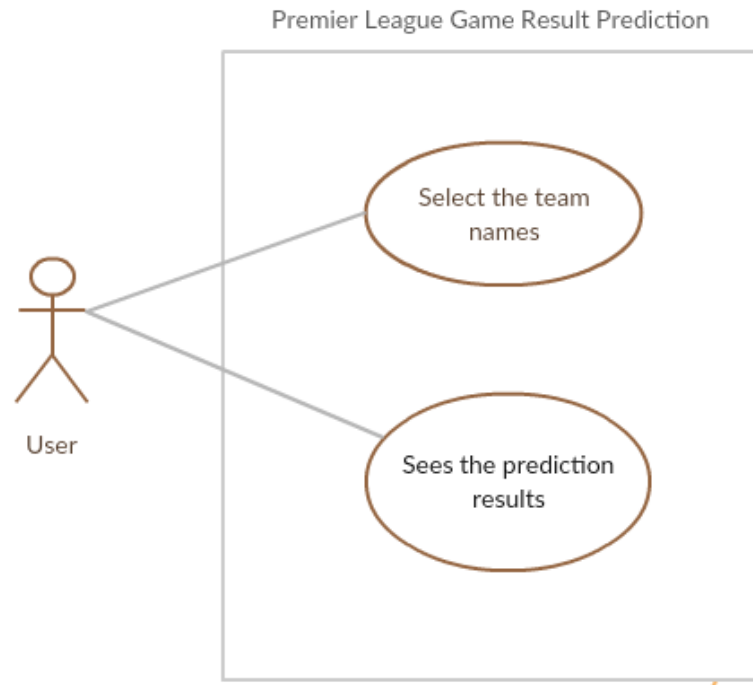


Figure 2.1 : Use Case Diagram Of Premier League Game Result Prediction

The Figure 2.1 describes about the use case diagram of the system. The figure illustrates the user who did the action i.e. Enter the team names, and sees the prediction results.

2.4 Feasibility Analysis

2.4.1 Technical Feasibility

Premier League Game Result Prediction is an application based on the implementation of the backpropagation algorithm. It uses gradient descent, multilayer perceptron concept and Java Servlet and Java Server Page (JSP) as a front end to display the content for the user. All of the technology that are need by Premier League Game Result Prediction are available and are open source which can be accessed freely so it is technically feasible.

2.4.2 Operational Feasibility

Premier League Game Result Prediction has a simple and user friendly user interface which makes it easy to use. The user can see the result by inserting the team names in the availability of the internet which makes it operationally feasible.

Premier League Result Prediction

2.4.3 Schedule Feasibility

The time allocated of the Premier League Game Result Prediction application is within 47 days and shown as below:

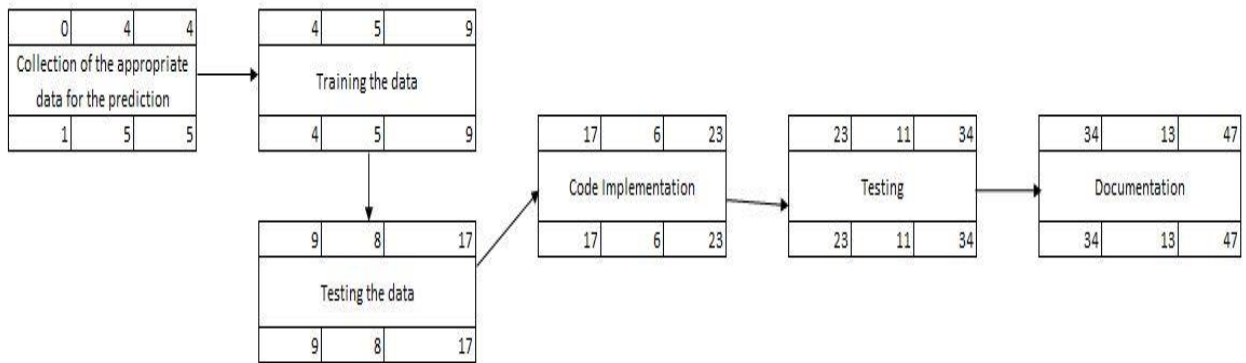


Figure 2.2 : Activity Network Diagram of Premier League Game Result Prediction

In Figure 2.2, we can see that Premier League Game Result Prediction application was completed within 47 days which is within schedule and hence it is schedule feasible.

Chapter 3: System Design

3.1 Methodology

With all the literature review, the most common thing was choosing the correct parameter is the first way to get the prediction. The more the parameters, the more chances of getting the result of the prediction correct. The prediction that was difficult for the experts have made some easy task due to several prediction methods. The parameters or factors such as home advantage, injuries of the players, cup game effect on league, team recent form and the head-to-head matches between the opponent need to be analyzed which adversely affect the result of the match. (Igiri, Chinwe Peace, Nwachukwu, Enoch Okechukwu, 2014) used the Hidden Markov Process Model and Ordered Probit Regression model with only three parameters like home advantage, injury of the key players and cup game effect on league. I will be using the same models but with more parameters like the team recent form and the head-to-head matches between the opponents which might affect the accuracy of the prediction then (Igiri, Chinwe Peace, Nwachukwu, Enoch Okechukwu, 2014) . With lots of data, the data-mining tool is used to extract the information. Artificial Neural Network (ANN) and Regression techniques are two data-mining techniques that will be used. I will first collect the previous results of the matches with every history of the team and the teams they played with. Then, from the collected data will extract the features such as Home and Away Goal Difference, Points, Attack and Defense Skills, which is not needed for the techniques. Then a collective database will be built to collect all the necessary data and will be kept in MS Excel spreadsheet. Using an algorithm will use the parameters to adjust the training algorithms to make it more useful. The data-mining tool, which holds the better results and improves the learning rate, also finds the optimal values and an improved model will be built. The two models Artificial Neural Network (ANN) and Logistic Regression (LR) are used to make the improved models for the system. So, the record of the match and the result of the match will be predicted by the use of ANN and LR model. The accuracy was 75.04% according to (Igiri, Chinwe Peace, Nwachukwu, Enoch Okechukwu, 2014) but after the addition of the parameters that helps to estimates the model more might increase the accuracy by 85%.

Premier League Result Prediction

So, with the known of the several factors, I came to use different input variables, output variables, algorithms for the process of prediction.

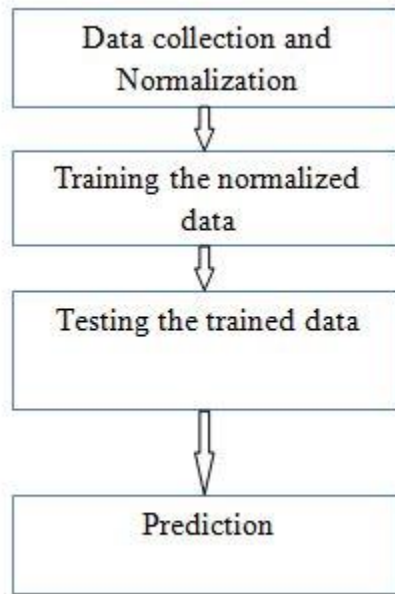


Figure 3.1: Method for Prediction

Figure 3.1 illustrates the method that needs to be followed while making the prediction. First, the data is collected and it is being normalized by using sigmoid function. Then the normalized data is trained by using neural network and the trained data are being tested for the prediction accuracy. Then after the testing the prediction can be done.

3.1.1 Data Collection and Normalization

The data are collected from the official premier league website i.e. www.premierleague.com & <http://www.football-data.co.uk/englandm.php>. The premier league started from 1993 and there are overall 9000 data of the premier league starting from 1993. The data that are collected contains unnecessary parameters that need to be filtered out. The necessary data are only collected. Then, the collected data are normalized. The process adopted for normalizing the data is given below:

$$\text{Normalized value} = (x_i - \min(x)) / (\max(x) - \min(x))$$

Premier League Result Prediction

After normalization of the data the input and output parameter are selected which have major impacts on the accuracy of the prediction.

The input variables chosen for training, will have a heavy impact on the prediction of the result. The input variables that were used are

- Team Name
- Previous game home score
- Previous game away score
- Head-to-head home games
- Head-to-head away games

The neural network architecture was trained with the given data set with the above five variables as input parameters. Supervised learning approach was adopted to train the network and the network can predict the result based on the output combination of three nodes of output layer. The pattern of the output would determine the result whether it is the case of win, lose, or the draw game.

Table 3.1: Original Data Taken From Website

Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	Referee	HS	AS	HST	AST	HF	AF	HC	AC	HY	
EO	16/08/14	Arsenal	Crystal Pa	2	1	H	1	1	D	J Moss	14	4	6	2	13	19	9	3	2	
EO	16/08/14	Leicester	Everton	2	2	D	1	2	A	M Jones	11	13	3	3	16	10	3	6	1	
EO	16/08/14	Man Unite	Swansea	1	2	A	0	1	A	M Dean	14	5	5	4	14	20	4	0	2	
EO	16/08/14	QPR	Hull	0	1	A	0	0	D	C Pawson	19	11	6	4	10	10	8	9	1	
EO	16/08/14	Stoke	Aston Vill	0	1	A	0	0	D	A Taylor	12	7	2	2	14	9	2	8	0	
EO	16/08/14	West Bror	Sunderlan	2	2	D	1	1	D	N Swarbr	10	7	5	2	18	9	6	3	3	
EO	16/08/14	West Ham	Tottenhar	0	1	A	0	0	D	C Foy	18	10	4	4	12	10	8	5	1	
EO	17/08/14	Liverpool	Southamp	2	1	H	1	0	H	M Clatten	12	12	5	6	8	11	2	6	1	
EO	17/08/14	Newcastle	Man City	0	2	A	0	1	A	M Atkinso	12	13	0	5	8	11	3	3	1	
EO	18/08/14	Burnley	Chelsea	1	3	A	1	3	A	M Oliver	9	11	2	3	6	7	4	3	1	

Table 3.1 shows the original data taken from the websites which needs to be refined and only usable parameters are selected.

Table 3.2: Selected Data

HomeTeam	AwayTeam	HomeGoals	AwayGoals	GD	Results
0.07	0.4	0	3	0.4	2
0.08	1.02	4	1	2.98	1
0.35	0.16	1	2	0.84	2
0.74	1.09	2	0	0.91	1
0.78	0.7	1	1	0.3	3
0.87	1.24	0	1	1.24	2
0.94	0.67	0	3	0.67	2
1.09	1.21	3	1	1.79	1
1.11	0.51	0	2	0.51	2
1.3	0.04	0	2	0.04	2

Table 3.2 shows the refined data with only the necessary parameters but without normalizing the data. Home Team, Away Team, Home Team Goals, Away Team Goals and Goal Differences are the necessary parameters chosen from the refined data.

Table 3.3: Normalized Data

0.16	0.45	0	0	2	0
0.4	0.35	2	1	0	1
0.51	0.08	0	0	2	0
0.8	0.71	2	2	2	0
0.82	0.7	0	0	2	0
0.87	0.34	0	0	2	0
1.09	1.3	0	1	1	1
0.04	1.24	3	1	0	2
1.11	0.74	1	2	1	1
0.07	0.91	2	1	0	1

Table 3.3 shows the refined data that are normalized.

3.1.2 Training

The data that were collected from official premier league website i.e. www.premierleague.com & <http://www.football-data.co.uk/englandm.php> of which 80% of the data were selected in this application for the training purpose. The normalized data were divided into two parts as an input and output parameter. Since, it is a supervised learning, the output and input is already known.

Table 3.4: Sample Training Data

HomeTeam	AwayTeam	HomeTeamGoals	AwayTeamGoals	GoalDifference
0.07	0.4	1	4	3
0.08	1.02	5	2	3
0.35	0.16	2	3	1
0.74	1.09	3	1	2
0.78	0.7	2	2	0
0.87	1.24	1	2	1
0.94	0.67	1	4	3
1.09	1.21	4	2	2
1.11	0.51	1	3	2
1.3	0.04	1	3	2

Table 3.4 shows the sample input data that are being trained. The selected data are kept as the input parameter for the training set.

Table 3.5: Trained Result Output

Win	Draw	Lose
0	1	0
1	0	0
0	1	0
1	0	0
0	0	0
0	1	0
0	1	0
1	0	0
0	1	0
0	1	0

Table 3.5 shows the output of the three nodes in the output layer which determines the result of the game. The above output represents win, loss, and draw depending upon the pattern of output nodes. The three combinations as 000 represents loss, 010 represents draw, and 100 represents win case during the prediction.

3.1.3 Testing

From the available 9000 data sets, 20% of the data were selected in this application for the testing purpose. The normalized data were divided into two parts as an input and output parameter. Since, it is a supervised learning, the output and input is already known.

Table 3.6: Sample Testing Data

Home Team	Away Team	Home Team Goals	Away Team Goals	Goal Difference
0.87	1.29	3	3	0
1.17	1.32	0	1	1
1.24	0.15	2	1	1
1.3	1.18	0	3	3
1.34	0.16	2	3	1
0.16	1.34	1	2	1
0.56	0.08	0	0	0
0.74	1.18	1	1	0
0.87	0.07	0	0	0
1.02	0.18	0	4	4
1.32	0.9	1	1	0

Table 3.6 shows the sample input data that are being tested. The data are kept are the input parameter for testing purpose.

Table 3.7: Test Output Result

Win	Draw	Lose
0	0	1
0	1	0
1	0	0
0	1	0
0	1	0
0	1	0
0	0	1
0	0	1
0	0	1
0	1	0

Table 3.7 shows the output that are been seen while testing data. The data represents win, loss and draw. 010 represents draw, 100 represents win and 000 represents loss. These estimated output condition is then compared with the original target output condition to check the validity of the system. Output condition given by these test cases data set must be similar with the target output condition for all the data set of testing data for the accuracy of the system.

3.1.4 Prediction

After the completion of the training and testing phase, the next phase is prediction which predicts the result of the game by adjusting weight neural network and by entering the team names in the input field.

A Multilayer perceptron was chosen because we have to train the input variables and generate the output variables that are significantly originated by the hidden layers used in the multilayer perceptron. Multilayer Perceptrons (MLPs) are layered feed forward network. Multilayer Perceptron helps to make the mapping of input and output but it requires more set of training data and also requires more time. In this multilayer perceptron, the hidden layers are selected by hit and trail and more number of hidden layers will have low processing capability. During the training phase, first the hidden layer is set to the minimal and checking the value of cost functions, the number of hidden layers and number of nodes in each hidden layers is determined. Finally, the hidden layer with number of nodes in each layer is determined as such that the cost function reached minimum and the process is stopped and the network topology is set up for the prediction.

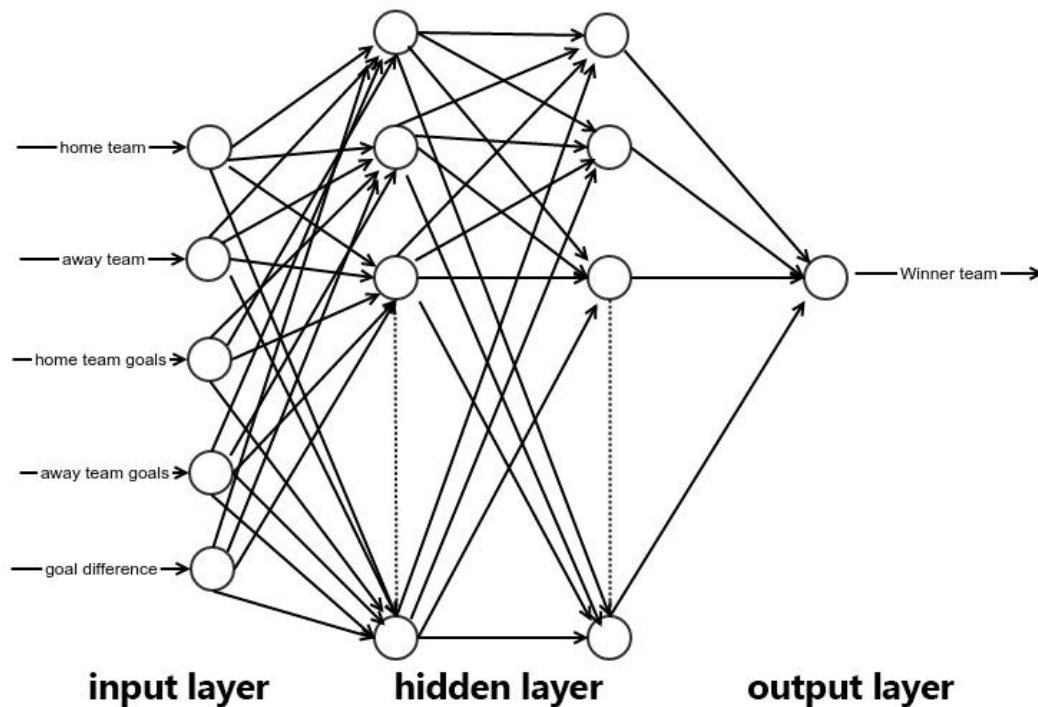


Figure 3.2: Neural Network Of Premier League Result Prediction

Premier League Result Prediction

Figure 3.2 shows that this application consists of five input parameters as home team, away team, home team goals, away team goals and goal difference with two hidden layers and the output layer contains three nodes. The result predicted by this network is listed in the appendices 1, 2, 3, and 4 respectively.

3.1.5 Algorithm

Back propagation algorithm was used for training the neural network.

Steps for back propagation:

- a. Initialization of the weight to random values.
- b. Feed the training sample through the network and determine the final output.
- c. Computation of the error for each output unit

for unit k, it is

$$\delta_k = (t_k - y_k) * f'(y_{ink}) = (t_k - y_k) * f(y_{ink}) * [1 - (f(y_{ink}))]$$

- d. Calculation of the weight correction term for each output unit

for unit k, it is

$$\Delta\theta_{jk} = \alpha \delta_k Z_j$$

- e. Propagate the delta terms (errors) back through the weights of hidden units where the delta input for jth hidden unit is

$$\delta_j = (\delta_{ink}) * f'(Z_{ink}) = (\delta_{ink}) * f(Z_{ink}) * [1 - (f(Z_{ink}))]$$

- f. Calculate the weight correction term or parameters for the hidden units

$$\Delta\theta_{ij} = \alpha \delta_j x_i$$

- g. Update the weights or parameter

$$\theta_{jk} \text{ (new)} = \theta_{jk} \text{ (old)} + \Delta\theta_{jk}$$

- h. Test for the stoppage

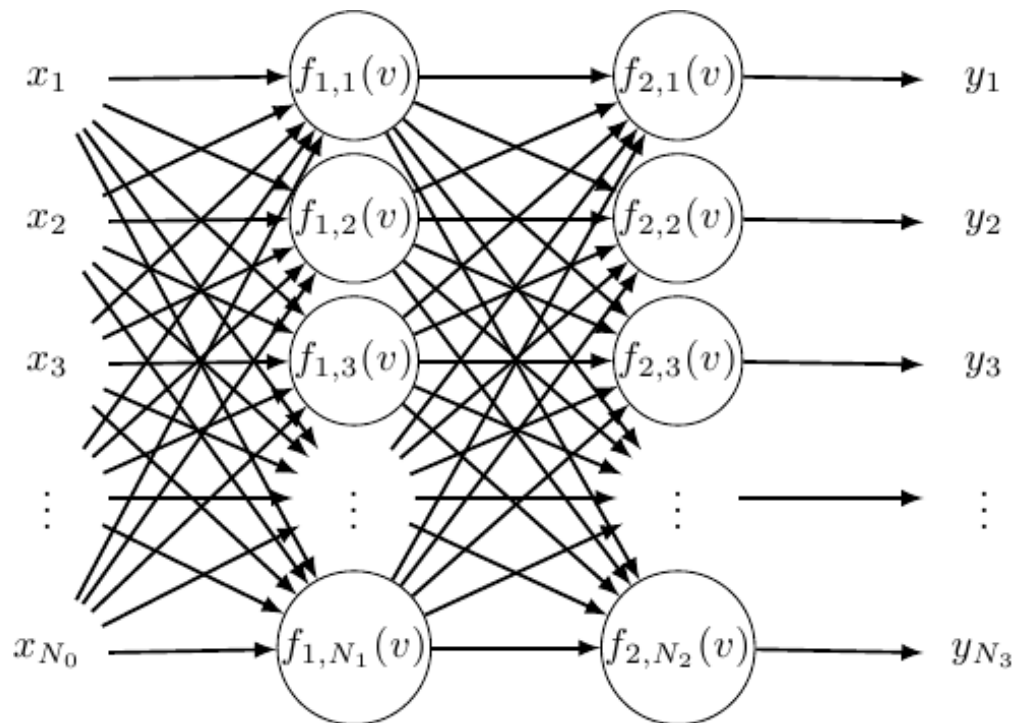


Figure 3.3: Back Propagation Algorithm

3.2 System Design

3.2.1 Sequence Diagram

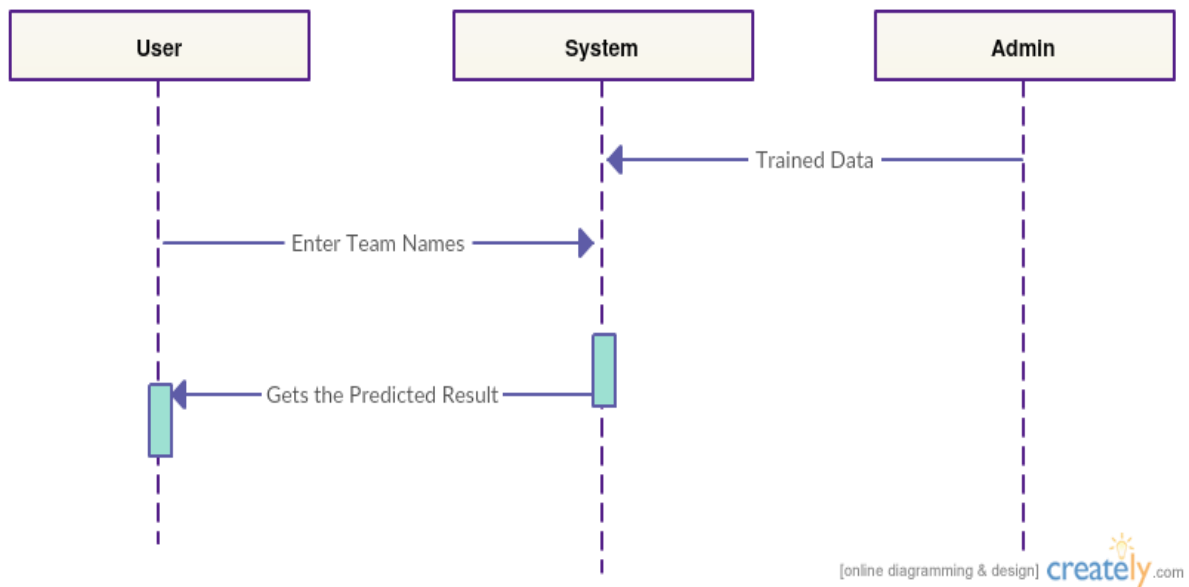


Figure 3.4 : Sequence Diagram of Premier League Game Result Prediction

Premier League Result Prediction

Figure 3.4 illustrates the sequence diagram of the application Premier League Game Result Prediction. It gives the sequence of the application from the user to the system and vice-versa. When the choose the team names from the available teams with trained set of data, it gets interpreted and the system returns the predicted result to the user. Further it is displayed on the application where admin gives trained and tested result.

3.2.2 Event Diagram

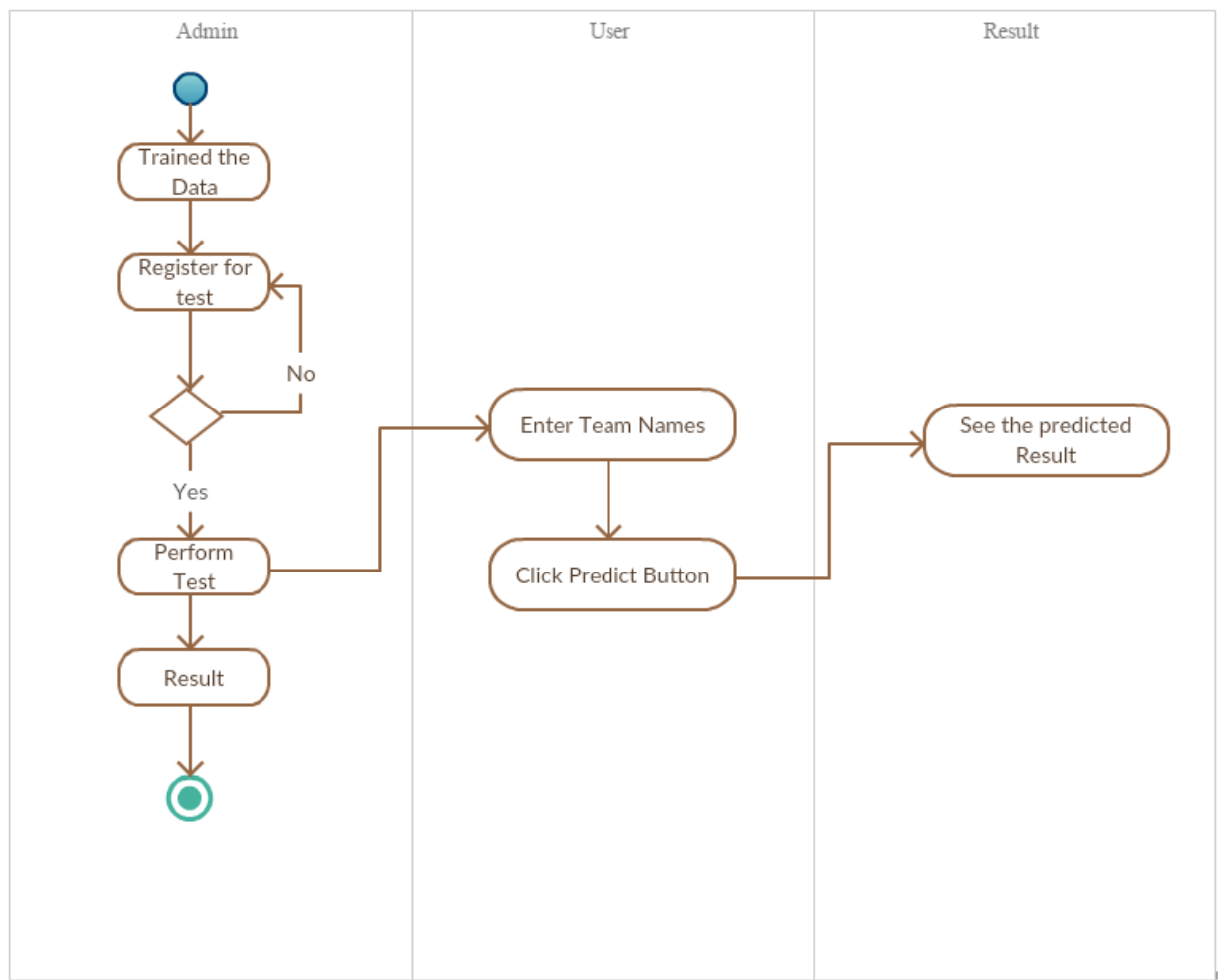


Figure 3.5 : Event Diagram Of Premier League Game Result Prediction

Figure 3.5 illustrates the event diagram of the Premier League Game Result Prediction which displays events and the process that are interrelated to each other. First, admin trained the data which is registered for the testing purpose and the test is performed. After the test is

Premier League Result Prediction

performed, a user-friendly User Interface is built and the user enters the team names into the input fields and clicks the predict button from which the result can be seen by the user.

3.2.3 Class Diagram

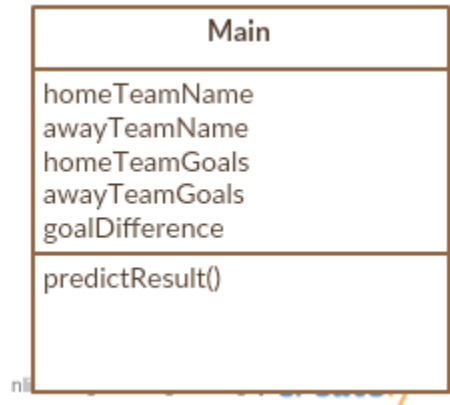


Figure 3.6 : Class Diagram Of Premier League Game Result Prediction

Figure 3.6 shows the class diagram of this application which consists of a Main Class only where a function named as predictResult() represents the prediction of the result. The parameters or variables includes Home Team Name, Away Team Name, Home Team Goals, Away Team Goals and Goal Differences.

Chapter 4: Implementation and Testing

4.1 Implementation

Premier League Game Result Prediction is the ability to predict the game result. The main purpose of implementing Premier League Game Result Prediction is to predict results of matches before the game. The past dataset needs to be trained and tested and with the trained dataset, we need to predict the result. In the implementation process, we have around 9000 data of which 80% are used for the training purpose whereas remaining 20% are used for the testing purpose.

Training is done by using Back Propagation algorithm in octave where the data are differentiated into Home Team, Away Team, Home Team Goals, Away Team Goals and Goal Differences in one file and the result is set in another file. The supervised learning is done so the test and train result output must be given before processing the application.

Table 4.1: Sample Trained Data Used For Implementation

HomeTeam	AwayTeam	HomeTeamGoals	AwayTeamGoals	GoalDifference
0.07	0.4	1	4	3
0.08	1.02	5	2	3
0.35	0.16	2	3	1
0.74	1.09	3	1	2
0.78	0.7	2	2	0
0.87	1.24	1	2	1
0.94	0.67	1	4	3
1.09	1.21	4	2	2
1.11	0.51	1	3	2
1.3	0.04	1	3	2

Table 4.1 shows the sample input data that are being trained. The data are kept as the input parameter for the training set.

Table 4.2: Trained Data Result Used For Implementation

Win	Draw	Lose
0	1	0
1	0	0
0	1	0
1	0	0
0	0	0
0	1	0
0	1	0
1	0	0
0	1	0
0	1	0

Table 4.2 shows the output that are been seen while data are trained. The data represents win, loss and draw. 010 represents draw, 100 represents win and 000 represents loss. testing is done as per the result obtained from the training sets in octave where the output pattern is already known by the application.

Table 4.3: Test Data Used For Implementation

Home Team	Away Team	Home Team Goals	Away Team Goals	Goal Difference
0.87	1.29	3	3	0
1.17	1.32	0	1	1
1.24	0.15	2	1	1
1.3	1.18	0	3	3
1.34	0.16	2	3	1
0.16	1.34	1	2	1
0.56	0.08	0	0	0
0.74	1.18	1	1	0
0.87	0.07	0	0	0
1.02	0.18	0	4	4
1.32	0.9	1	1	0

Table 4.3 shows the input data that are being tested. The data are kept are the input parameter for the testing purpose.

Table 4.4: Test Data Result Used For Implementation

Win	Draw	Lose
0	0	1
0	1	0
1	0	0
0	1	0
0	1	0
0	1	0
0	0	1
0	0	1
0	0	1
0	1	0

Table 4.4 shows the output that are been seen while testing the data. The data represents win, loss and draw. 010 represents draw, 100 represents win and 000 represents loss. Testing is done as per the result obtained from training sets in octave where the output pattern is already known by the application.

4.1.1 Tools Used

Different tools and technologies has been used in this project. Some of them are discussed below.

Octave

It is a high-level interpreted language, primarily intended for numerical computations. It provides capabilities for the numerical solution of linear and nonlinear problems, and for performing other numerical experiments. Basically, Octave is used to train and test the data sets which are useful in predicting the result. Octave uses different functions and some custom functions which sets up the input, hidden layers and generate output. The main reason for using octave is it makes easy to interpret the result and used for testing and training the dataset in the project.

Premier League Result Prediction

Creately

It is the diagramming and designing software which helps to design the various technical diagrams. It is a cloud-based diagram tool built on Adobe's Flex/Flash technologies and provides a visual communication platform for virtual teams. It can be used to create infographics, flowcharts, Gantt charts, organizational charts, website wireframes, UML designs, mind maps, circuit board designs, doodle art and many other diagram types. In our project, there are many technical diagram like UML class diagram, event diagram, sequence diagram and use case diagram which is built from creately.com which helps to make the easy diagram simply by drag and drop.

HTML/CSS/JQuery

HTML(Hyper Text Mark Up Language) and CSS(Cascading Style Sheet) is used to give the view of the application or websites. It is used to make the user interface for the user which makes them user friendly to use it. JQuery is used to hold the answer, process it and display the answer in this applicaton.

Java Servlet

It is used to create the web-based application in java using Java Server Page (JSP). In this project, the training set of data is kept as the theta values where the Java Servlet is used to interpret it and display the result of the prediction.

IDEA IntelliJ

It is a Java integrated development environment (IDE) for developing computer software. The project is written, compiled and run in this software which made easy to write code of the application.

4.1.2 Description / Listing of major classes / methods / any other modules

Main Class

It is the main class where the main program is included. This class consists of a method which interpret the values that are coming from the trained data and processing it to the user input and displaying the output which consists of predictResult() as the main function

predictResult() method predicts the result of the user input on which team will win the game with interpreting the values of theta with the help of matrix multiplication.

```
double [][] theta2Transpose = new double[11][3];

//Transpose of a matrix

for(int i=0;i<3;i++){
    for(int j=0;j<11;j++){
        theta2Transpose[j][i] = theta2DoubleArray[i][j];
    }
}

double [][] finalResult = new
double[firstResultAfterBias.length][theta2Transpose[0].length];

System.out.println(finalResult.length + "x" + finalResult[0].length);

for(int i=0;i<firstResultAfterBias.length;i++) {
    for(int j=0;j<theta2Transpose[0].length;j++){
        finalResult[i][j] = 0.0;
        for(int k=0;k<theta2Transpose[0].length;k++) {
            finalResult[i][j] += firstResultAfterBias[i][k] * theta2Transpose[k][j];
        }
    }
}
```

Premier League Result Prediction

```
}
```

```
double [][] secondModifiedresult = new  
double[firstResultAfterBias.length][theta2Transpose[0].length];  
  
for(int i = 0; i < finalResult.length; i++){  
    for(int j = 0; j < finalResult[0].length; j++){  
        secondModifiedresult[i][j]= (1 / (1 + Math.pow(Math.E, (-1 * finalResult[i][j]))));  
        System.out.println(secondModifiedresult[i][j]);  
    }  
}
```

The above code shows the interpretation of theta values that are obtained from the training set. It interprets the theta values doing matrix multiplication. The two dimensional matrix multiplication is used for analyzing the theta values and the result is being displayed. For the multiplication of matrix, the size of the multidimensional array are determined and if it doesn't matches the matrix as of the theta value then transpose of the matrix is used.

```
costFunction = @(p) nnCostFunction(p, input_layer_size, hidden_layer_size, num_labels, X,  
y, lambda);
```

The code calculates the cost function with given input layer, hidden layer, the output. y represents what will be the output look like and X is the training set of data.

4.2 Testing

4.2.1 Unit Testing

Each unit of the system was tested for its correct and proper functionality.

Table 4.5: Unit Testing of Premier League Game Result Prediction

Test no.	Unit	Test	Expected Result	Test Outcome	Evidence
1	Cost Function	Cost function minimizing test	Minimize the value of cost function in every iteration	Successful	Test 1.1
2	Error Check	Target value test	Output should be similar to target value of testing data set	Successful	Test 1.2
3	Predicted Value Check	Output value test	Output should be in pattern either 000, 100, or 010	Successful	Test 1.3

Premier League Result Prediction

Test Case 1:

Test 1.1: Cost function successfully minimized

Input: Values of attributes variables like Home Team, Away Team, Home Team Goals, Away Team Goals and Goal Differences are provided to the neural network as input variables.

Output: After a 20,000 iterations, cost functions was finally stabilized with a minimum value of 5.3247832.

Test 1.2: Output for testing dataset successfully implemented

Input: Values of attributes variables like Home Team, Away Team, Home Team Goals, Away Team Goals and Goal Differences from testing data set was provided as an input for validating the output of system.

Output: Output obtained with the target values of testing data set.

Test 1.3: Predicted Output Checked.

Input: Two teams were selected from the given team names by the user for the prediction of the game.

Output: Result obtained meet the criteria of the result as home team wins, away team wins or draw as shown to the user.

Chapter 5: Maintenance and Support

This application will be maintained over a period of time to provide the future needs. Some of the technique for maintenance are:

5.1 Adaptive Maintenance

This application needs to update the data and it needs to be trained within every year which makes the prediction more accurate.

5.2 Perfective Maintenance

The interface for user can be changed frequently and more user friendly design is developed. The game result can be predicted with the score line rather than win, lose or draw.

Chapter 6: Conclusion and Recommendation

6.1 Conclusion

This application estimates the prediction of the result of premier league game between two teams. Also, it predicts only the win, lose or draw of the game. The result is overall satisfactory with 47% of the accuracy of the prediction using Back Propagation algorithm.

6.2 Recommendation

This application consists of the less number of parameters which definitely cannot predict more accurate result. However, more number of attributes cannot improve accuracy of the prediction. Thus, selection of good attributes can only improve the accuracy of the prediction system. So, the best parameters still needs to be selected and hence it can improve the accuracy of predicting the game result.

References

AdityaSrinivasTimmaraju, AdityaPalnitkar, VikeshKhanna, “Game ON! Predicting English Premier League Match Outcomes”, Stanford University, 2014.

A. Joseph, N.E. Fenton, M. Neil, “Predicting football result using Bayesian nets and other machine learning techniques”, Computer Science Department, Queen Mary, University of London, UK, April 2006.

AlbinaYezus, “Predicting Outcome of Soccer Matches Using Machine Learning”, Saint-Petersburg State University Mathematic and Mechanics Faculty, 2014.

Brijesh Kumar Bhardwaj, Saurabh Pal, “Data Mining: A prediction for performance improvement using classification”, (IJCSIS) International Journal Of Computer Science and Information Security, Vol. 9, No. 4, April 2011.

DouweBuursma, “Predicting Sports Events From Past Results”, University of Twente, 7500AE Enschede The Netherlands, 2013.

FarzinOwramipur, ParinazEskandarain, and Faezeh Sadat Mozneb, “Football Result Prediction with Bayesian Network in Spanish League – Barcelona Team”, International Journal of Computer Theory and Engineering, Vol 5, No. 5, October 2015.

F. Halicioglu, “Can we predict the outcome of the international football tournaments: The case of EURO 2000,” Doğuş Üniversitesi Dergisi, vol. 6, no. 1, 2005.

Francisco Louzadal, Adriano K.Suzuki and Luis E.B. Salasar, “Predicting Match Outcomes in the English Premier League: Which Will Be the Final Rank?”, Journal Of Data Science, December 2014.

GianlucaBaio, Marta A. Blangiardo, “Bayesian hierarchical model for the prediction of football result”, Dec 2010.

Premier League Result Prediction

Havard Rue and Oyvind Salvesen, "Prediction and Retrospective Analysis Of Soccer Matches In a League", Department of Mathematical Sciences, NTNU, Norway, June 1999.

Igiri, Chinwe Peace, Nwachukwu, EnochOkechukwu, "An Improved Prediction System for Football a Match Result",IOSR Journal Of Engineering (IOSRJEN), Vol 4, Issue 12, December 2014.

Jeffrey Alan Logan Snyder, "What Actually Wins Soccer Matches: Prediction of the 2011-2012 Premier League for Fun and Profit", May 6, 2013.

Reilly, Thomas; Gilbourne, "Science and football: a review of applied research in football code", Journal of Sports Science, 2003.

YueWengMak, "Prediction on Soccer Matches using Multi-Layer Perceptron", ID = 903-051-7735, 2013.

Appendix

Appendix - 1: Home Page

The screenshot displays the home page of a Premier League game result prediction website. The page is divided into three main sections:

- Table of Contents:** A table listing the current Premier League standings, including position, club, points per game (PPG), goal difference (GD), and total points (Pts).
- Prediction Form:** A central section titled "Premier League Game Result Prediction" where users can select a home team and an away team from dropdown menus and click "See the Prediction".
- Matchweek 4 Fixtures:** A list of upcoming games for Matchweek 4, showing the home and away teams with their respective logos.

The background features a large, stylized logo of a lion standing on a shield, with the text "BARCLAYS PREMIER LEAGUE" below it.

Pos	Club	PPG	GD	Pts
1	Man City	3	+6	9
2	Chelsea	3	+5	9
3	Man Utd	3	+5	9
4	Everton	3	+2	7
5	Hull	3	+2	6
6	Middlesbrough	3	+1	5
7	Spurs	3	+1	5
8	Arsenal	3	+1	4
9	Leicester	3	0	4
10	West Brom	3	0	4
11	Liverpool	3	-1	4
12	West Ham	3	-2	3
13	Burnley	3	-2	3
14	Swansea	3	-2	3
15	Southampton	3	-2	2
16	Sunderland	3	-2	1
17	Crystal Palace	3	-2	1
18	Watford	3	-3	1
19	Bournemouth	3	-3	1
20	Stoke	3	-4	1

Premier League Game Result Prediction

Premier League

Home Team: Manchester United

Away Team: Accrington Stanley

See the Prediction

Matchweek 4 Premier League

Home Team	Away Team
MUN	MCI
ARS	SOU
BOU	WBA
BUR	HUL
MID	CRY
STK	TOT
WHU	WAT
LIV	LEI
SWA	CHE
SUN	EVE

This screenshot shows the home page or the landing page which consists of the current premier league table standings of the club. In the middle section, the teams are being selected through the dropdown. The right side consists of the upcoming game fixtures in the premier league.

Premier League Result Prediction

Appendix - 2: Selecting the teams

Premier League Game Result Prediction

Home Team: Manchester United

Away Team: Chelsea

See the Prediction

Matchweek 4 Premier League

Pos	Club	Pt	GD	Pts
1	Man City	3	+6	9
2	Chelsea	3	+5	9
3	Man Utd	3	+5	9
4	Everton	3	+2	7
5	Hull	3	+2	6
6	Middlesbrough	3	+1	5
7	Spurs	3	+1	5
8	Arsenal	3	+1	4
9	Leicester	3	0	4
10	West Brom	3	0	4
11	Liverpool	3	-1	4
12	West Ham	3	-2	3
13	Burnley	3	-2	3
14	Swansea	3	-2	3
15	Southampton	3	-2	2
16	Sunderland	3	-2	1
17	Crystal Palace	3	-2	1
18	Watford	3	-3	1
19	Bournemouth	3	-3	1
20	Stoke	3	-4	1

Matchweek 4 Premier League

Club	Club
MUN	MCI
ARS	SOU
BOU	WBA
BUR	HUL
MID	CRY
STK	TOT
WHU	WAT
LIV	LEI
SWA	CHE
SUN	EVE

This screenshot shows the selection procedure of the teams to predict their results.

Premier League Result Prediction

Appendix - 3: Result

The screenshot displays a web application for Premier League result prediction. A modal window titled "Result" is centered on the screen, showing the word "Draw" in green. Below the result, there are two links: "View Adjusted Weight 1" and "View Adjusted Weight 2", and a "Close" button. In the background, a table lists the Premier League teams and their statistics. A "See the Prediction" button is visible below the modal. On the right, a "Matchweek 4" fixture list is shown. The background also features the Barclays Premier League logo.

Pos	Club	Pl	GD	Pts
1	Man City	3	+6	9
2	Chelsea	3	+5	9
3	Man Utd	3	+5	9
4	Everton	3	+2	7
5	Hull	3	+2	6
6	Middlesbrough	3	+1	5
7	Spurs	3	+1	5
8	Arsenal	3	+1	4
9	Leicester	3	0	4
10	West Brom	3	0	4
11	Liverpool	3	-1	4
12	West Ham	3	-2	3
13	Burnley	3	-2	3
14	Swansea	3	-2	3
15	Southampton	3	-2	2
16	Sunderland	3	-2	1
17	Crystal Palace	3	-2	1
18	Watford	3	-3	1
19	Bournemouth	3	-3	1

Matchweek 4
Premier League

MUN MCI
ARS SOU
BOU WBA
BUR HUL
MID CRY
STK TOT
WHU WAT
LIV LEI
SWA CHE
SUN EVE

Waiting for localhost...

This screenshot shows the result of the prediction of the teams