IOSR Journal of Engineering (IOSRJEN) www.iosrjen.org
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 04, Issue 12 (December 2014), ||V4|| PP 12-20

An Improved Prediction System for Football a Match Result

Igiri, Chinwe Peace[1]; Nwachukwu, Enoch Okechukwu[2]
chynkemdirim@gmail.com

enoch.nwachukwu@uniport.edu.ng
[1, 2] *(Department of Computer Science, University of Port Harcourt, Nigeria.)*

***ABSTRACT*:** *Predictive systems have been employed to predict events and results in virtually all walks of life. Football results prediction in particular has gained popularity in recent years. Statistical approaches have shown complex and low prediction results. Data mining tools with insufficient features, however, have also yielded low predictions. In our research, knowledge discovery in databases (KDD) has been used to develop a football match result predictive model by gathering 9 features that affect the outcome of football matches. We constructed a more comprehensive system with an improved prediction accuracy by using the features that directly affect the result of a football match. Our prediction system for football match results was implemented using both artificial neural network (ANN) and logistic regression (LR) techniques with Rapid Miner as a data mining tool. The technique yielded 85% and 93% prediction accuracy for ANN and LR techniques respectively. With this output, it is observed that the prediction accuracy is higher than those of existing systems.*

***KEYWORDS*:** *ANN, data mining, KDD, models, prediction*

I. INTRODUCTION

Today, so much information is available that the challenge has become deciphering what is relevant. The Internet is updated frequently with information. Information about the stock market, databases of student results, telecommunication information, and weather records are a click away. The entertainment industry is no exception; for instance, football statistics flood the Internet every now and then. The English Premier League, in particular, produces a great deal of data because it is so popular. Fixed odds betting markets and researchers make use of these data to analyse and predict football match results. Different statistical techniques have been used to develop models for football match result prediction. Although some of these predictions have reasonable levels of accuracy, limitations remain, and including the fact that some features affecting matches are not considered due to their complexity. In spite of several studies in this area, there is room for improvement in terms of developing a system with higher prediction accuracy without much complexity. This research attempts to contribute to the existing literature by developing and implementing an improved model using the Artificial Neural Network (ANN) and Logistic Regression classifier through Knowledge Discovery in Database (KDD). A data mining tool will be used to implement the model as well as to evaluate and predict the result of a football match. The improvement is that efforts will be made to overcome some of the limitations in the previous work through Knowledge Discovery in Database. Mining football data with the aid of data mining software, such as Rapid Miner, creates room to include as many features as possible due to the scalability of the tool.

II. RELATED WORK ON RESULT PREDICTION SYSTEMS

[1], developed a system that can predict students' performance based on their past performances by employing classification in data mining. Their analysis was carried out on a data set of student information, such as gender and marks scored at various levels of examination. They applied an ID3 (Iterative Dichotomiser 3) and C4.5 classification algorithm on these data sets to predict the general and individual performances of freshly-admitted students on future examinations. Their prediction was 75.145%. However, this work is limited by the fact that its implementation was not dynamic in that the prediction parameters could not train the new dataset when fed into the web application.

[2], used a training set from an eleven-year period to train World Chess Federation rating systems with 2000 chess players as the data set. They applied the Hidden Markov Process Model and fitted it with Newton Raphson's method. The success of the prediction was 55.64%. However, according to the authors, a more accurate prediction would have been obtained by following another model [2].

[3], developed a system with the intention to "beat" bookmakers' odds on football. In his work, he attempted to ascertain the important features in predicting football match and to calculate the probability of the proposed features in order to identify bets to maximize profit. He employed seven machine learning algorithms to classify the matches into home win, draws or away wins: MultiClassClassifier, RotationForest,

LogitbooST, BayesNet, Naive Bayes and Home Wins. The accuracy of his algorithm was 55%. He acknowledged that, in the long run, his system did not meet the goal of "beating" the bookmakers. The limitation in this research is that the prediction accuracy was relatively low. The author later suggested an improved system that would include features, such as all bookings during the match, the players composing each team, their managers and more.

[4], employed the Bayesian Network Model to predict the results of football matches involving the Barcelona team in the 2008-2009 Spanish League. They divided the data set for the project into two: (a) non-physiological factors (weather, history of five previous matches, results against/for team, home game and players' psychological state; and (b) physiological factors (average age of the players, the number of injured main players, average match in a week, performance of main players, the performance of all players and average number of goals for all home and away matches). NETICA software was used to build the model, which yielded values for average age of the players as a medium, history of the last five games as win, injured main players, psychological state of the of the players and weather conditions during the match. A prediction accuracy of 92% was obtained when used to predict the 2008-2009 season matches. The limitation in this research is that only one team was considered.

[5], proposed a Bayesian hierarchical model to predict football results. The data set was based on scoring intensity determined by the attack and defence strength of the two teams involved. The team, playing from home or at away matches, also was used to determine the goals scored for each season. They applied an MCMC-based procedure to estimate the value of the main effect, which was used to explain the scoring rate. Although their predictions was 95% accurate, their work only highlighted the teams with the highest propensity to score or concede goals, a major limitation in the research. They also attempted to reduce the challenge of over-shrinking caused by Bayesian hierarchical model by introducing a mixture model, thereby making the model more complex and time consuming.

[6], developed four football result prediction models: (1) toto-models (random probability and team grouping); (2) multi-independent score model (3) single-independent model; (3) dependent score model; (4) pseudo least-square estimator score model. For the English Premier League of 2007-2008 through 2010-2011, they applied each of these models on the number of goals for both teams playing and the number of goals scored by the home and away teams, forming their data set.

Summary of Related Literature on Result Prediction Systems: The summary of the work related to this study is given in Table 2.1. Four literatures related to result prediction system have been critically reviewed and summarized in a table for the purpose of visualization and interpretation.

Table 2.1: Summary of Related Works for Result Prediction Systems

Authors	Research Title	Data Set Used	Techniques Applied	Prediction Accuracy	Limitations
[1]	Predicting student performance using ID3 and C4.5 classification algorithms	Student gender, Marks scored at various levels	ID3 and C4.5 Classification	75.2%	The system is not dynamic.
[2]	Chess game result prediction system	2000 chess players	Hidden Markov Process Model & Newton Raphson's method	55.6%	The model did not perfectly fit the system.
[3]	Predicting sports events from past results: Towards effective betting on football matches	Home win, draw and away win results	Multiclassifier, Rotationforest, LogitbooST, BayesNet	55%	Insufficient features that affect the outcome of a football match were used for the system.
[4]	Football result prediction with Bayesian Network in Spanish league-Barcelona team	Psychological and non-psychological features	NETICA software – a data mining tool	92%	The system can only predict results for one team.

Factors that affect the result of a football match: Several factors affect the result of football match. These factors include; Home advantage, the effect of injuries on key players and the effect of external cup on the league.

Home advantage effect on team's performance

Home advantage is "*the consistent finding that the home team wins over 50% of the games played under balanced home and away schedule"* [7]. Categorically, travel and crowd factors could also account for the home advantage effect on team's performance.

Experimental study by [8] reported the evidence of home team favouritism in the English Premier League and German Bundesliga. More so, [9] and [10] drew the same conclusion based on their respective studies: that it makes sense to model home and away performance separately.

The effect of injuries of key player on team performance

[11], reviewed the effect of injuries of individual players on the performance of the entire team. In their analysis, they measured the team quality as a sum of players' quality scores from a sample of five seasons of English Premier League. This research showed that the injuries of key players plays significant role on team's performance. However, [12] argued that such information is subjective in nature and should not be considered in statistical forecasting.

Effect of external cup on league performance

The explanation to the effect of external cup can be described by the financial incentives cups offer. [13], propounded that teams are more interested to perform well in league with direct revenue than that without direct financial incentives. [9], claimed that cup participation would imply an increase in league performance, consequently a cup exit would imply decrease in league performance.

Concluding remarks on relevant factors to football result prediction : In this literature, different factors that have shown to have effect on match results: home advantage, effect of injuries on key players, and external cup effect. Information about historical league matches and public sources are relevant in football result prediction.

III. ANALYSIS AND DESIGN

Analysis of Existing systems: We analyse two approaches used in football result prediction in existing systems, these include; statistical and machine learning approaches. For statistical approach, we analyse Hidden Markov Process Model and Ordered Probit Regression model. Also, a detailed analysis of machine learning approach by [3] has been done. Football predictive system is made up of two main components, namely: feature sets/ data sets and implementation techniques. We therefore analyse the data sets and the techniques used in the implementation of existing system.

Football Feature Sets: The type of feature sets used to build a model determine how accurate its predictions would be. If the data sets that directly affect the match results are used according to the appropriate technique, there is no doubt that the prediction accuracy will be high. Otherwise, the reverse is true. The following feature sets have been used to develop and implement football models in existing systems: number of goals scored in home and away matches, number of goals conceded in home and away matches, attack and defence strength, players' injuries, weather conditions, etc. [2] used win, loss, and draw values of the chess game match result as feature sets while [9] used the same data sets for football match result prediction.

Analysis of implementation approach of existing system: Analysis of implementation approach in existing system is a method or technique used in the implementation of the existing system. The methods can be broken down into the following:

Statistical techniques
Machine learning techniques.

Statistical Model for Football Predictive system: Sports betting uses statistical models for prediction by means of statistical tools. The hidden Markov model and the Ordered Probit Regression model are examples of statistical models that have been used to develop predictive systems.

Hidden Markov Process Model: [2], modelled a chess match results prediction system. In the model, each chess player i has a rating in month t, denoted as $X_{i,t}$. The chess match result, $Y_{t,j1,j2}$, was observed and denoted by players $_{j1}$and $_{j2}$ in month t. $Y_{t,j1,j2}$ could take 3 values for win, draw and loss. Equations (1), (2), and (3) describe the relationship.

$$X_{i,t} = \omega_0 X_t + \sum_{1=k}^{k} \omega X_{i,t-l} + \epsilon_{i,t} \epsilon_{i,t} \backsim N(0, \sigma^2_{i,t}) \quad (1)$$

$$Y_{t,j1,j2} \alpha \ \phi_{t,j1,j2}^{Y_{t,j1,j2}} (1- \phi_{t,j1,j2}) \quad ^{(1-Y_{t,j1,j2})} \quad (2)$$

$$\log \frac{\phi_{t,j1,j2}}{1-\phi_{t,j1,j2}} = X_{j1,t} - X_{j2,t} + F \quad (3)$$

In order to model the winning probability (3), a hidden Markov process model was used to implement a dynamic rating in (1) and a Bernoulli-like distribution in (2). A relatively low prediction accuracy of 56% was reported when fitted with the Newton Raphson method.

Ordered Probit Regression Model for Match Results: An ordered probit regression was used to model and predict football match results. The result of the match between teams i and j, denoted $y_{i,j}$, depends on the unobserved variable $y_{i,j}$ and a normal independent and identically distributed disturbance term, ε_{ij} as follows:

Home win: $y_{i,j} = 1$ if $\mu_2 < y_{i,j} + \varepsilon_{ij}$

Draw: $y_{i,j} = 0.5$ if $\mu_1 < y_{i,j} + \varepsilon_{ij} < \mu_2$

Away win: $y_{i,j} = 0$ if $y_{i,j} + \varepsilon_{ij} < \mu_1$ (1)

$y_{i,j}$depends on the following systematic influences on the result of the match between teams i and j:$p^{a}_{i,j,s}$=Home team i's average win ratio (1=win, 0.5=draw, 0=loss) from matches played 0-12 months (y=0) or 12-24 months (y=1) before the current match; within the current season(s=0) or the previous season (s=1) or two previous seasons (s=2); in the team's current division (d=0) or one (d=$\pm$1) or two (d=$\pm$2) divisions above or below the current division.

$R^{H}_{i,m}$= Result (1=win, 0.5= draw, 0 =loss) of m'th most recent home match played by home team i (m=1…M).

$R^{A}_{i,n}$= Result of n'th most recent away match played by home team i (n=1…N) [9].Although, they claimed that the probit regression model was easy to implement, a low prediction accuracy remains.

Data Mining (Machine Learning) Technologies for Predictive System: Data mining does not replace statistics. In fact, statistics are a good complement to data mining. Traditional statistical techniques, such as regression, should therefore be used alongside data mining technologies, such as neural networks, LogitBoost classifiers, Bayesian network, support vector machine etc. Statistics are also necessary to validate data mining results. Ananalysis of the classifiers (LogitBoost, Bayesian network, ClassificationViaRegression, etc.) and the Bayesian hierarchal model will be conducted for the purpose of completing this research.

Buursma's Classifiers Model: [3], used six machine learning algorithms to build a model that selected the best features (home wins, draw or away wins) of the classifier algorithms: ClassificatioViaRegression, MultiClassClassifier, RotationForest, LogistBoost,BayesNet, and NaivBayes. Fig. 3.1 illustrates the schematic representation of Buursma's classifiers model.

Advantages and Disadvantages of Existing Systems: A number of techniques and feature sets have been used to develop and implement football results prediction systems. Due to the variations observed in these systems, it is difficult to identify which system outperforms the others. These variations include: number of goals prediction, win-draw-loss prediction, propensity to score or concede goals, total points earned in a season, etc.

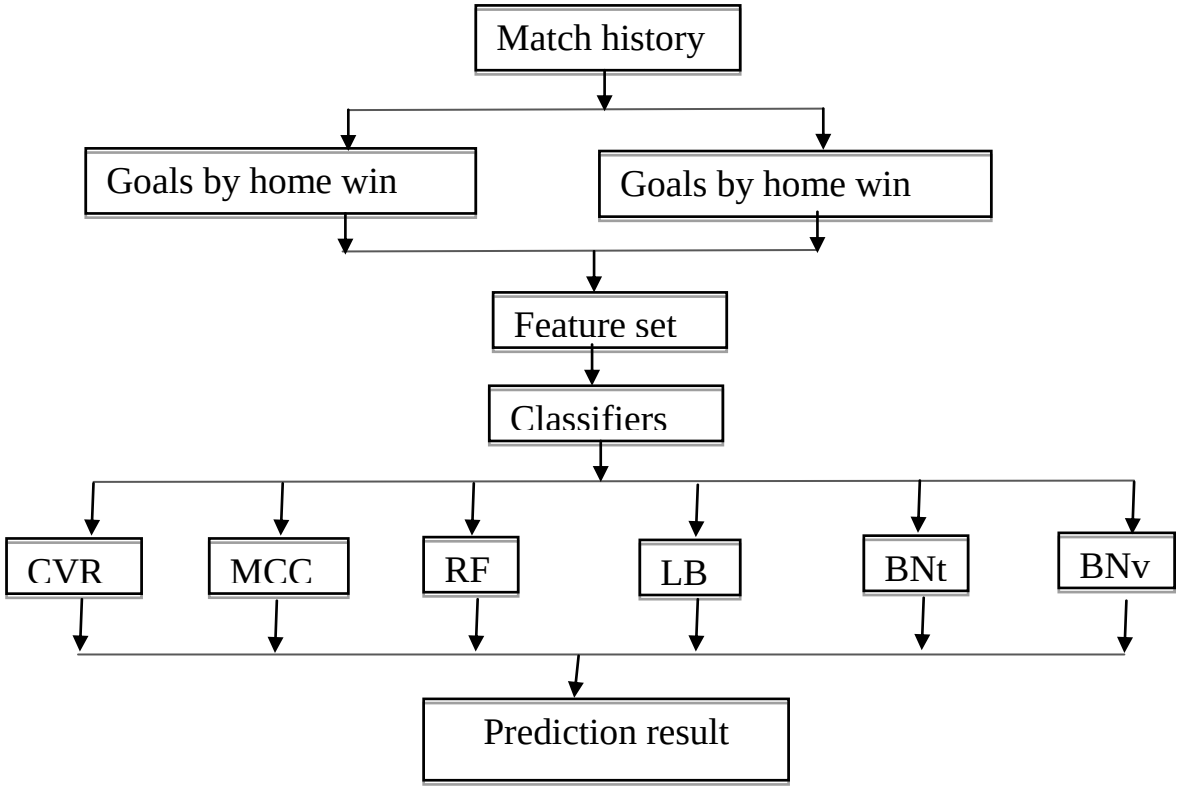

Figure 3.1 Buursma's classifiers model (Existing system)

Table 3.1 Summary of advantages and disadvantages of existing systems

Approach	Advantages	Disadvantages
Statistics	1. Low cost implementation, 2. Data availability, 3. Does not involve the rigors of data selection and cleansing.	1. Complex and could result in some errors, 2. Difficult to identify errors 3. Low prediction 4. Time consuming
Data mining	1. Free and open source data mining tools 2. Low error risk 3. Data availability 4. Enhanced commercial data mining tools	1. Limited capabilities 2. Tedious data cleansing task 3. Cost intensive

Analysis of the Proposed System: The framework for our proposed system is knowledge discovery in databases (KDD). The model will be developed using a data mining tool with enhanced capabilities to extract players' and managers' indices, which seeks to improve on the problems of implementation complexities and low prediction rate inherent in exiting systems. Our proposed system would be implemented using two different data mining techniques: Artificial Neural Network (ANN) and Logistic Regression techniques. Number of goals score, moving average of teams 'performance within a season, players' and managers' performance indices as well as other features used in the existing systems will be used to design and build the model. The data set will be extracted from 110 matches played in the 2014-2015 English Premier League season with the aid of the commercial data mining tool, Rapid Miner Studio. The same means will also be used to cleanse and integrate the data set. As shown in Fig. 3.2, data will be collected from three different sources, namely the match history record, the performance index record, and general football database. Excel

spreadsheets will be used to select and organise the data. Data cleansing and consolidation will be conducted using the data mining tool. Data quality assessment will also be carried out with the aid of Rapid Miner. The prepared data will be used to build a predictive model for football match results. Fig. 3.2 is an illustration of the model for our proposed system. The model consists of five major building blocks:

Building Block1: Football match results are collated and gathered from three major source these include; Match history, Performance index, and match day record for feature sets preparation.

Building Block2:Football feature sets: Data concerning the match history record, performance index record and football spreadsheet was extracted from different data sources in the internet.Such features sets as; Home and Away goals (GHA), Home and Away shorts (HAS), Home and Away corner (HAC), Home and Away Odds (HAOD), Home and Away attack strength (HAAT), Home and Away Players' performance index (HAPPI), Home and Away Managers' performance index (HAMPI), Home and Away managers' win (HAMW), Home and Away streak (HASTK).A data mining database will be built by collating all the data gathered from these different sources in an MS Excel spreadsheet.

Building Block3: Data Cleansing and Consolidation: This is another key data pre-processing feature introduced in our model in other to ensure improved efficacy. This module cleanses and consolidates data to build a data mining database as shown in Fig. 3.2 by using the RapidMiner data mining data cleansing tool. A precise impute missing value data cleansing operator will be used to execute this operation in our proposed system. It is a nested operator that always takes in data sets and returns a model. This operator calculatedly guesses the value of missing values by learning models for each attribute (excluding the label) and applying those models to the data sets. The Artificial Neural Network (ANN) technique, which is the learner in this case, will be used for estimating missing values and will be placed in the sub-process of the impute missing value operator. The cleansed data is further consolidated and integrated into the data mining database.

Building Block4: Weighting/Parameter Optimisation: The weighting and parameter optimisation module, as shown in Fig. 3.2, will be used adjust the training algorithms in order to produce a better learner RapidMiner Optimise Weights (Evolutionary) operator. This operator calculates the weights of the features of our football data sets by using a genetic algorithm (GA). The higher the weight of an attribute, the more relevant it is considered. According to [14], a GA is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to breed useful solutions to optimisation and search problems. Genetic algorithms are an aspect of a larger class of evolutionary algorithms (EA), which produce solutions to optimisation problems using such techniques as inheritance, mutation, and selection, as well as crossover [14].

Parameter optimisation, another data mining tool that finds the optimal values for a set of parameters using an evolutionary approach, actually yields better results. This tool will be used to adjust the learning rate, momentum and other hidden expert parameters until an improved model is built.

Building Block5: Model Building using ANN/LR: This module, as shown in Fig. 3.2, will be used to build two improved models for the proposed system using the ANN and logistic regression techniques respectively.

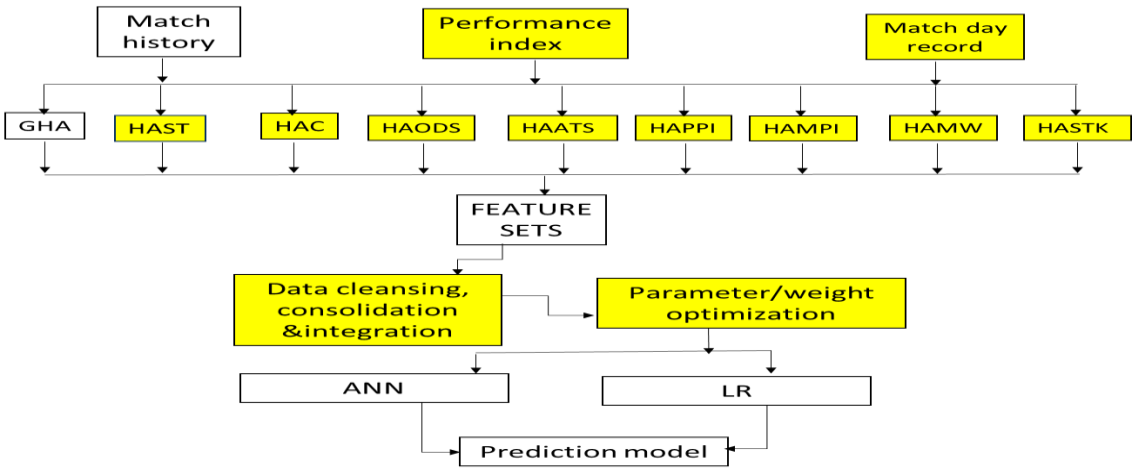

Figure 3.2: An improved model for football results prediction (The proposed model)

Building Block6: Predictive Model: This module, as shown in Fig. 3.2, describes our novel predictive model for football match results.

IV. RESULT AND FINDINGS

Comparison of the New System and Existing Systems: Since our third research objective is to compare the new system with existing systems, we will compare [6] and [3] approaches. The criteria for the comparison are: the data set, approaches/techniques, and prediction accuracy. Table 4.3 shows the comparison of the new system and existing systems.

Results Analysis and Findings: From our results, we could see that a higher prediction accuracy is obtained when nine features were optimised by weighting. In fact, without optimisation by weighting, the prediction accuracy is 75.04%, while optimisation increased the accuracy to 85% for current season (2014-2015) data. Fig. 4.1 shows the actual and predicted result for 20 matches played in 10th and 11th week of 2014/2015 English premier league football season. Further, the system could predict win, draw or loss results using ANN techniques. The prediction accuracy is almost two times higher than that achieved by [3]. Additionally, from our findings, it was discovered that logistic regression techniques could only predict win or lost results. Further, the system could predict win, draw or loss results using ANN techniques. The prediction accuracy is almost two times higher than that achieved by [3].

Row No.	WLD	prediction(...	confidence(...	confidence(...	confidence(...
1	WIN	WIN	0.999	0.001	0.000
2	WIN	WIN	1.000	0.000	0.000
3	DRAW	LOSS	0.000	0.124	0.876
4	LOSS	LOSS	0.000	0.015	0.985
5	LOSS	LOSS	0.000	0.008	0.992
6	WIN	WIN	0.998	0.000	0.002
7	DRAW	DRAW	0.000	0.549	0.451
8	LOSS	LOSS	0.000	0.304	0.696
9	WIN	WIN	1.000	0.000	0.000
10	LOSS	LOSS	0.000	0.370	0.630
11	WIN	WIN	0.961	0.000	0.039
12	LOSS	LOSS	0.000	0.029	0.971
13	WIN	WIN	0.953	0.000	0.047
14	DRAW	LOSS	0.000	0.133	0.866
15	WIN	WIN	1.000	0.000	0.000
16	DRAW	DRAW	0.000	0.981	0.019
17	DRAW	LOSS	0.000	0.034	0.966
18	WIN	WIN	0.998	0.000	0.002

Arsenal	Burnley
Chelsea	QPR
Everton	Swansea
Hull	Southampton
Leicester	West Brom
Newcastle	Liverpool
Stoke	West Ham
Aston Villa	Totten ham
Man City	Man United
Crystal Palace	Sunderland
Burnley	Hull
Liverpool	Chelsea
Man United	Crystal Palace
QPR	Man City
Southampton	Leicester
West Ham	Aston Villa
Sunderland	Everton
Swansea	Arsenal

Figure 4. 1: Screen shot of actual/ predicted result for 20 matches played in 10th and 11th week of 2014/2015 English premier league football season.

Table 4.3 Comparison of the new system with existing systems.

Criteria	Existing system	New system
Approaches	1.Statistical approach, 2. Machine learning	Knowledge Discovery in Data base
Data set used	Number of goals for both teams,	1. PPI, 2. Manager Index, 3. Attack strength 4. Defence strength, 5. Corner kicks, 6. shots, 7. shot on target, 8. Betting Odds, 9. Home and away win
Prediction accuracy	53.55%, 55%	75.8% 85% 93%

V. SUMMARY, CONCLUSION AND RECOMMENDATIONS

Summary: This thesis employed a data mining tool (Rapid Miner) to predict the results of football matches. Specifically, football data was extracted from English Premier League matches, which serve as the population of the football feature sets. When the ANN technique and logistic regression were used to build the model, it yielded 75.04% accuracy, 85% and 93% prediction accuracy when ANN and LR techniques were applied respectively. The ANN technique could predict win, lose or draw, while logistic regression could only predict a win or loss result. In any case, if a match draw must be predicted, the ANN technique is a better option compared to the logistic regression technique. Analysis was carried out to ensure data quality by using the data cleansing tool provided by Rapid Miner, which consequently improved the prediction accuracy. Optimisation plays a key role in a predictive system. In our research for instance, a prediction accuracy of 75.04% was recorded without appropriate optimisation. But the accuracy increased to 85% with ANN technique while Logistic Regression technique yielded 93% prediction. The optimisation by weighting also showed that the Player Performance Index (PPI) has a higher weighting compared to the other features when implemented with the ANN technique.

Conclusion: The improved football result Prediction System explores the use of machine learning techniques in the framework of Knowledge Discovery in Database. This research is driven by the overwhelming increase in the pool of available sports data in English premier league. The datasets collected was successfully implemented using data mining technique in different aspects of the work. In many instances, predicting the results of sporting procedures has always been a challenging and rewarding venture, therefore forecasting problem provides a growing need to conduct research in this area. Sports outcomes predictive techniques arise and this motivates the need to find more valuable datasets to improve the prediction accuracy and make precise decisions at key. Past comprehensive statistical data has been kept to assist English premier league games and other sporting events. Both players and teams' presents varying forms of these statistical facts kept as data season in and off-season. As the dataset set grows with the EPL games, it has become the preferred test platform. This pool of information will keep motivating different groups, ranging from public, statisticians and sports enthusiasts to discover embedded knowledge in it. This project also provide an ideal data mining environment, data mart containing comprehensive EPL game information, which can be reused by future research.

Finally, this project can be considered as a successful exploration of using Knowledge Discovery in Database for sports result prediction and it provides a good mainstay for future research works.

Recommendations: Prediction systems, particularly football match prediction, are growing and offer a wide range of technology for implementation. The advent of data mining in the framework of knowledge discovery in databases opens up new opportunities for development. Due to the high level of demand for football prediction systems, organisations are well positioned to take advantage of all that our improved prediction system offers. We can provide a few recommendations for ensuring the maximum utility of our prediction system for football match results:

A. Since the data necessary for the football match features is spread throughout different websites, we suggest that software able to collate all the features and organise them in Excel or CV files be developed.
B. Automated maintenance of the database of the prediction system might be necessary to ensure the system is constantly updated.
C. One of the major challenges in this project was the ability to understand the features that affect football match results. Further research could be carried out in this area using specific literature to assist novice and young researchers to develop a proper sense of direction.
D. Research could also be carried out on other data mining techniques to analyse their strengths and weaknesses in football result prediction.

Contribution to knowledge: The resultant findings can be accentuated as contributions to the knowledge in the area of sports and predictive analysis. This contribution to knowledge can be summarized as follows:

A. Developing an improved model for a football match result prediction that resulted to higher prediction (85%) than existing systems (55%).

B. Identifying new feature sets (players' performance index, managers' index, and book maker odds) that affect the result of a football match and consequently the prediction which have not been used in previous system.

VI. ACKNOWLEDGEMENTS

I would like to first and foremost offer my appreciation to Almighty God for the wisdom and understanding He granted everyone that contributed to the success of this work. My sincere gratitude goes to my supervisor, Prof. E.O. Nwachukwu, for his painstaking, thorough and detailed supervision during the course of this project. I must also thank Prof. H.C. Inyiama for introducing me to the data mining and modelling ideas that motivated this research. My gratitude also goes to my uncle, Prof. Lawrence Nkemdirim, for his immense contribution to this work. Thanks to all my MSc lecturers (whose contributions to my knowledge have led to the success of this research work): Dr. P.O. Asagba, Dr. B.O. Eke, Dr.Oghnenevo, Dr. William, Dr. (Mrs) V. Ejiofor, Mr. WobidiEchebiri, my hardworking, H.O.D, Dr. C. Ugwu, and our P.G. coordinator, Dr. (Mrs). Onyejiegbu. My special appreciation also goes to the family of Pst. Dele Adesida, Pst. Vincent and the entire family of Deeper Life Bible church for their motivation, enthusiasm, prayers and support during the course of this programme. I must not forget to specially thank my family; my beloved husband Mr. Chima Igiri, my lovely kids (Hephzibah, Joshua, Beulah, and Christian), and my aunt: Mrs Ola C. Alaribe. Remain blessed for being there for me. Last but not the least; I appreciate all those who in one way or another contributed to the success of this research work.

REFERENCES

[1] **Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R. &Honrao, V.** Predicting students' performance using ID3 and C4.5 classification algorithms. *International Journal of Data Mining & Knowledge Management Process*, *3*(5), 2013, 39-52.

[2] **Zheyuan F., Yuming K., Xiaolin, L.** Chess game results prediction system. *Stanford University CS Machine Learning Project Report, 2013*,Retrieved from:http://cs229.stanford.edu/proj2013/FanKuangLin-ChessGameResultPredictionSystem.pdf .

[3] **Buursma, D.,**Predicting sports event from past result: Towards effective betting on football matches. *Preceding 14thTwente Student Conference on IT. University of Twente, Faculty Electrical Engineering, Mathematics and Computer Science, Netherlands*, 2011. Retrieved from: http://referaat.cs.utwente.nl/conference/14/paper/7226/predicting-sports-events-from-past-results.pdf

[4] **Farzin, O., Parinaz, E., &Faezeh, S. M.** Football result prediction with Bayesian network in Spanish league-Barcelona team. *International Journal of Computer Theory and Engineering*, *5*(5), 2013, 812-815.

[5] **Baio, G. &Blangiardo, M.** Bayesian hierarchical model for prediction of football results. *Discovery, University College London*., **2010,** retrieved fromhttp:discovery.ucl.ac.uk.

[6] **Nivard, W. & Mei, R. D.***Soccer analytics: Predicting the of soccer matches*. (Master thesis: UV University of Amsterdam), 2012,Retrieved from www.few.vu.nl/en/Images/werkstuk-wijk_tcm39-315158.pdf

[7] **Courneya, K. S., & Carron, A. V.** The home advantage in sport competitions: A literature review. Journal of Sport and Exercise Psychology, *14(1*), 1999, 13-27.

[8] **Buraimo, B., Forrest, D., & Simmons, R.,**The 12th man? Refereeing bias in English and German football. *Journal of Royal Statistics Society. Series A: Statistician in Society*,*173*(2), 2012, 431-449.

[9] **Goddard, J. &AsimakopoulosI.,** Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting 23*(1), 2004, 51–66.

[10] **Clarke, S. R. & Norman, J. M.** Home ground advantage of individual clubs in English football. *Statistician, 44*, 1995, 509-521.

[11] **Drawer S. & Fuller C. W.** Evaluating the level of injury in English professional football using a risk based assessment process. *Br J Sports Med 36.* , 2002, 446–451.

[12] **Webby, R. & O'Connor, M.**. Judgmental and statistical time series forecasting: a review of the literature. *International Journal of' Forecasting 12, 1996*, 91-118.

[13] **Battle R., Bridge T., Bull A., Hanson C. Taylor R. & Thorpe A**. The top 20 Money League clubs generated combined revenues of €4.4 billion in 2010/11, over a quarter of the European football market. *Football Money League 2012 Sports Business Group*.Deloitte, 2012, retrieved from: http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/sports-business-group/deloitte-uk-deloitte-football-money-league-2012.pdf.

[14] **Akhtar, F. &Hahne, C.***RapidMiner 5 Operator Reference*. (Rapid-I GmbH.*)*.2012, retrieved from: http://rapidminer.com/wp-content/uploads/2013/10/RapidMiner_OperatorReference_en.pdf.

AUTHORS' BRIEF

Igiri Chinwe Peace holds a B.Eng. in Computer Science and Engineering from Enugu State University of Science & Technology, in 2001. She is a currently at the concluding stages of Master's degree programme at the department of Computer Science, University of Port Harcourt, Nigeria. Her research area are data base technology, data mining and soft engineering. She is a member of Nigeria Society of Engineering, ACM and IEEE-Computer Society. She can be reached via chynkemdirim@gmail.com and +2348036012095.

Enoch O. Nwachukwu is a professor of Computer Science and former Head of Department of Computer Science, Faculty of Physical Science and Information Technology, University of Port Harcourt, Nigeria. His research interests focus on Artificial Intelligence applications, Expert Systems, Software Engineering, and Network Security. He has a first degree in Electronics/Electrical Engineering, University of Ife, Nigeria, and received his MSc and PhD in Computer Science from The University of Manchester UK. He is a Fellow of Nigeria Computer Society and a Member of Nigeria Society of Engineers. Contact him at Department of Computer Science, University of Port Harcourt, P. O. Box 258, Uniport, Nigeria, and Email: ***enoch.nwachukwu@uniport.edu.ng***