NIPES

Journal of Science and Technology Research

Journal homepage: www.nipesjournals.org.ng

Evaluation of Team's False '9' for Match Winner Prediction

Abiodun Tolulope Oluwayomi, Ajayi Olusola Olajide, Adegbite Adewuyi Adetayo, Aju Omojokun Gabriel, Orogun Adebola Okunola, Omomule Taiwo Gabriel

Department of Computer Science, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria

Article Info

***Keywords*:**
False '9', Support Vector Machine (SVM), Prediction, Win, Algorithm

Received 25 March 2021
Revised 18 April 2021
Accepted 07 June 2021
Available online 13 March 2022

https://doi.org/10.37933/nipes/4.1.2022.24

https://nipesjournals.org.ng

Abstract

The quest to develop a concrete analysis on evaluation of teams false '9' for match winner prediction depends on the predictability of match results. The main concern of this research is on the efficiency of team manager tactics that deals with the usage and feeding of false '9' position in the football tactics used for a match. The effect of the false '9' role in match prediction was also evaluated. The recent applied statistics literature has focused primarily on modelling goal scoring. The prediction of the match depends on variables like team statistics, historical data etc. which is used by managers and club directors to decide who is going to win the match and what is needed to win the match but football result prediction has gained lots of popularity in recent years due to sports betting markets. A manager's tactics is now being used to develop a football match result predictive model by gathering the features that affect the outcome of football matches. Data extracted from site and their features (Attacking, Tackling, Technicality, Creativity and Defense) were used to compute the metric values of the prediction. Support Vector Machine (SVM) algorithm was used in prediction of team's performance using false '9' abilities to determine winning for the team. The system structure and the visible activities that take place within the system were also presented using appropriate design. The support vector machine system was implemented using Python as the software tool. The performance of the SVM is used to ascertain the level of accuracy of the prediction. Though many soccer predictive works have been done, this study was able to establish a prediction model using Support Vector Machine (SVM) algorithm for the evaluation of match winning considering false '9' roles, which is the first predictive model from that angle/specific aspect of football prediction and was able to give a 97% accuracy.

1. Introduction

The predictability of match results is the main concern of research on the efficiency of sports betting markets. The recent applied statistics literature has focused primarily on modelling goal scoring [1][2][3]. Football has been the most popular sport to be played and viewed in most European and South American countries. It is a game that unites everyone around the world because of its fun-filled attribute and it is the most watched generally not minding the league with goal scored an important decider of the winner or loser of the game. Figure 1, shows the data of average goal scored per match from 1992/1993 season to 2017/2018 season of the English Premier League.

Today, the team managers, fans and analysts give a prediction about who is going to win the match in football [4]. Nowadays, different statistics are used for the prediction of football match result. The prediction of a football match is done by using some previous data. The prediction of the match depends on variables like Team stats, Player stats, Historical data etc. which is used by managers and club directors to decide who is going to win the match and what is needed to win the match. Football result prediction has gained lots of popularity in recent years. Previously, knowledge discovery in databases (KDD) was used to develop a football match result predictive model by gathering the features that affect the outcome of football matches [5].
Data mining techniques have also been used in the past [6]. In existing systems, they have analyzed two approaches used in football match result prediction, statistical and machine learning [7][8][9]. For team management, the manager and the other coaching staff members need to know about their own team as well as the opponent team in detail. They should be able to identify strengths and weaknesses of different teams and accordingly prepare for the matches. The management should also be able to monitor a match in real time to take a decision during a match. For all these purposes, the prediction application should be able to suggest ideas based upon the provided conditions [10]. Traditional predictive methods have simply use match results to evaluate team performance and build statistical models to predict the results of future games [11]. It allows one to say that the presence of a risk factor increases the probability of a given outcome by a specific percentage. This study proposed a model of football match prediction by using the data of every match in premier league for the last 25 years and training the algorithm based on that data.

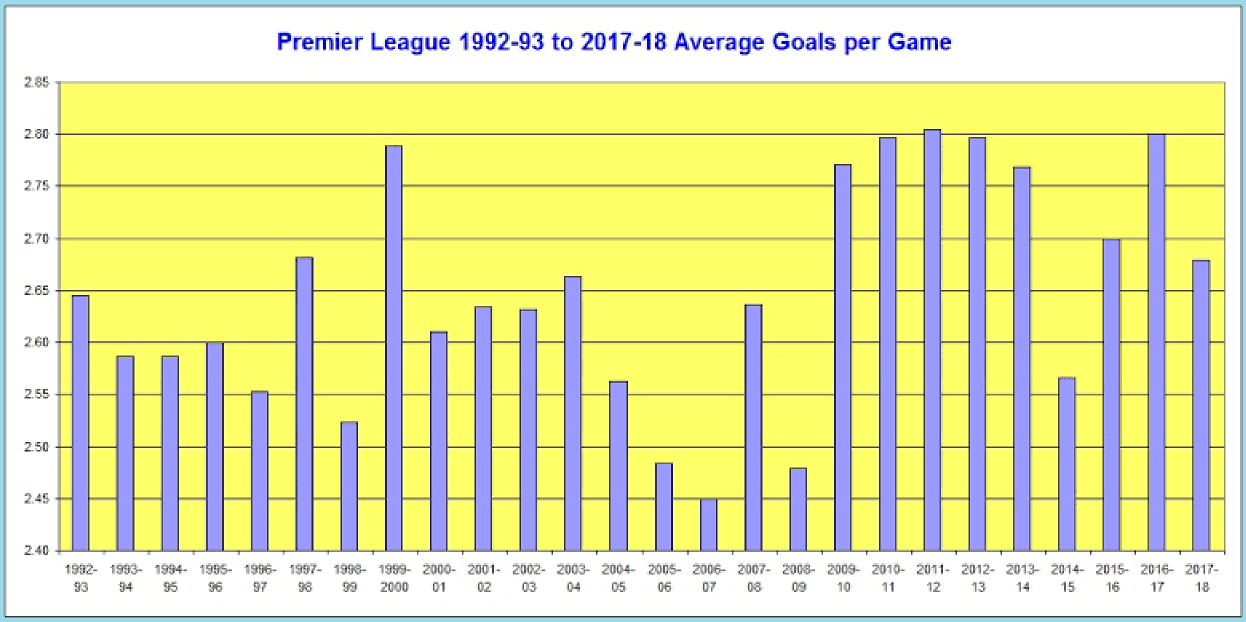

Figure 1: Average number of goals scored per game in the English Premier League (http://www.myfootballfacts.com/Premier_League_Goal_Statistics_1992-93_to_2017-18)[12]

The algorithm allows training data to form different rules and pattern and based on that they make a decision on some new data. With every input, the algorithm keeps making this decision by using the feedback. These results can be used by the management staff while selecting the team to use or in the type of tactics to use, and also by the fans playing fantasy leagues in order to build an efficient team or for the purpose of betting [13]. Another difference is that it can be used by all types of users (fans, management staff and bet placers) because it provides a wide range of functionalities rather than just prediction of win, lose or draw. As one of the most popular sports on the planet, football has always been followed very closely by a large number of people. In recent years, new types of data have been collected for many games in various countries, such as play-by-play data including information on each shot or pass made in a match.

According to every football team manager, different tactics is implemented based on the intensity of the game at hand but this paper considers the False '9' approach in any game no matter the intensity [14].

The false '9', in some ways similar to a more advanced attacking midfielder or playmaker role, is an unconventional lone striker or centre forward, who drops deep into the midfield. The purpose of this is that it creates a problem for opposing centre-backs who can either follow the false '9', leaving space behind them for onrushing midfielders, forward or wingers to exploit, or leaving the false '9' to have time and space to dribble or pick out a pass.

The term comes from traditional number for centre forwards (nine), and the fact that a centre forward traditionally stayed near the line of defenders until they get an opportunity to move past them towards goal. Key attributes for a false '9' are similar to those of a deep-lying striker: dribbling ability to take advantage of space between the lines, good short passing ability to link up with the midfield and vision to play through teammates making runs from deep to goal. In essence the use of a false '9' is an approach which, at its best, employs a terrific talent to the finest possible benefit to the team as a whole. The likes of Lionel Messi, when used in this position, will drift and dart, flick and spin, creating chances, spaces and opportunity for his fellow attacking teammates. This study therefore aim at evaluating team's false '9' performance to predict the winning probability of the team. The authors decided to use Support Vector Machine, which is one of the multiple machine learning algorithms that gives best accuracy [15][16][17][18]. The strength of Support Vector Machine lies in the fact that, decision boundaries are determined to minimize the classification error of both training data and unknown data [19][20][21][22].

1.2. Developmental Trends of Soccer Game

Humans possess the ability to adapt, and development programs must also adapt in order to continue to deliver results and keep up with the changes in football. Nowadays, it sometimes seems like football has been taken over by scientific research and analysis, but nevertheless creativity must remain at the nucleus of youth development. Of course, some of the conclusions of said research may be helpful in developing projects and merit consideration. For this reason, all of the suggestions presented should be adapted to the reality of the situation and to the qualities of the players. Clear but flexible systems all teams at the top level are well organized and adapt to the individual qualities of their players. The system of play forms the basic structure from which a team can evolve. However, systems are applied flexibly so that players can flourish and fulfill their potential. The best teams find the perfect balance between having clearly recognizable systems and strategies, allowing players a certain amount of freedom to take individual initiative and display creativity. The ability of a team to act with intelligence and flexibility by modifying tactics depending on the situation in the game, added to a positive aggression and the technical qualities of the players, are all key elements of success.

1.3. Different Approaches to Predicting Outcomes of Soccer Game

1.3.1. Tactical modelling: concepts and ideas

For coaches and researchers, tactical modelling can be helpful because it offers the opportunity to identify match regularities and random features of game events according to the offensive and defensive play. Obviously, the information about performance is crucial to achieve individual and team efficacy, because it constitutes a basic criterion for the training process. Several authors have been trying to outline significant tactical performance features in TS [23]. In this range, game modelling has been used to provide detection of patterns among match play events, according to the characteristics that afford players and team's success or failure. As stated by Lames and Hansen [24], it is important to ask whether models contain the essential attributes of the original game sport observed. That's why, recently, game sports research has become aware that another aspect of the model building process has perhaps not been given enough attention: the purpose of the model. In order to achiever deeper insight into the tactical game, it is necessary to record the substantial tactical actions in a chronological order, so the stream of tactical behavior can be recognized according to Pfeiffer et al [25]. In view of TS as the composite of complex interactions, systemic approach brings us to consider, among others, two main organizational levels: A match constitutes a complex system and the central notion of opposition leads us to consider two teams as interacting organized systems. The game may be thought of as living in the regions of meta-stability [26], where individual actions may serve to destabilize or (re)stabilize the system. The facility with which an attacker or a defender may destabilize or (re)stabilize the system would be considered a hallmark of quality in sport competition. In general terms, the ability of a team to destabilize or (re)stabilize a system might be examined at critical junctures of a game, say on the occurrence of an unexpected change of ball possession. Modelling a dynamic system means mapping not only its components and input-output behavior but also in particular its components interaction. From this viewpoint, the information about the interaction processes generated by the interactivity by teammates and opponents happens to show an outstanding relevancy because observing how interaction in a concurrent and competitive situation occurs can facilitate the design of specific and advantageous preparation. To date, research does not progress significantly to develop new and inclusive methods of dynamic analysis of sports contests, and particularly in TS further than the original work of Hughes et al [27]. Nevertheless, dynamic systems analyses may hold the key to unlocking the “hidden logic” of sports performance and variability within. The potential of these models to concentrate enormously complex behavior into simple expressions has been confirmed and offers a significant advantage over the labour intensive and inefficient approach required within traditional notational analysis. In order to describe and interpret game sequences in different sports, Anguera et al [28] suggest a noteworthy tool - the Observational Methodology. In this scope some authors have been using sequential analysis and polar-coordinates technique in their works. Amaral and Garganta [29] put forward an approach to game observation based on a double level analysis plan: i) the creation of a theoretical map with relevant match performance indicators regarding tactical organization; ii) the observation of game sequences and exploitation of data coming from both qualitative and quantitative analysis of team’s and player’s organization. Such an intention is very challenging due the nature and diversity of the constraints that compete for the success in TS, namely: i) the complexity concerning the plentiful relationships among the players: ii) the fact that game events do not correspond to a predictable sequence of actions; iii) the acute sensitivity of team and player’s behaviors to the initial conditions, taking into account the large number of variables and its interaction. For instance, in sports disciplines such as Soccer, Basketball or Handball, the teams compete for possession of the ball, which must be passed through a goal, while in Volleyball; the teams pass the ball in an attempt to place it in contact with an area of the opponent’s playing field. The teams involved in a match behave similar to self-organized systems searching for order and shape in a macroscopic plan, according to the interactions produced by the players. The individuality and degrees of freedom of team’s performance are dependent on a number of players and their possible interactions in game. Each team aims to disturb or to break the opponents’ balance, with the intention to generate disorder in its organization. On the other hand, teams intend to assure their own stability and organization. This way, the actions performed along the matches tend to assure space and time advantage over the contender, which means that the confrontation determines, usually, a winner and a loser. Because teams represent dynamical systems organized in accordance with principles and prescriptions, players and team’s behavior are generated from the tension among regularities and the production of novelty. In this sense, teams proceed as specialized systems strongly dominated by strategy and heuristic competences. Some years ago, Leon Teodorescu claims that it is not advisable to reduce TS to any algorithm model, because team action does not represent predictable sequences. Gréhaigne [30] appeals for a type of heuristic reasoning and he reinforces this idea referring that if the cascade of decisions will be restricted to an algorithm of binary choice, an impoverishment necessarily takes place, bringing about a limitation in game analysis. Lames and Hansen [24], alleged that the multi-causal structure of diagnosis in TS demands an interpretative rather than algorithmic approach [31]. The swat up of team's and player's tactical organization afford the possibility to identify game events, namely the identification of some pattern expressing preferential ways or forms of action, and the distinctive characters showing the variability of behaviors' and events . Lames and McGarry [23], asserts that what we see by observing a sports game is a dynamical interaction process in which measures and countermeasures are taken in an attempt to overcome the opponent. This implies that the behavior produced is not primarily the expression of stable properties of the individual players. In this context, the decision-making behavior is best considered at the level of the performer environment relationship and viewed as emerging from the interactions of individuals with environmental constraints over time specific functional goals. Therefore, the difficulty is that an adequate interpretation of numerical and visual data has to consider individual circumstances (tactics, strategy), but also situational aspects like physical and cognitive processes during the game, the quality of opponent and the preparation level.

1.3.2. The goal-ratio compare model

This model relies on the nested if-else combinations over the statistical data (goal-ratio data), to produce a result for the possible matchup of two teams. GR: Goals scored per match by a team
IF GRA – GRB ≥ 0.3 THEN higher wins ELSE IF 0.1 < GRA – GRB < 0.3 IF GRA > GRB THEN team A wins ELSE team A wins or draws ELSE team A wins or draws (GRA – GRB ≤ 0.1).

1.3.3. Last six matches comparison model

It is based on the scores of teams at the match date.(Score: Total points of a team) IF ScoreA – ScoreB ≥ 6 THEN higher wins ELSE IF Score A – ScoreB = 5 IF ScoreA > ScoreB THEN team A wins ELSE team A wins or draws ELSE IF ScoreA – ScoreB ≥ 2 THEN higher wins ELSE team A wins or draws (ScoreA – ScoreB ≤ 1)

2. Related Works

[32] proposed a prediction model for football matches using Fuzzy Logic. In the proposed system, results of matches between two football teams are generated based on the providence of several parameters associated with every football team. Fuzzy Logic Gaussian Membership function (Gmf) technique is adopted for the computation of membership grade for each input parameter. Datasets were obtained from the twenty (20) football teams in the English Premier League, 2017/2018 Season and were used to test the system predictability. The proposed model was implemented using MATLAB version R2017. Evaluation results showed the practicability of the system to the prediction of football outcomes based on the consideration of many input parameters for higher prediction accuracy when compared to reported literatures. [33] implemented its model using different machine learning algorithms and were able to reach the accuracy of 80.6% with Random Forest Algorithm on the Match History Database of one year along with the Team vs. Team Database. There are a lot of improvements that can be made to this system for improving the accuracy of both match result prediction as well as goal prediction. [34] worked on a model that predicts football match outcomes from all over the world. The author adopted both Dynamic ratings and Hybrid Bayesian Networks techniques. The techniques Dolores ranked 2nd in the competition with a predictive error 0.94% higher than the top and 116.78% lower than the bottom participants. With this output, it is observed that the prediction accuracy is higher than those of existing systems. However, this paper extends the empirical evidence to historical training data that does not just include match results from a single competition but contains results spanning different leagues and divisions from 35 different countries.

[35] worked on predicting football results using machine learning techniques. The author developed an 'expected goals' metric which helps to evaluate a team's performance, instead of using the actual

number of goals scored. The paper built a classification model predicting the outcome of future matches, as well as a regression model predicting the score of future games. The model performance compares favorably to existing traditional techniques and achieves a similar accuracy to bookmakers' models. [36] used several classifiers such as logistic regression, support vector machine and XGBoost algorithm devised on statistical analysis of past football games. The study able to make fairly accurate predictions. Although the accuracy of the model was pretty good, it is not guaranteed to be always right and there is a lot of scope for future work in this research. Sentiment analysis, features such as individual player and team performance metrics to further enhance the accuracy of the model.

[37] presents prediction of football match score and decision making process using machine learning algorithms such as Naive Bayes, Bayesian network, LogitBoost, The k-nearest neighbors algorithm, Random forest and Artificial neural networks. The implementation only includes teams from Spanish La Liga over the last 5 seasons. Prediction of the outcome of matches between Home Team and Away Team which would include the final score, the starting 11 players, the substitutes and the names of probable goal scorers. The best accuracy that they could reach was 65%. The drawback of the system is that it uses a very small dataset (96 matches) i.e. the data of 1 complete UCL season and the testing was done on the same. They implemented the model using different machine learning algorithms and were able to reach the accuracy of 71.63% with Logistic Regression. Referring to related works, there are a lot of improvements that can be made to improve the accuracy of both the match result prediction as well as goal prediction. Also, the two predictions can be later combined to further improve the model. [38] investigated the performance of a Support Vector Machine (SVM) with respect to the prediction of football matches. Gaussian combination kernel type is used to generate 79 support vectors at 100000 iterations. 16 example football match results (data sets) were trained to predict 15 matches. The findings showed 53.3% prediction accuracy, which is relatively low. Until proven otherwise by other studies, an SVM-based system is not good enough in this application domain.

[39] solved the problem of loss of interest in fantasy football over the season, a game-changing strategy was thought of which led to the creation of this idea. Powered by an exhaustive dataset of all football statistics from 1992 i.e. the start of the Premier League era, it seemed exciting to allow the use of Data Mining techniques to forecast future statistics. A points system based on the success of predictions, which in turn allow buying/auctioning better players adds a greater interactive feeling to the existing FPL system. This would prevent the churning of players of the season, since they would be attracted to getting more points and better players through such predictions. [40] implemented a Score-Time based Model for Handling Deadlocks in Football Group Matches, the study adopted the following methodologies, Phase 1 (Model Design): Score-time-based model, with appropriate algorithms was adopted. The Unified Modeling Language (UML) was used. Phase 2 (Model Implementation): An object-oriented approach was adopted for the implementation of the proposed model. Phase 3 (Model Demonstration) Java programming technique was deployed to achieve the implementation of the model. The paper was been able to establish the need to remodel the existing models of resolving deadlock in football soccer match. The article has to some extent, argue for the need to phase out the coin tossing method of deciding deadlock in football group matches. The major limitation of the proposed model is that it might not work for a situation whereby the teams in question all played goal-less draw in all their matches (as the name of the model suggests – score-time).However, the work can be bettered by employing a hybrid method, inculcating goal attempts.

[41] carried out the research work using object oriented approach. Predicting the outcome of soccer matches poses an interesting challenge for which it is realistically impossible to successfully do so

for every match. Despite this, there are lots of resources that are being expended on the correct prediction of soccer matches weekly, and all over the world. Soccer Match Result Prediction System Model (SMRPSM) is a system that is proposed where by the result of the match between two soccer teams are auto-generated, with the added excitement of given users a chance to test their predictive abilities. As many as possible, soccer teams are loaded in different leagues, with each team's corresponding manager and other information like team location, team logo and nickname. The user is also allowed to interact with the system by selecting the match to be predicted and viewing of the results of completed matches after registering/logging in. This work was unable to cover referee decision, team decision and team schedules. [42] worked on an improved prediction system for a football Match Result**.** The author adopted both artificial neural network (ANN) and logistic regression (LR) techniques with Rapid Miner as a data mining tool. The techniques yielded 85% and 93% prediction accuracy for ANN and LR techniques respectively. With the output, it was observed that the prediction accuracy was higher than those of existing systems. However, the study could not establish the accuracy of other prediction techniques.

[43] worked on predicting the results of soccer matches in the English Premier League (EPL) using artificial intelligence and machine learning algorithms. From historical data we created a feature set that includes game day data and current team performance (form). Using our feature data we created five different classifiers: Linear from stochastic gradient descent, Naive Bayes, Hidden Markov Model, Support Vector Machine (SVM), and Random Forest. Our prediction is in one of three classes for each game: win, draw, or loss. Our best error rates were with our linear classifier (.48), Random Forest (.50), and SVM (.50). Our error analysis focused on improving hyper parameters and class imbalance, which we approached using grid searches and ROC curve analysis respectively. [44] worked on predicting football results using Bayesian nets and other machine learning techniques**.** The authors adopted Bayesian networks (BNs) with additional ML techniques which were: MC4, a decision tree learner; Naive Bayesian learner; Data Driven Bayesian (a BN whose structure and node probability tables are learnt entirely from data); and a K-nearest neighbor learner. The technique accepted that the expert BN is generally superior to the other techniques for this domain in predictive accuracy. With the output, the results are even more impressive for BNs given that, in a number of key respects, the study assumptions place them at a disadvantage. The ability to provide accurate predictions without requiring much learning data are an obvious bonus. Moreover, the BN was relatively simple for the expert to build and its structure could be used again in this and similar types of problems. Prediction is of lesser accuracy compared to some other techniques.

[45] reviewed previous research on data mining systems to predict sports results and evaluates the advantages and disadvantages of each system. In the current world, sports produce considerable statistical information about each player, team, games, and seasons. Traditional sports science believed science to be owned by experts, coaches, team managers, and analyzers. However, sports organizations have recently realized the abundant science available in their data and sought to take advantage of that science through the use of data mining techniques. Sports data mining assists coaches and managers in result prediction, player performance assessment, player injury prediction, sports talent identification, and game strategy evaluation. The study in [46] applied the bayesian network model to predict the results fixtures involving the FC Barcelona team in the Spanish La Liga. They split the data-set into Non-physiological factors such as record of five previous matches, weather conditions, results for or against team; and physiological factors such as the players' mean the number of key players injured in the team, number of goals scored in all home and away NETICA software was used to design which produced values for players' average age medium, history of the last few games as win, injured main players, psychological state of the players and weather forecast during the match. They accomplished a prediction accuracy of 92% applied to

predict the 2014-2015 season fixtures. It was however noticed, that the study only considered using one team for the experimentation. This certainly was not enough to justify the outcome of the research.

[47] carried out the implementation of a system that can predict the result of matches in the UEFA Champions League with about 60% accuracy. They have used algorithms such as Naive Bayes, k-nearest neighbors, Random forest and Bayesian networks in order to obtain a suitable combination of features and classifiers required to make predictions. The software system was implemented in the Java programming language with use of Weka API. The study recommended other research work to look into the aspect of feature selection. Also, larger data sets for learning would prove beneficial to predict future outcomes with better accuracy. [48] presented a system that facilitates prediction of the winner in a sport game. The system consists of methods for: collection of data from the Internet for games in various sports, preprocessing of the acquired data, feature selection and model building. Many of the algorithms for prediction and classification implemented in Weka (Waikato Environment for Knowledge Analysis) have been tested for applicability for this kind of problems and a comparison of the results has been made. [49] presented an approach to estimate football match results with Neural Networks. Initially, they have classified a match into three categories using a Learning Vector Quantization neural network to determine the strength contrast between the two opponents. Then they use specific Back Propagation networks on the data they have designed according to the classifying result. They have trained and tested their model on actual football match outcomes from the famous Italian league, Serie A and have achieved a great performing model with good accuracy.

[50] studied multiple techniques in data mining and their prediction results are correlated to devise a good model for predicting matches of the Dutch football team. They use three major models namely Generalized Boosted Models (GBM), K-nearest neighbor and Naive Bayes classification. Using GBM, they attained 60.22% accuracy on average, while the other models were not as accurate. The results of the paper were based on a data-set that only included information about the Dutch team but no data regarding the opponent except for their FIFA ranking. To further improve this research, more data and statistics could be taken into account such as the opponent team's overall form in that season, and other factors such as head-to-head results or information about each team's previous games.

The existing literatures prove the use of several machine learning approach in football prediction such as Artificial Neural Network (ANN) and Logistics Regression (LR) to improve prediction system of a football match result [42], applied statistics used in modelling goal scoring [1],[2],[3], a model that predicts football outcome using Hybrid Bayesian approach and dynamic rating [34], using decision making process [37] data mining technique for result prediction [26], and using Artificial Intelligence and Machine Learning [43].
Prediction of the result of a football match has been placed upon the different method and analysis, which include a decision suggestion system for the football team managers to improve the team's performance. The decision support system would suggest team selection, tactic selection, analysis of opponent team, player evaluation etc. Traditionally however, when strategizing for match win, much emphasis is laid on the strike force compare to any other segment of the team. Much as the base team is tactically preparing for a win, the opposing team is also estimating the strength of the other's striking force. This has in the recent years, seen teams coming up with false '9' rather than relying solely on the real '9'. With little or no predictive analytical research work from this perspective, this study predicts match win based on evaluating team's false '9'.

2. Methodology

The implementation phase of this study shows the prediction of a football match win using evaluation of false 9', which is carried out using one machine learning algorithm SVM and Python 3.7 IDLE script. The implementation data was pre-processed from www.sofascore.com designed with the credibility criteria and further prepared using Microsoft Excel Spreadsheet.
The training and testing are done on the complete dataset using train_test_split, which splits out your data into a training set and a test set. We provide the proportion of data to use as a test set and we can provide the parameter random_state, which is a seed to ensure repeatable results. The main part of our implementation was concentrated on team's false 9' performance evaluation to determine win for their respective clubs. The Player Ability Database includes 5 attributes: Technical, Defending, Creativity, Tactical, and Attacking.
A list of several players who played as a false 9 across several leagues in the Europe have been listed. These players have been assigned a Sample_ id from 1-160 based on their abilities in the seasons in which they feature where the player Sample_ id starts from 1.

This study adopts the Machine Learning Approach using Support Vector Machine (SVM). The aim of this research work was achieved using the following procedures:

1. **Data collection**: Obviously, to start testing algorithms for this problem a data set of players' abilities is needed. Thorough research helps to find the result and the statistics of all players' abilities on the internet, to our knowledge there is no publicly available dataset that can be downloaded and imported into some database. While a great deal of sports data can be effective in data mining, lack of a general dataset forces the researcher to collect the required data manually from valid sports websites.
2. **Data preprocessing**: After all required data is stored in a relational database, it must be preprocessed. The preprocessing may refer to: normalization and/or discretization of some parameters in a given range; or generating new parameters that did not exist in the original database. New parameters are generated by reviewing the data for the players. Previous games refer to games that were played before the date of the game whose data is being preprocessed and can contain data of games played by player and teams playing that particular game, but also games played by other teams. This means that none of the generated parameters uses "future" data, i.e. data that was not own before the beginning of a particular game. In other words, each generated parameter is time dependent, player dependent and team dependent. Some of the parameters that are generated in this module of the system are:

   - *Number of injured players in Team A before a particular game.* This parameter can be generated by reviewing the data from the previous game that Team A played, because it contains information why a particular player did not play – either it was coach's decision or the player was injured. Additionally, there are websites that publish information about injured players on a daily bases. The information retrieval from this kind of websites can also be automated.
   - *Winning streak (w) of Team A before a particular game*. This is done by counting how many games in a row have won *(w is positive)* or lost *(w is negative)* before that game.
   - *Fatigue of Team A before a particular game.* We are introducing this parameter to indicate how many times Team A have to travel in order to play the previous 7 games. Traveling a lot contributes to fatigue of the team. On the example shown on Figure 2 the fatigue of Team A before game 27 (the particular game) is estimated to be 5/6 because it had to travel 5 times. The maximum fatigue is 1 (if the team traveled 6 times) and the minimum is 0 (if the team played all the relevant games at home).

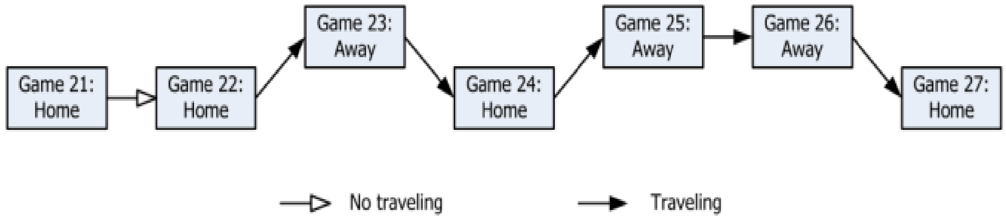

Figure 2. Example of fatigue estimation

• *Home, away and overall winning percentage*. The number of games won at home, divided by the number of games played at home, gives the home winning percentage. The calculation of the other parameters is similar.

• *Offensive, defensive and overall ratings of the team.* These ratings are calculated by formulas which are described in more details for various sports.

3. **Data analysis:** Data collected must be inspected, cleansed, transformed and modelled with the goal of discovering useful information, informing conclusions and supporting decision making.
4. **Feature selection:** After data collection, data analysis is executed and adding new features to the existing ones, the accuracy and speed of predictions will depend on proper automatic selection of the most significant, highly correlated features.
5. **Parameter optimization**: A Gaussian combination kernel type is used. Parameters including kernel sigma, kernel sigma2, kernel sigma3, kernel cache, constant C, convergence epsilon, and maximum iteration have been set to yield optimal prediction accuracy.
6. **Model :** We are trying a specific method to address our three-class classification problem:
   - *Support Vector Machines:* Literature suggested that the most successful model for this task was an SVM, so we began by implementing one with an RBF (Gaussian kernel): K(x, z) = e ||x−z||2/2σ 2. We used features of the false 9' abilities to determine the result of the game.

2.2.Proposed Model: Conceptual Model

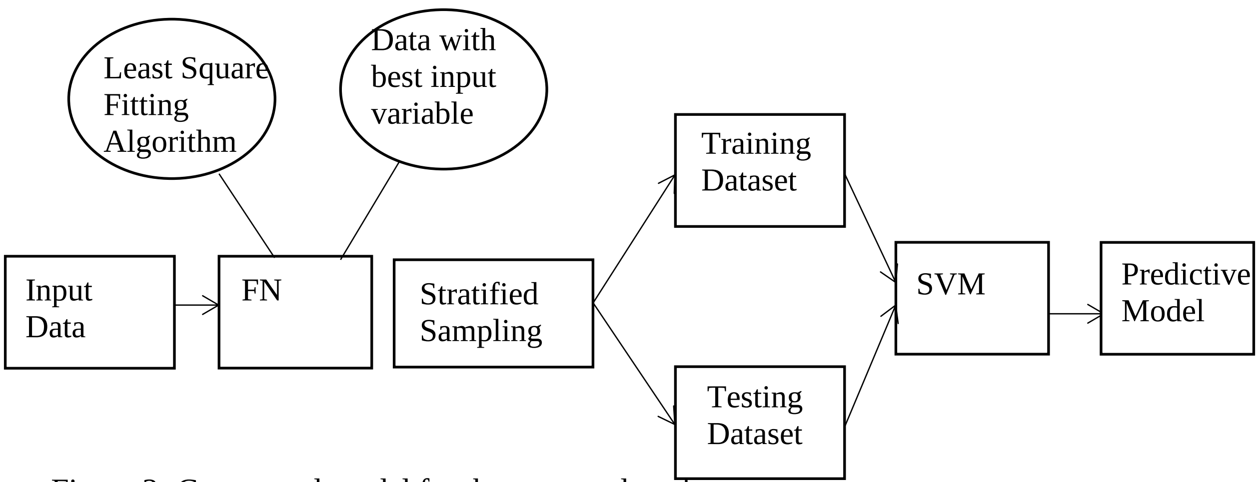

Figure 3**:** Conceptual model for the proposed study

2.2.1. Mathematical Model

A support vector machine constructs a hyper-plane or set of hyper-planes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyper-plane that has the largest distance to the nearest training data points of any class (so-called functional margin) seen in Figure 4, since in general the larger the margin the lower the generalization error of the classifier.

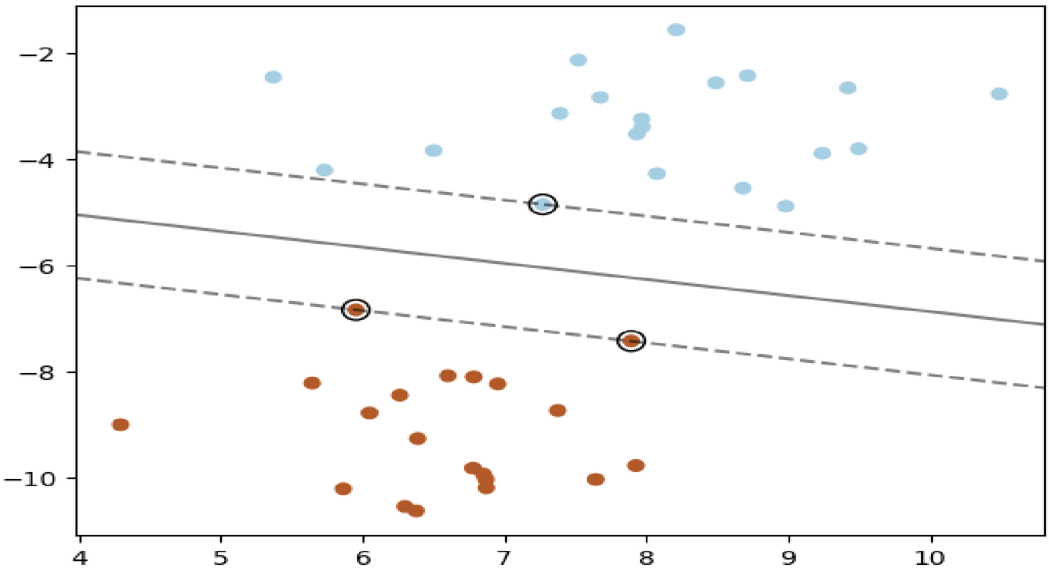

Figure 4: Functional Marginalization Graph

Training involves the minimization of the error function:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i$$

Subject to the constraints:

$$y_i\left(w^T\phi(x_i)+b\right)\geq 1-\xi_i \text{ and } \xi_i \geq 0, i=1,\dots,N$$

Where C is the capacity constant, w is the vector of coefficients, b is a constant, and ξ_i represents parameters for handling non-separable data (inputs). The index i labels the N training cases. Note that $y \in \pm 1$ represents the class labels and xi represents the independent variables. The kernel ϕ is used to transform data from the input (independent) to the feature space. It should be noted that the larger the C, the more the error is penalized as seen in Equation 1. Thus, C should be chosen with care to avoid over fitting.

Kernel Functions

$$K(\mathbf{X}_i,\mathbf{X}_j)=\begin{cases}\mathbf{X}_i\cdot\mathbf{X}_j & \text{Linear}\\ (\gamma\mathbf{X}_i\cdot\mathbf{X}_j+\mathrm{C})^d & \text{Polynomial}\\ \exp(-\gamma\,|\mathbf{X}_i-\mathbf{X}_j|^2) & \text{RBF}\\ \tanh(\gamma\mathbf{X}_i\cdot\mathbf{X}_j+\mathrm{C}) & \text{Sigmoid}\end{cases}$$ Equation 1

Where $K(\mathbf{X}_i,\mathbf{X}_j)=\phi(\mathbf{X}_i)\bullet\phi(\mathbf{X}_j)$

That is, the kernel function, represents a dot product of input data points mapped into the higher dimensional feature space by transformation ϕ.

According to Denzin and Lincoln [51], a research methodology or strategy is determined by the nature of the research question and the subject being investigated. As a result, the research format used in an investigation should be seen as a tool to answer the research question.

A methodology does not set out to provide solutions—it is therefore, not the same as a method. Instead, a methodology offers the theoretical underpinning for understanding which method, set of methods, or best practices can be applied to a specific case, for example, to calculate a specific result.
It has been defined also as follows:

1. "the analysis of the principles of methods, rules, and postulates employed by a discipline";
2. "the systematic study of methods that are, can be, or have been applied within a discipline";
3. "The study or description of methods".

This project aimed at exploring and understanding the meanings constructed by the participants. The study did not aim to provide the ultimate truth about the research topic but rather to investigate a particular way of looking at and deriving meaning on the phenomenon under investigation. This study was guided by the following questions:

- Who is a false 9?
- How does false 9's function?
- What are false 9's abilities?

2.3. Data Collection

Data collection method can either be qualitative or quantitative or mixed. Quantitative method examines numerical data and often requires the use of statistical tools to analyze data collected. This allows for the measurement of variables and relationships between them can then be established. This type of data can be represented using graphs and tables. The Technique for Qualitative research are interviews and observations, while quantitative researches are experiments and surveys. For the purpose of this project work we shall adopt quantitative research technique.
Quantitative Method is used to gather the data since it is numerical and it deals with evaluation.

2.4. Samples Data: The data which comprises of players used as false 9' with their abilities in past and present season for this study is extracted from www.sofascore.com manually by typing it into Microsoft Excel spreadsheet as the data will be imported into Python programming language.The sample of the study comprised data with the list of names and attributes of each players and the statistical numbers used in the attributes from all football clubs participating in the League across Europe over the period from 1980 to 2019.

The predictive attributes are enumerated below:

i. **Attacking**: making a forceful attempt to score or otherwise gain an advantage.
ii. **Technicality:** deals with mastering of the techniques of the game (passing, trapping, shooting, dribbling, tackling the ball.) having excellent touch, precise control, a fast work-rate of touches on the ball, and uses all sorts of body surfaces or edges.
iii. **Tactical:** involves dropping deep, allowing the opposition to have the ball and come forward with it, committing players forward and leaving gaps in behind as they go.
iv. **Creativity:** requires players to know themselves, work with each other, and understand the very best ways in which they can combine.
v. **Defending:** moving quickly to the player with the ball and preventing goal.

Considering all this attributes, a false '9' player has an average percentage or value of each of the attributes as shown in Figure 5.

Figure 5: Average attributes of a false '9' player

SVM: The Support Vector Machine (SVM) for classification has led to their application in a growing number of problem domains and to increasingly larger data sets. Linear classification using kernel matrix is used for the prediction [52].
Output: The result of the trained data in the SVM model will be the prediction accuracy of the match-winning false 9' player.
Dataset: The clinical data of 160 records considered for analysis has been taken from www.sofascore.com manually. The data will be implemented using Microsoft Excel Spreadsheet as the database and will be coded using Python programming language. There are 6 attributes in the dataset which include name, technical, defending, creativity, tactical, attacking. We filtered these attributes into a final list of 5 attributes which proved to be the most influential for predicting the outcome
Number of Instances: 160
Number of Attributes: 6
Class: Figures
Attribute Values: 1 to 100
Label Values: 0 and 1.

Table 1. Dataset for football players' abilities.

Sample_id	Name	Attacking	Technical	Creativity	Defending	Tactical	Mean	Label
1	firmino	85	75	77	26	57	64	1
2	messi	99	99	99	31	90	83.6	1
3	ronaldo	98	85	93	33	88	79.4	1
4	tevez	75	67	66	32	60	60	1
5	harzard	80	80	75	30	55	64	1
6	sanchez	75	70	60	20	70	59	1
7	eto'o	87	71	78	32	65	66.6	1
8	torres	70	56	35	15	50	45.2	0
9	insigne	80	65	50	33	50	55.6	1
10	van persie	70	60	70	24	55	55.8	1

2.4.Data Pre-processing

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors.
The data-set obtained from www.sofascore.com has several attributes. A lot of these features are pretty much unnecessary for making outcome predictions. Hence, the datasets was cleaned and only the features or attributes most needed was retained as shown for a player in Figure 6.

Figure 6: A sample player attributes showing the strength and weakness

Table 2: Pre-processed data

Attacking	Technical	Creativity	Defending	Tactical	Mean	Class_Label
85	75	77	26	57	64	1
99	99	99	31	90	83.6	1
98	85	93	33	88	79.4	1
75	67	66	32	60	60	1
80	80	75	30	55	64	1
75	70	60	20	70	59	1
87	71	78	32	65	66.6	1
70	56	35	15	50	45.2	0
80	65	50	33	50	55.6	1
70	60	70	24	55	55.8	1
80	60	44	18	49	50.2	1
77	44	60	34	33	49.6	0
70	44	56	37	55	52.4	1
77	46	56	23	67	53.8	1
75	54	55	22	45	50.2	1
80	77	65	25	55	60.4	1
70	70	55	25	67	57.4	1
60	60	69	32	45	53.2	1
83	40	23	28	44	43.6	0
88	67	71	29	65	64	1
60	64	50	12	33	43.8	0
50	50	30	22	23	35	0
40	55	23	28	65	42.2	0
74	34	66	20	45	47.8	0
65	32	54	23	44	43.6	0
78	44	20	32	22	39.2	0
33	39	45	35	55	41.4	0
88	45	22	10	21	37.2	0

Class Label: to get the relative attribute values each response was coded (1 and 0) and to get the class label for each sample the mean value for each row was taken (1, 2…5)
The following are the steps involved in data pre-processing:

Step 1: Gather the data
Step 2: Derive the class labels for each sample
Step 3: Check out the missing values
Step 4: See the Categorical Values
Step 5: Split the data-set into Training and Test Set
Step 6: Feature Scaling

2.5.Data splitting

After finishing the building of our new set of crucial attributes, we split the data into training and testing data. 70% of the data were used for training while 30% were used for the testing with an epoch of 105.

2.6.Detailed Architectural Model

Architectural model is popularly defined as a rich and rigorous diagram, created using available standards, in which the primary concern is to illustrate a specific set of tradeoffs inherent in the structure and design of a system or ecosystem.
A model is a complete, basic, and simplified description of software architecture which is composed of multiple views from a particular perspective or viewpoint. A view is a representation of an entire system from the perspective of a related set of concerns. It is used to describe the system from the viewpoint of different stakeholders such as end-users, developers, project managers, and testers.
In this module, we apply the machine learning classifier required for making our prediction.

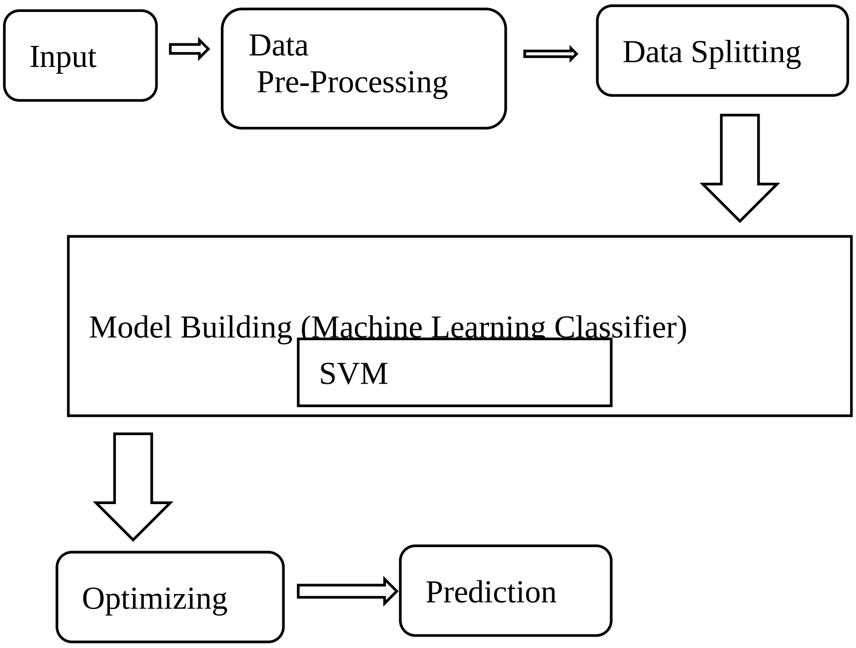

Figure 7: Architectural Design

2.7. **Design**

2.7.1. Activity Diagram

An activity diagram is a behavioral diagram i.e. it depicts the behavior of a system. An activity diagram portrays the control flow from a start point to a finish point showing the various decision paths that exist while the activity is being executed.

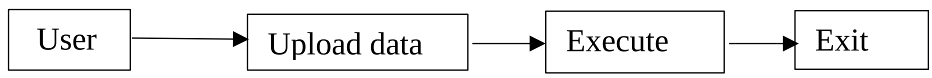

Figure 8: Activity diagram of the proposed system

2.8. **Sequence Diagram**

A *sequence diagram* simply depicts interaction between objects in a *sequential* order i.e. the order in which these interactions take place. We can also use the terms event *diagrams* or event scenarios to refer to a *sequence diagram*. *Sequence diagrams* describe how and in what order the objects in a system function.

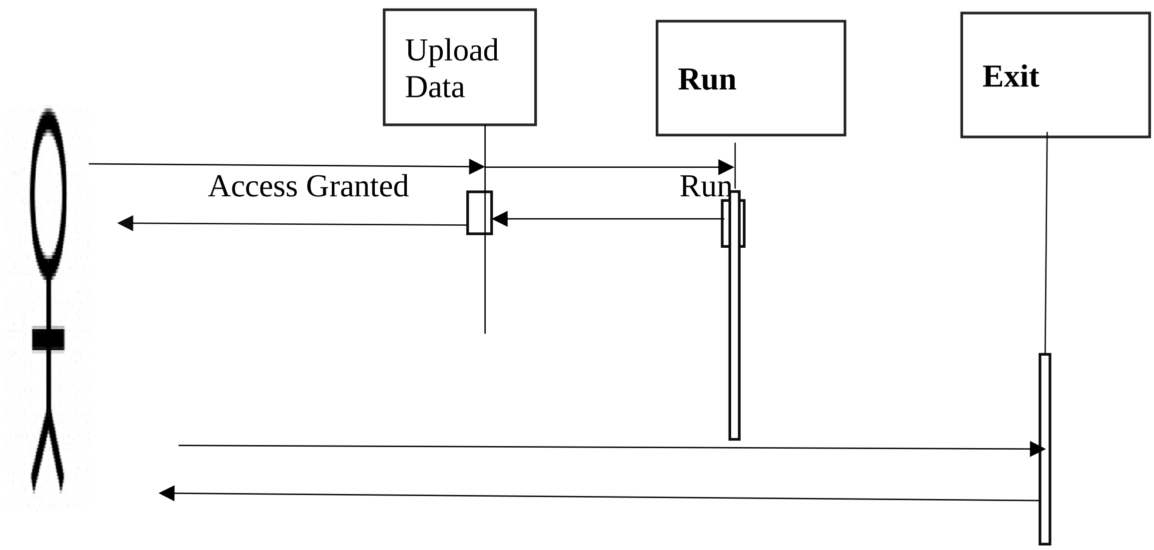

Figure 9: Sequence diagram of the proposed system

3. Results and Discussion

Input Interface

```
Tolu implementation.py - C:\Users\PROGRESSIVE\Desktop\MY_IMP\TOLEX\Tolu implementation.py (3.7.1)
File  Edit  Format  Run  Options  Window  Help
# -*- coding: utf-8 -*-

import pandas as pd
import numpy as np
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC

mydataset = pd.read_excel("C:\\Users\\PROGRESSIVE\\Desktop\\MY_IMP\\TOLEX\\Tolu_Data.xlsx")

print(mydataset.head(10))

print(" ")
print(" ")

x_train,x_test,y_train,y_test = train_test_split(X,Y,test_size=0.3, random_state = 10)

print(x_test)

print(" ")
print(" ")

print(x_train)

print(" ")
print(" ")

print(x_test.shape)
print(x_train.shape)
print(y_train.shape)
print(y_test.shape)

clf1 = SVC()
clf1.fit(x_train,y_train)
predict1 = clf1.predict(x_test)
percent1 = accuracy_score(y_test,predict1)
percent = percent1 * 100
print(" ")

print("SVC_score: ", percent1, "or", int(percent),"%")
print(" ")

print("confusion_matrix:\n", confusion_matrix(y_test,predict1))
print(" ")
print(" ")

print(mydataset.groupby("Class_Label").size())

print(" ")
print(" ")

summary = mydataset.describe()
print(summary)

mydataset = mydataset.values
print(" ")

print(mydataset)
```

Figure 10: Input interface

3.1. Output Interface

```
Python 3.7.1 Shell
File  Edit  Shell  Debug  Options  Window  Help
Python 3.7.1 (v3.7.1:260ec2c36a, Oct 20 2018, 14:05:16) [MSC v.1915 32 bit (Inte
l)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
======== RESTART: C:\Users\PROGRESSIVE\Desktop\MY_IMP\TOLEX\imple.py ========
   Attacking  Technical  Creativity  Defending  Tactical  Mean  Class_Label
0         85         75          77         26        57  64.0            1
1         99         99          99         31        90  83.6            1
2         98         85          93         33        88  79.4            1
3         75         67          66         32        60  60.0            1
4         80         80          75         30        55  64.0            1
5         75         70          60         20        70  59.0            1
6         87         71          78         32        65  66.6            1
7         70         56          35         15        50  45.2            0
8         80         65          50         33        50  55.6            1
9         70         60          70         24        55  55.8            1
```

Figure 11: Output Interface

```
       Attacking   Technical   Creativity  ...    Tactical        Mean  Class_La
bel
count  162.000000  162.000000  162.000000  ...  162.000000  162.000000    162.000
000
mean    72.771605   62.216049   59.629630  ...   53.623457   55.038272      0.666
667
std     14.794356   17.665884   17.869746  ...   17.623042    9.828177      0.472
866
min     22.000000    4.000000   20.000000  ...   21.000000   25.600000      0.000
000
25%     66.000000   50.000000   49.000000  ...   44.250000   48.650000      0.000
000
50%     75.000000   64.000000   61.000000  ...   53.000000   55.600000      1.000
000
75%     83.000000   75.750000   74.000000  ...   65.000000   60.600000      1.000
000
max     99.000000   99.000000   99.000000  ...   98.000000   83.600000      1.000
000

[8 rows x 7 columns]

[[85.  75.  77.  ... 57.  64.   1. ]
 [99.  99.  99.  ... 90.  83.6  1. ]
 [98.  85.  93.  ... 88.  79.4  1. ]
 ...
 [64.  51.  76.  ... 49.  54.8  1. ]
 [86.  71.  67.  ... 47.  59.2  1. ]
 [67.  64.  56.  ... 32.  49.6  0. ]]
```

```
(49, 6)
(113, 6)
(113,)
(49,)
```

```
SVC_score:  0.9795918367346939 or 97 %
```

Figure 12: Result Interface

From the result gotten from the model, it was able to predict effectively up to 97% accuracy the effect of including false ‘9’ in a match result. This accuracy is high enough to predict the match winning streaks with introduction of false ‘9’player by a manager in a match to be played.

3.2.Implementation Phases of the SVM model

3.2.1. Training Data

The training data is the information used to train the prediction algorithm. It includes both the input data and the expected output. The checking data is used to compare the performances of the prediction algorithms that were created based on the training data, help pick the one with the best performances. The testing data on the other hand, is used to assess how well the algorithm was trained, and confirm the actual predictive power of the algorithm.

It is the learning process of the developed model. The model is trained until the results obtained arrive at minimum error.

The training error is the difference between the training data output and the SVC output of the same training data input. The training set is a subset of the data set used to train a model. **x_train** is the training data set. **y_train** is the set of labels to all the data in **x_train**. The data used for the training stage is 70%.

```
[[67.  76.  44.  20.  40.  49.4]
 [74.  69.  74.  19.  65.  60.2]
 [73.  33.  74.  19.  84.  56.6]
 [80.  60.  44.  18.  49.  50.2]
 [45.  34.  77.  34.  54.  48.8]
 [69.  71.  74.  30.  54.  59.6]
 [88.  45.  22.  10.  21.  37.2]
 [78.  93.  64.  27.  73.  67. ]
 [88.  67.  71.  29.  65.  64. ]
 [78.  44.  20.  32.  22.  39.2]
 [90.  89.  75.  29.  74.  71.4]
 [78.  44.  20.  32.  22.  39.2]
 [90.  89.  47.  24.  53.  60.6]
 [66.  76.  53.  56.  22.  54.6]
 [67.  64.  56.  29.  32.  49.6]
 [84.  83.  60.  22.  74.  64.6]
 [75.  54.  55.  22.  45.  50.2]
 [65.  32.  54.  23.  44.  43.6]
 [67.  67.  40.  29.  34.  47.4]
 [75.  67.  66.  32.  60.  60. ]
 [83.  40.  23.  28.  44.  43.6]
 [67.  76.  44.  20.  40.  49.4]
 [88.  45.  22.  10.  21.  37.2]
 [56.  53.  62.  15.  53.  47.8]
 [79.  50.  80.  40.  33.  56.4]
 [33.  39.  45.  35.  55.  41.4]
 [38.  36.  49.  27.  45.  39. ]
 [86.  71.  67.  25.  47.  59.2]
```

Figure 13: Training data

3.2.2. Testing Data

A **test set** is therefore a **set** of examples used only to assess the performance (i.e. generalization) of a fully specified classifier. The testing set is a subset of the data set used to test a model. **x_test** is the testing data set. **y_test** is the set of labels to all the data in **x_test**. The data used for the testing stage is 30%.

```
[[86.   71.   67.   25.   47.   59.2]
 [70.   60.   70.   24.   55.   55.8]
 [63.   73.   63.   30.   57.   57.2]
 [65.   32.   54.   23.   44.   43.6]
 [90.   89.   47.   24.   53.   60.6]
 [80.   60.   44.   18.   49.   50.2]
 [87.   71.   78.   32.   65.   66.6]
 [80.   60.   44.   18.   49.   50.2]
 [83.   75.   67.   40.   67.   66.4]
 [63.   58.   83.   34.   74.   62.4]
 [83.   85.   85.   22.   85.   72. ]
 [75.   75.   85.   21.   69.   65. ]
 [88.   79.   76.   15.   73.   66.2]
 [60.   64.   50.   12.   33.   43.8]
 [63.   58.   83.   34.   74.   62.4]
 [67.   67.   40.   29.   34.   47.4]
 [70.   70.   55.   25.   67.   57.4]
 [99.   99.   99.   31.   90.   83.6]
 [85.   81.   48.   33.   70.   63.4]
 [77.   44.   60.   34.   33.   49.6]
 [88.   67.   71.   29.   65.   64. ]
 [70.   44.   56.   37.   55.   52.4]
 [75.   67.   66.   32.   60.   60. ]
 [60.   60.   69.   32.   45.   53.2]
 [50.   50.   30.   22.   23.   35. ]
 [73.   25.   67.   34.   35.   46.8]
```

Figure 14: Testing data

Output

```
        Attacking   Technical  Creativity  ...    Tactical        Mean  Class_La
bel
count  162.000000  162.000000  162.000000  ...  162.000000  162.000000   162.000
000
mean    72.771605   62.216049   59.629630  ...   53.623457   55.038272     0.666
667
std     14.794356   17.665884   17.869746  ...   17.623042    9.828177     0.472
866
min     22.000000    4.000000   20.000000  ...   21.000000   25.600000     0.000
000
25%     66.000000   50.000000   49.000000  ...   44.250000   48.650000     0.000
000
50%     75.000000   64.000000   61.000000  ...   53.000000   55.600000     1.000
000
75%     83.000000   75.750000   74.000000  ...   65.000000   60.600000     1.000
000
max     99.000000   99.000000   99.000000  ...   98.000000   83.600000     1.000
000

[8 rows x 7 columns]

[[85.  75.  77.  ... 57.  64.   1. ]
 [99.  99.  99.  ... 90.  83.6  1. ]
 [98.  85.  93.  ... 88.  79.4  1. ]
 ...
 [64.  51.  76.  ... 49.  54.8  1. ]
 [86.  71.  67.  ... 47.  59.2  1. ]
 [67.  64.  56.  ... 32.  49.6  0. ]]
```

Figure 15: Output

4. Conclusion and Recommendations

The false 9 are players that have the abilities to play like a striker; they have all it takes to perform using their abilities in their respective positions. Not all coaches adopt this method due to one or two reasons unknown but the managers that have used and those still using it are the likes of Jurgen Klopp, Pep Guardiola, Johan Cryuff, Sir Alex Fergusson etc., have all enjoyed the efficiency of this tactics which has helped them win major trophies across Europe. This study was developed to predict the usefulness of this tactics using support vector as its approach, with intent to make the tactics efficient in its usage and more accurate especially in Europe. The model was able to give a 97% accuracy in predicting of the effect of feeding false ‘9’ tactics in a football match.

This study is recommended to be used by Team managers across the world in other to successfully evaluate the tactics in helping them achieve their aim of winning trophies as expected of them.
In this present study, data used for the prediction were collected online. Data of football players that were used for this study was insufficient and few parameters were considered for this study. Other abilities can be used for further study to produce better results and accuracy. Aside support vector machine that this study has proved to be important in the football prediction, other soft-computing techniques like, artificial neural network (ANN), etc. can be deployed in false ‘9’ predictions for match winner.

References

[1] M.J. Dixon. and S.C. Coles. Modelling association football scores and inefficiencies in the football betting market. Applied Statistics 46: 265-280, 1997.
[2] H. Rue, and O. Salvesen. Prediction and retrospective analysis of soccer matches in a league. The Statistician 49: 399-418, 2000.
[3] M. Crowder, M. Dixon, A. Ledford, and M., Robinson. Dynamic modelling and prediction of English Football League matches for betting. The Statistician 51: 157-168, 2002.
[4] A. Tsakonas , G. Dounias , S. Shtovba and V. Vivdyuk. “Soft Computing-Based Result Prediction of Football Games”. // Proc. of the First International Conference on Inductive Modelling ICIM'2002, Lviv, Ukraine, 20-25 May 2002, Vol. 3. P. 15-21.
[5] D. Sharma, and V. Sharma. “Optimal Predictive Model for FIFA World Cup”. International Journal of Recent Technology and Engineering (IJRTE).ISSN: 2277-3878, Volume-7 Issue-6C, April 2019.
[6] F. Akhtar and C. Hahne. Rapid Miner. Rapid-I GmbH, 2012, Retrieved February13,2015from:http://rapidminer.com/wpcontent/uploads/2013/10/RapidMiner_OperatorReference_en , 2012.
[7] A.Boz. “Large Scale Machine Learning using NVIDIA CUDA,” retrieved May 7, 2015, from http://www.codeproject.com/Articles/336147/Large-Scale-Machine-Learning using- NVIDIA-CUDA, 2012. Technical blog., 2015
[8] J. R. Jang, Sun, C., and E. Mizutani. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, Prentice-Hall, Englewood Cliffs, NJ, pp. 7-9, 1997.
[9] L. Breiman, “Random Forests,” Machine Learning, vol. 45, 2001, 5–32.
[10] Teodorescu L (1985). Contributi al concetto di gioco sportivo di squadra. / Contribution to the concept of team sport. In In, L’insegnamento dei giochi sportivi, Rome, CONI, Scuola dello Sport, 1985, p. 19-35. Italy.
[11] K. Dimitris, and I. Ntzoufras. "Analysis Of Sports Data By Using Bivariate Poisson Models". “Journal of the Royal Statistical Society”: Series D (The Statistician) 52.3 (2003): 381- 393.
[12] http://www.myfootballfacts.com/Premier_League_Goal_Statistics_1992-93_to_2017-18.html
[13] S. M. Aqil Burnley, Nadeem Mahmood,Kashif Rizwan , Usman Amjad. “A Generic Approach for Team Selection in Multi– player Games using Genetic Algorithm”. International Journal of Computer Applications (0975 – 8887) Volume 40– No.17, February 2012.
[14] Tsakonas A., Dounias G., Shtovba S., Vivdyuk V. Soft Computing-Based Result Prediction of Football Games // Proc. of the First International Conference on Inductive Modelling ICIM'2002, Lviv, Ukraine, 20-25 May 2002, Vol. 3. P. 15-21
[15] B. Scholkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, “New Support Vector Algorithms,” Neural Computation 12, °c 2000 Massachusetts Institute of Technology, (2000), 1207–1245.
[16] P. Hall, J. Dean, I.K. Kabul, J. Silva, “An Overview of Machine Learning with SAS Enterprise Miner,” SAS Institute Inc, 2014. Retrieved May 6, 2015 from http://support.sas.com/resources/papers/proceedings14/SAS313-2014.
[17] S. Rüping (2000), mySVM-Manual, University of Dortmund, Lehrstuhl Informatik 8, http://www-ai.cs.unidortmund.de/SOFTWARE/MYSVM/).
[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, “Section 16.5. Support Vector Machines”. Numerical Recipes: The Art of Scientific Computing (3rd ed.), 2007, New York: Cambridge University Press. ISBN 978-0-521-88068-8.
[19] Alex J. Smola, Bernhard Schölkopf (2004) - A Tutorial on Support Vector Regression”, Statistics and Computing archive Volume 14 Issue 3, August 2004, p. 199-222.

[20] N. Ivars, A. Ludmila , and P. Inese. Neural Network Modelling for Sports Performance Classification as a Complex Socio-Technical System. ISSN 2255-9094 (online) ISSN 2255-9086 (print) December2016,vol.19,pp. 45–52 doi: 10.1515/itms-2016-0010 https://www.degruyter.com/view/j/itms.

[21] Q. S. Zhang, "A Multi-Label Learning Based Kernel Automatic Recommendation Method for Support Vector Machine," DOI: 10.1371/journal.pone.0120455, 2015.

[22] S.S. Subashka Ramesh, Nadeem Hassan, Anushka Khandelwal, Ritwiz Kaustoob, Sonal Gupta. "Analytics and Machine Learning Approaches to Generate Insights for Different Sports". International Journal of Recent Technology and Engineering (IJRTE). ISSN: 2277-3878, Volume-7, Issue-6, March 2019.

[23] M. Lames, and T. McGarry. On the search for reliable performance indicators in game sports. International Journal of Performance Analysis in Sport 7(1): 62-79, 2007.

[24] M. Lames, and G. Hansen. Designing observational systems to support top-level teams in game sports. International Journal of Performance Analysis in Sport 1(1): 83- 90, 2001.

[25] M. Pfeiffer, J. Perl. Analysis of tactical structures in team handball by means of artificial neural networks. International Journal of Computer Science in Sport 5(1): 4-14, 2006.

[26] S. Kelso. Dynamic Patterns: The self-organization of brain and behavior. Champaign, Illinois: Human Kinetics, 1995.

[27] M. Hughes, N. Dawkins, D. David, and J. Mills. The perturbation effect and goal opportunities in soccer. Journal of Sports Sciences 16: 20-21, 1998.

[28] J. A. Anguera, J. Boccanfuso, J. L. Rintoul, O. Al-Hashimi, F. Faraji, J. Janowich, E. Kong, Y. Larraburo, C. Rolle, E. Johnston, and A. Gazzaley. Video game training enhances cognitive control in older adults. Nature 501(7465):97-101, 2013.

[29] R. Amaral, and J. Garganta. Game modeling in futsal. Sequential analysis of 1x1 in attacking process. Revista ias do Desporto, 2005(3), 298-310, 2005.

[30] J. F. Grehaigne. A new method of goal analysis. Science and Football 5: 10-16, 1991.

[31] A. Malaise, A. Malibari, M. Alkhozae, (2014) "Student Performance Prediction System Using Multi-agent Technique," International Journal of Data Mining and Knowledge Process (IJDKP), vol. 4(5), 2014, 1–20.

[32] T. G. Omomule, A. J. Ibinuolapo, O. O. Ajayi. Fuzzy-Based Model for Predicting Football Match Results. International Journal of Scientific Research in Computer Science and Engineering Vol.8, Issue.1, pp.70-80, February (2020) E-ISSN: 2320-7639, 2020.

[33] R. G. Sonkamble, T. A. Dasare, S. S. Patil, S. B. Patil, I. C. Pawar and A. M. Phadtare. Sports Result Prediction System. International Journal of Research in Engineering, Science and Management 2(1),1-2, 2019

[34] A. C. Constantinou. Dolores: a model that predicts football match outcomes from all over the world. Machine Learning 108 (1), 49-75, 2019.

[35] C. Herbinet. Predicting Footbal Results using Machine Learning Techniques. M. Eng Thesis, Imperial College London, 2018.

[36] G. Anand, and M. Harini. English Footbal prediction using Machine Learning Classifiers. International Journal of Pure and Applied Mathematics 118(22): 533-536, 2018.

[37] Z. Nilay, S. Utkarsh, T. Shubham, S. Pramila, and K. T. Lalit. Prediction of Football Match Score and Decision Making Process. International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 6 Issue: 2 162 – 165, 2018.

[38] C. P. Igiri. Support Vector Machine–Based Prediction System for a Football Match Result. IOSR Journal of Computer Engineering 17(3), 21-26, 2015.

[39] A. Gangal, A. Talnikar, A. Dalvi, V. Zope, A. Kulkarni. Analysis and Prediction of Football Statistics using Data Mining. International Journal of Computer Applications 975, 8887, 2015.

[40] O.O. Ajayi. Implementing a Score-Time based Model for Handling Deadlocks in Football Group Matches. International Journal of Computer Applications (0975 – 8887) Volume 122 – No.21, July 2015.

[41] O.O. Ajayi and Alonge. Soccer Match Result prediction system: international science index journal vol:8 NO:11 2014, page 1173 – 1178, 2014.

[42] C. P. Igiri, and E. O. Nwachukwu. An Improved Prediction System for Football a Match Result". IOSR Journal of Engineering (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 2278-8719 Vol. 04, Issue 12 (December 2014): 12-20, 2014.

[43] B. Ulmer, M. Fernandez, and M. Peterson. Predicting Soccer Match Results in the English Premier League. Ph. D. Thesis, Doctoral dissertation, Stanford University, 2013.

[44] A. Joseph, A. E. Fenton, and M. Neil, (2006) "Predicting football results using Bayesian nets and other machine learning techniques." Knowledge-Based Systems 19.7 (2006): 544-553, 2006.

[45] M. Haghighat, H. Rastegari, and N. Nourafza. A Review of Data Mining Techniques for Result Prediction in Sports. Advances in Computer Science: An International Journal 2(5), 7-12, 2013.

[46] O. Farzin, Parinaz, E., and S. M. Faezeh. Football result prediction with Bayesian network in Spanish league-Barcelona team. International Journal of Computer Theory and Engineering, 5(5), 2013, 812-815, 2013.
[47] J. Hucaljuk and A. Rakipović . Predicting football scores using machine learning techniques. 2011 Proceedings of the 34th International Convention MIPRO,, 1623-1627, 2011.
[48] E. Zdravevski, and A. Kulakov. System for Prediction of the Winner in a Sports Game, In: ICT Innovations 2009, Part 2, 2010, pp. 55–63, 2009.
[49] T. Cheng, D. Cui, Z. Fan, J. Zhou, and S. Lu, “A new model to forecast the results of matches based on hybrid neural networks in the soccer rating system,” Proceedings Fifth International Conference on Computational Intelligence and Multimedia Applications. ICCIMA 2003.
[50] A. Hijmans and S. Bhulai. Dutch football prediction using machine learning classifiers. Research Paper Business analytics, 1-24, Unpulished.
[51] N. K. Denzin and Y. S. Lincoln.Introduction: The discipline and practice of qualitative research. In The Sage handbook of quantitative research, ed. Norman K. Denzin and Yvonna S. Lincoln, Thousand Oaks: SAGE Publications, 1-32, 2005.
[52] C. Souza, “Kernel Functions for Machine Learning Applications,” retrieved May 7, 2015 from http://crsouza.com/2010/03/kernelfunctions-for-machine-learning-applications/#anova.