International Journal of Computer Applications (0975 – 8887)
Volume 176 – No. 28, June 2020

Selecting Forward Players in a Football Team using Artificial Neural Networks

Abraham E. Evwiekpaefe
Department of Computer Science
Nigerian Defence Academy
Kaduna, Nigeria

Emmanuel Bitrus
Department of Computer Science
Nigerian Defence Academy
Kaduna, Nigeria

Fiyinfoluwa Ajakaiye
Department of Computer Science
Nigerian Defence Academy
Kaduna, Nigeria

ABSTRACT
The success of any football team lies in the performance of its players. Determining the best player among a pool of players is a very difficult task. The purpose of this research is to assess the performance skills of forward football players in a football game. To conduct this research, players were randomly selected from different teams across Europe based on their play positions. One hundred (100) forward players were selected for the analysis. Performance analysis was conducted using Artificial Neural Networks (ANN) Multilayer Perception and compared with the J48 classifier. A model based on the ANN Multilayer Perception was trained and developed using secondary data collected from the online Complete Dataset of the FIFA 2017/2018 football season. The analysis was done with the aid of the WEKA data mining tool. The results show that the Multilayer Perception classification had a better performance than the J48 classification.

General Terms
Classification, Data mining, FIFA, Football

Keywords
Artificial Neural Networks (ANN), Forward, J48, Multilayer Perceptron, Player, Selection, WEKA.

1. INTRODUCTION
Football is a team sport that is popular in almost every country in the world. The player selection process for football teams is crucial in the quest for making a good team. So much so that a wrong formation of a team can cost a football team the loss of several titles and awards and even a lot of money if the players selected do not live up to the team's expectations. There is no doubt that the assessment of a player is of great benefit and extremely useful when trying to form a good football team.

Putting together a successful football team depends on the management's ability to determine and select the best players amongst the pool of players. The management of the football team, in a bid to aid their decision-making process, has to make use of a decision support machine (system) to determine the best player using certain attributes. These attributes may include the player's speed, technique, physical fitness, defensive power; attacking power, current form, age, dribbling power, injuries, among others.

A football game has several formations of how the players are to play a match and each player has his own role to perform in the field which also depends on the position the player is playing. Figure 1 shows an example of one of the formations in a football game and the players circled in yellow are the forward players. The maximum or minimum number of forward players mainly depends on the formation. These are named as right and left wing forwards, right and left strikers and center forwards. They are primarily responsible for scoring goals.

In this research, a model for forward player selection in a football team shall be built from a pool of players using artificial neural network Multilayer Perceptron techniques. Some attributes, with the aid of the WEKA data mining tool, will be used to decide which players are good, average and below-average.

Figure 1: Football formation 5-4-1

2. REVIEW OF RELATED WORK
Torgler & Schmidt [1] investigated the pay-performance relationship of football players utilizing data from eight seasons of the German Bundesliga. The results of the panel analysis showed that salaries have a positive impact on players' performance but with diminishing returns tendency. Furthermore, their empirical findings demonstrated a strong impact of a player's relative income level on his performance. Their analysis provided evidence of a direct impact of teammates' attributes, like age, nationality and position he is playing, on individual player performance. Correlation matrix was used to analyze the performance of the players based on their salary and position. The position of a player had a strong impact on his performance.

Feng et al. [2] proposed a novel method for member selection of cross-functional teams where both the individual performance of the candidates and the collaborative performance between candidates were considered. An improved non dominated sorting genetic algorithm II (INSGA-II) was built to solve the model.

Dey et al. [3] discovered that team selection is very vital in cricket as players are chosen according to their total contributions which hinge on numerous factors and it becomes more essential when a large sum of money is required. In the paper, the Analytical Hierarchy Process (AHP) was used to estimate a player's price in cricket. This was determined by the previous performance, experience and other characteristics of the individual player's form. The proposed model consisted of two main parts. In the first part, Analytical Hierarchy Process was used to compute the relative importance of the attributes while in the second, Artificial Neural Network Back Propagation (ANN-BP) was applied to train the model and generate the player's price as an output.

Uzochukwu & Enyindah [4] used Artificial Neural Network to build a decision support system for players' selection in a football team. Attributes of players used in the study included player's resistance, speed, physical status and technique. Neural Network was used as the method to determine which player was the best. The system was developed and implemented in MATLAB data mining tool.

Al-Shboul et al. [5] were concerned with creating a tool that allowed football team management conduct analysis on a collection of players and generated a ranking based on the analysis. The paper presented a design of an ANN that was tailored to assist team managers in selecting a team that will provide the best performance against a given opposition. Semi-supervised learning approach was used in order to quantify and predict player performance from team data with mutual influence among players, and the report win accuracies was around 60%.

Passi & Pandey [6] predicted players' performance in One Day International (ODI) matches by analysing their characteristics and statistics using supervised machine learning techniques. According to the authors, the performance of the players relied on many attributes such as the opposition team, the venue, their current form, etc. They attempted to forecast the performance of players as how many runs each batsman will score and how many wickets each bowler will take for both the teams. Both the problems were targeted as classification problems where several runs and the number of wickets were classified in different ranges. Naïve Bayes, Random Forest, Multiclass SVM and Decision Tree classifiers were used to generate the prediction models for both problems. The results demonstrated Random Forest as the most accurate classifier for both the problems.

From the review of literature, researches have been conducted to determine the performance of players in a football game using selected parameters. However, with regards to forward players not much has been done. This study intends to fill this gap. Hence this study will determine the ratings (good, average, or below-average) of forward players using a combination of attributes from a pool of players available for selection.

3. METHODOLOGY

The study evaluated the players based on 'good', 'average', or 'below-average' rating. The forward players were selected from a pool of players, using a combination of attributes. A dataset of one hundred (100) players was selected and divided into two parts: 90 for training and 10 for testing using Multilayer Perceptron in the WEKA (Waikato Environment for Knowledge Analysis) environment.

3.1 Data Mining Technique Used

Diverse types of data mining techniques are available such as association rule mining, K-means clustering, Artificial Neural Network etc. Among the various approaches, Artificial Neural Network Multilayer Perceptron was chosen for this research. Multilayer Perceptron is one of the most commonly used Neural Network architecture due to its low complexity and ability to produce satisfactory result for non-linear relationships. This network structure is usually trained using supervised learning (Hemalatha and Rani, [7]).

A neural network is a group of interconnected input/output elements in which every connection has a weight connected with it. During the learning phase, the network learns by regulating its weights so it can correctly predict the class label of the input tuples (Han & Kamber, [8]).

According Sebastian [9], a Multilayer Perception possesses the following:

i. Has any number of inputs.

ii. Has one or more hidden layers with any number of units

iii. Uses generally sigmoid activation functions in the hidden layers.

iv. Has connection between the input layer and the first hidden layer, between the hidden layers, and between the last hidden layer and the output layer. See Figure 2.

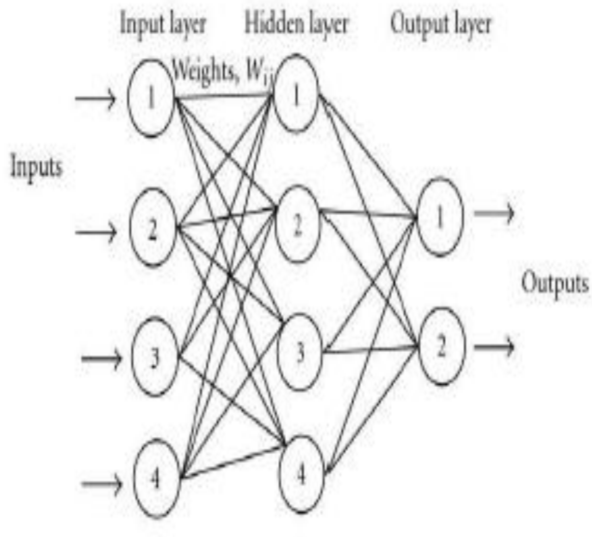

Figure 2: A Multilayer Perception Source: Sebastian [9]

3.2 Attributes Selection

In the ANN model creation, different attributes needed for player selection were analysed. These attributes were used to make the final assessments on the class of player to be selected.

The attributes used in this research are a combination of those used by Uzochukwu & Enyindah [4], Passi & Pandey [6]. Table 1 shows the combined attributes adopted.

Table 1: The adopted Attributes, Description and Values

Attribute	Description	Values
Name	The name of the player	Name
Age	The age of the player	Good Average Below average
Position	The position he is playing	Striker (ST) Left Winger (LW) Right Winger (RW)
Acceleration	The average speed of the player	Good Average Below average
Agility	The average ability of how the player moves quick and easy	Good Average Below average
Balance	The average of how upright and steady.	Good Average Below average
Ball control	The average ability to tackle the ball	Good Average Below average
Composure	The average state of being calm in the field	Good Average Below average
Crossing	The average over hit balls	Good Average Below average
Dribbling	The average number of disposing of a player	Good Average Below average
Finishing	The average number of goals score	Good Average Below average
Free kick	The average number of free kick taken by the player	Good Average Below average
Heading	The average number of ball headed by the player	Good Average Below average
Interception	The average number of ball stopped or prevented	Good Average Below average
Jumping	The jumping ability	Good Average Below average
Long passing	The average long pass accuracy	Good Average Below average
Long shot	The average long shot power accuracy	Good Average Below average
Penalty kick	The average number of goals score in penalty	Good Average Below average
Positioning	The ability of always being in the right position of play	Good Average Below average
Short passing	The average accurate short passing	Good Average Below average
Shot power	The shot power of a player	Good Average Below average
Stamina	The ability to endure physical or mental activity over a long time	Good Average Below average
Strength	state of being physically strong	Good Average Below average

3.3 Designing Artificial Neural Network (ANN) Models

Designing ANN models follow a number of systemic procedures. In general, there are four basic steps: (1) data collection, (2) data pre-processing, (3) training model (4) Testing the performance of model.

3.3.1 Data Collection

The data used in this research was collected online from the FIFA 2017/2018 Complete Dataset. A dataset of 100 players was used for the evaluation and Artificial Neural network was used for predicting the performance skills of the targeted forward players.

3.3.2 WEKA Data Mining Tool

WEKA is an acronym for Waikato Environment for Knowledge Analysis. It was built by the University of Waikato in New Zealand. It has a collection of machine learning algorithms for data mining and machine learning tasks. WEKA provides implementations of learning algorithms that can be applied easily to one's dataset. It also includes a variety of tools for transforming datasets, such as the algorithms for discretization (Ian & Eibe, [10]).

WEKA supports several data mining tasks such as data pre-processing, classification, clustering, visualization, regression, association rules mining, feature selection to name a few. Commonly supported data formats in WEKA are ARFF and CSV. Furthermore, one can also import from a URL or an SQL database.

3.3.3 Data Pre-Processing and Training model

The training of the dataset was done using the Multilayer Perceptron neural network. 10 fold cross validation was used for training and testing the data set. Here, the entire data set was split into ten (10) equal subsets (folds). That is, split the dataset into 10 parts (folds), hold out each piece in turn, and average the outcomes. So each data piece in the dataset is used once for testing and 9 times for training. This is 10-fold cross-validation.

3.3.4 Testing the performance of model in WEKA

To evaluate the performance of neural network model, different parameters are available such as Accuracy, Precision, Recall, F-Measure, Kappa score etc (Powers, [11]). Here accuracy, precision and recall were measured. According to Sebastian [9], these parameters are defined thus:

i. Accuracy is the outcome of the correctly predicted observation to the total dataset or actual (true) value.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

ii. Precision is the outcome of the correctly predicted positive values to the overall predicted positive values.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

iii. Recall: It is also called sensitivity. It refers to the true positive rate.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Where:

TP: True Positive: Predicted values correctly predicted as actual positive

FP: Predicted values incorrectly predicted an actual positive. i.e., Negative values predicted as positive

FN: False Negative: Positive values predicted as negative

TN: True Negative: Predicted values correctly predicted as an actual negative.

3.3.5 Model Performance Evaluation

To measure its fitness, Multilayer Perceptron was consequently compared with J48 decision tree classifier. This is because J48 has a good ability to deal with default data and data with noise, and has higher classification accuracy. In addition, it is a non-linear classifier suitable for the condition that the judgment of factors is relatively less, in the same time, the relationship of logic combination is not complicated (Chun yan et al., [12]).

4. RESULTS AND DISCSSIONS

4.1 Results

4.1.1 Results of 10-fold cross validation

Here the results of the 10-fold cross validation of the training analysis carried out using Multilayer Perceptron and J48 classifiers are demonstrated. This is indicated via the following headings: Method, learning rate, momentum rate, accuracy, precision and recall as shown in Table 2, Figure 3 and Figure 4.

Table 2: Result of 10-fold Cross Validation of Training and testing for both classifiers

Classifier	**Method**	**Learning Rate**	**Momentum Rate**	**Accuracy (%)**	**Precision**	**Recall**
Multilayer Perception	10-fold Cross Validation of Training	0.3	0.2	68	0.623	0.680
J48	K-fold Cross Validation of Train	-	-	57	0.520	0.570

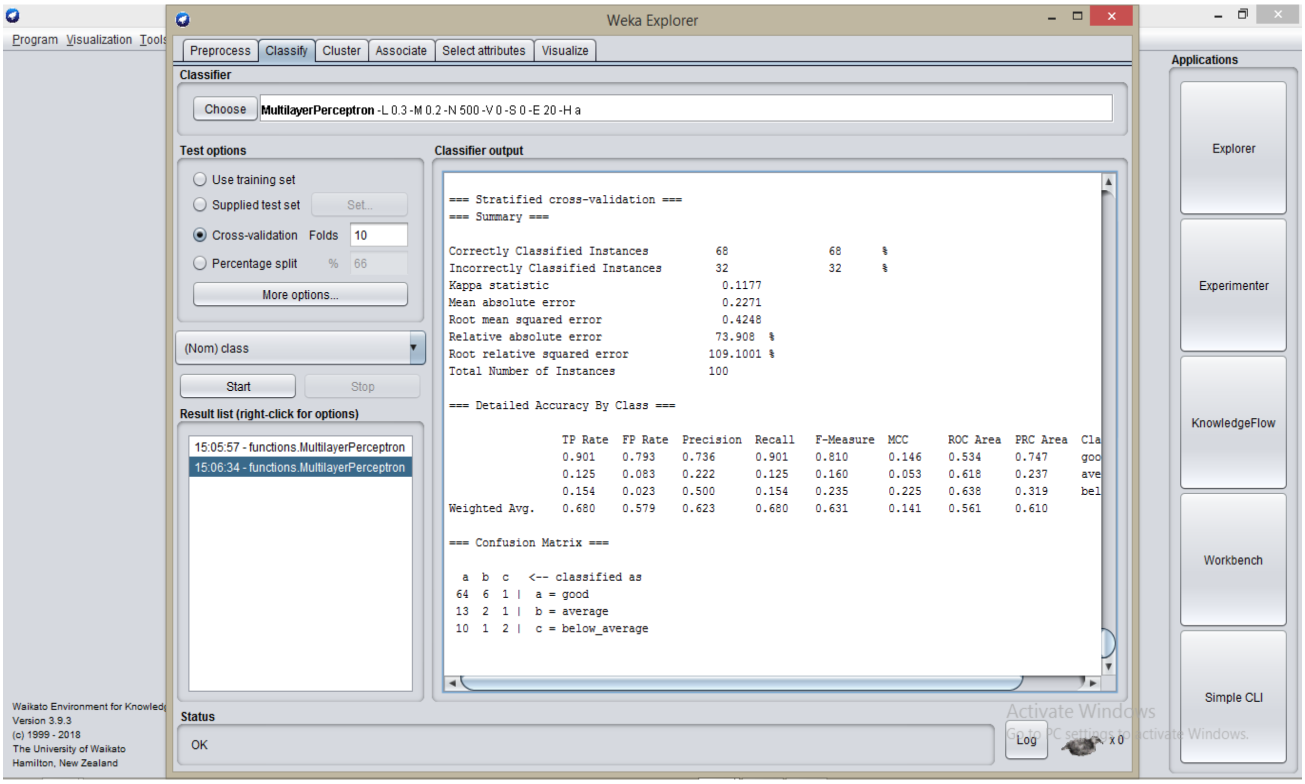

Figure 3: Result of 10-fold Cross Validation of Training and testing and confusion matrix for Multilayer Perceptron

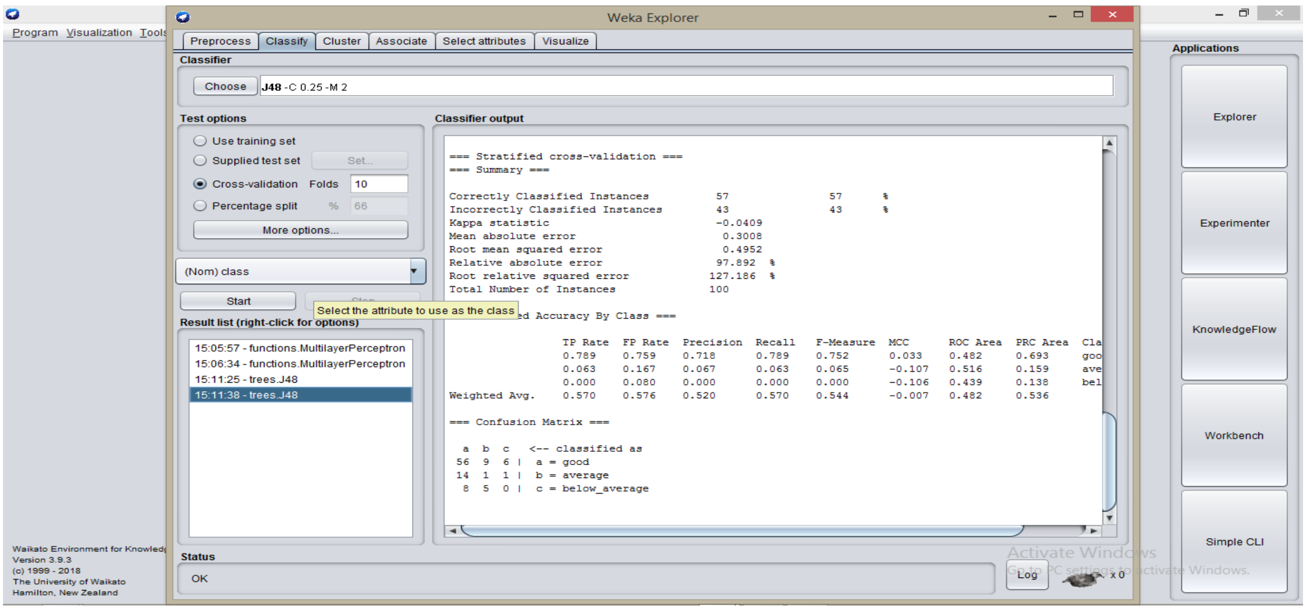

Figure 4: Result of 10-fold Cross Validation of Training and testing and confusion matrix for J48 classifier

4.1.2 Confusion matrix

Table 3 and Figure 3 display the confusion matrix of the Multilayer Perceptron while Table 4 and Figure 4 demonstrate the confusion matrix of the J48 classifier.

Table 3 Result of confusion matrix of Multilayer Perceptron

Output		Predicted			
		Good	Average	Below average	
Actual	Good	64	6	1	71
	Average	13	2	1	16
	Below average	10	1	2	13
		87	9	4	

Table 4 Result of confusion matrix for J48

Output		Predicted			
		Good	Average	Below average	
Actual	Good	56	9	6	71
	Average	14	1	1	16
	Below average	8	5	0	13
		78	15	7	

4.2 Discussion
The performance of the classification algorithms was evaluated on the basis of accuracy, recall and precision. Accuracy is defined as the number of all correct predictions over total number of the dataset. The best accuracy is 1.0, whereas the worst is 0.0. Precision is defined as number of correct positive prediction over total number of positive prediction and recall is defined as number of correct positive prediction over total number of positive cases. A high precision indicates that algorithm returns more relevant results than irrelevant and high recall means that most of the results retuned by the algorithms are relevant.

Table 2 considered parameters as method, learning rate, momentum rate, the accuracy, precision and recall. Here the results show that 10-fold cross validation for Multilayer Perceptron classification is better than that of J48 classification. Multilayer Perceptron had accuracy, precision and recall values of 68%, 0.623 and 0.680 respectively as against 57%, 0.520 and 0.570 respectively for J48.

Tables 3 and 4 indicated the confusion matrix displaying the total number of players that are good, average and below-average. The Multilayer Perceptron classification was able to predict accurately 64 out of the 87 Good players, 2 out of the 9 Average players and another 2 out of the 4 Below-average players. This gives an accuracy of 73.6% for Good, 22.2% for Average and 50% for the Below-average player prediction. J48 classifier on the other hand predicted accurately 56 out of the 78 Good players, 1 out of the 15 Average players and none out of the 7 Below-average players giving an accuracy of 71.8% for Good, 6.67% for Average and 0% for the Below-average player prediction. Again, Multilayer Perceptron classifier outperformed the J48 classifier.

5. CONCLUSION
Artificial Neural Network Multilayer Perceptron was used in this paper to predict the performance skills of forward football players on a rating of 'good', 'average', and 'below-average'. A model was developed and subsequently compared with that of the J48 classifier. The findings of Multilayer Perceptron classification outclassed that of the J48 classification.

6. ACKNOWLEDGMENTS
Special thanks go to Janet Chiwetelu for carrying out the initial proof reading of this paper. Many thanks also to the staff members of the Department of Computer Science, Nigerian Defence Academy, Kaduna, Nigeria for their support.

7. REFERENCES
[1] Torgler, B., & Schmidt, S. 2005. What Shapes Player Performance in Soccer? Empirical findings from a panel analysis. Center for Research in Economics, Management, and the Arts.

[2] Feng, B., Jiang, Z. Z., Fan, Z. P., & Fu, N. 2010. A method for member selection of cross-functional functional teams using the individual and collaborative performances. European Journal of Operational Research 203, 652–661.

[3] Dey, P., Banerjee, A., Ghosh, D. & Mondal, A. 2014. AHP-Neural Network-Based Player Price Estimation in IPL. International Journal of Hybrid Information, 7.

[4] Uzochukwu, O. C., & Enyindah, P. 2015. A Machine Learning Application for Football Players' Selection. International Journal of Engineering Research & Technology (IJERT), 4(2278-0181).

[5] Al-Shboul, R., Syed, T., Memon, J., & Khan, F. 2017. Automated Player Selectionfor sports Team using Competitive Neural Network. (IJACSA) International Journal of Advances Computer Science and Applications., 8(8).

[6] Passi, K., & Pandey, K. 2018. Predicting players' performance in One-Day International cricket matches using machine learning. Computer Science & Information Technology, 111-126.

[7] Hemalatha K. and Rani K. U. 2017. Advancements in Multi-Layer Perceptron Training to Improve Classification Accuracy. International Journal on Recent and Innovation Trends in Computing and Communication, 5(6): 353 – 357

[8] Han, J., & Kamber, M. (2006). Data Mining: Concept and Techniques. (2nd. edition, Ed.) 500 Sansome Street, Suite 400, San Francisco, CA 94111: Morgan Kaufmann Publishers is an imprint of Elsevier.

[9] Sebastian, S. 2016. Performance Evaluation By Artificial Neural Network Using WEKA. International Research Journal of Engineering and Technology (IRJET), 3 (3).

[10] Ian, W. H., & Eibe, F. n.d. Data Mining Practical Machine Learning Tools and Techniques. (2nd Edition, Ed.) Morgan Kaufann Publishers.

[11] Powers, D.M.W.2011. Evaluation: From Precision, Recall and F-Measure to Roc, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies. 2 (1) :37-63

[12] Chun yan, N., Ju, W., Fang, H. and Reika, S. 2015. Application of J48 Decision Tree Classifier in Emotion Recognition Based on Chaos Characteristics. International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2015).