NIPES Journal of Science and Technology Research 4(1) 2022 pp. 266-288 ISSN-2682-5821

NIPES

Journal of Science and Technology Research

Journal homepage: www.nipesjournals.org.ng

Assessing Team's Back '4' for Predicting Match Draw

Babalola Abdulhafeez Oluwabunmi, 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*:**
Prediction, Draw, Support Vector Machine (SVM), Back '4', Playing Position

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

doi Crossref NIPES JOURNALS

https://doi.org/10.37933/nipes/4.1.2022.23

https://nipesjournals.org.ng

Abstract

Many researches have tried to predict soccer outcomes using different methods such as supervised learning and unsupervised learning approaches. Many factors and features have been used to carry out this prediction. However, none so far has predicted the outcome of match results with emphasis and information on critical position of play in the entire team. Most teams lose by a slim margin, and this is vivid in the fact that 81.57% of matches have a goal difference per match that is less than two goals. This research critically considered the back '4' of teams, and was able to predict whether the match will be won, lost or end in stalemate with much emphasis on deadlock. Support Vector Machine was used to predict the result of matches using only the back '4' information of the team. The result was evaluated by predicting the whole of 2018/2019 season of the English Premier League. The prediction results give a 76.8% accuracy and error rate of 0.23.

1. Introduction

Team back '4' includes the centre-back, sweeper, full-back, and wing-back. The centre-back and full-back positions are essential in most modern formations. The sweeper and wing-back roles are more specialized for certain formations. Furthermore, in team back '4', a centre-back (also known as a central defender or centre-half) defends in the area directly in front of the goal, and tries to prevent opposing players, particularly centre forward from scoring. Centre-backs accomplish this by blocking shots, tackling, intercepting passes, contesting headers and marking forwards to discourage the opposing team from passing to them. There are two main defensive strategies used by centre-backs: the zonal defence, where each centre-back covers a specific area of the pitch; and man-to-man marking, where each centre-back has the job of tracking a particular opposition player. In the modern game, most teams employ two or three centre-backs in front of the goal keeper. The 4-2-3-1, 4-3-3, and 4-4-2 formations all uses two centre-backs.

The sweeper (or *libero*) is a more versatile centre-back who "sweeps up" the ball if an opponent manages to breach the defensive line. This position is rather more fluid than that of other defenders who man-mark their designated opponents. Because of this, it is sometimes referred to as libero. A recent and successful use of the sweeper was made by Otto Rehhagel, Greece's manager, during UEFA Euro, 2014. Rehhagel utilized Traianos Dellas as Greece's sweeper to great success, as Greece surprisingly became European champions. Sweepers also uses 5-3-2 formation, though sweepers may be expected to build counter-attacking moves, and as such require better ball control

and passing ability than typical centre-backs, their talents are often confined to the defensive realm. For example, the catenaccio system of play, used in Italian football in the 1960s, employed a purely defensive sweeper who "roamed" only around the back line.
The full-backs (the left-back and the right-back) take up the holding wide positions and traditionally stayed in defence at all times, until a set-piece. There is one full-back on each side of the field except in defense with fewer than four players, where there may be no full-backs and instead only Centre-backs. The wing-back is a variation on the full-back, but with heavier emphasis on attack. This type of defender focuses more heavily on attack than defence, yet they must have the ability, when needed, to fall back and mark opposing players to lessen the threat of conceding a goal-scoring opportunity.
Statistical models for football predictions started appearing from the 90s, but the first model was proposed much earlier by Moroney, who published his first statistical analysis of soccer match results in 1956 [1]. According to his analysis, both Poisson and negative binomial distribution provided an adequate fit to results of football games. The series of ball passing between players during football matches was successfully analyzed using negative binomial distribution by [2]. The method was improved in 1971, and [3] indicated that soccer game results are to some degree predictable and not simply a matter of chance.
[4] created a methodology in order to analyze player's movement and work rate within the different playing positions/roles in a first division football team. A total of 51 (both home and away) competitive games were analyzed over a course of a season. Player’s movements were subdivided into several distinct movement classifications. A tactical evaluation of soccer was also undertaken by [5] who performed a computerized notational analysis of 8 games in the 1994 World Cup Asian Qualifying. The respective playing patterns of the teams were analyzed, with a particular emphasis upon the Japanese national team. The analysis assessed 32 actions of players in relation to an 18 cell division of the pitch.

On Oct. 26, 1863, representatives from a group of clubs met at the Freemason's Tavern in England to draw up the first official rules. The rules were accepted on December 18, 1863. Rule 6 stated “when a player has kicked the ball any one of the same sides who is nearer to the opponent’s goal line is out of play” (offside!). The lifeblood of the early games was the skill of dribbling. The forward pass was banned. Rule 6 was changed in 1866 to permit advanced players to receive a pass, providing there were at least three opponents between themselves and the goal line. The first international match saw Scotland played a 1-2-7 and the Scots a 2-2-6. It was the Scots who realized the potential of the 1866 rule change and began to employ the short pass. Despite the large of forwards in the game, the result was a 0 – 0 ties. By 1890 the favoured system of play in England had evolved to the 2-3-5 formation. A pyramid shape was described, as one would draw lines from the two wings on either side of the field back to the goalkeeper.

[6] developed a specialized computer analysis system, to undertake a comprehensive technical evaluation of performance. Two (2) distinct levels of performance, the 1990 FIFA World Cup and the 1990 World Collegiate Soccer Championships were analyzed using 38 key events entered in real time by a trained analyst. From the results it can be inferred that collegiate coaches must be selective when selecting World Cup teams as an appropriate model of performance as many differences do occur, which makes any comparison invalid. [7] undertook a technical analysis of playing positions within elite level international soccer at the European Championships 2004. The qualitative data were gathered, post event, based on the relative successful execution of techniques performed. Players were classified by position as goalkeepers, defenders, midfielders or strikers. A comparison was also made between the technical distributions of both a successful and unsuccessful team.

According to [8], improving the process of constructing operational definitions within performance analysis and using a standard set of definitions for a sport would enhance the quality of data sets and promote future research and analysis. Billy Bean [9] defined these processes and definitions for baseball and used them, with large objective databases, to recruit players more efficiently and economically, and hence achieve success far in excess of the expectation of his club's financial standing.

The first model predicting outcomes of football matches between teams with different skills was proposed by [10]. According to his model, the goals, which the opponents score during the game, are drawn from the Poisson distribution. The model parameters are defined by the difference between attacking and defensive skills, adjusted by the home field advantage factor. The methods for modeling the home field advantage factor were summarized in an article by [11]. Time-dependency of team strengths was analyzed by [12]. He used recursive Bayesian estimation to rate football teams and this method was more realistic in comparison to soccer prediction based on common average statistics.

[13] and [14] have further extended these Poisson models by incorporating a large set of potential influence variables as well as team-specific (either random or fixed) ability parameters. By using different regularization technique, they discovered a sparse set of relevant covariates, which were then used to predict the European championship (EURO) 2012 and FIFA World Cup 2014 winners, respectively.
In the present work, we use Support Vector Machine for ordinal response (Draw). Different techniques have been used to develop result prediction systems. In particular, football match result prediction systems have been developed with techniques such as artificial neural networks, Poisson distribution [15], [16], [73], support vector machine [17], naive Bayesian system [18], [80], random forest [19], k-nearest neighbor algorithms (k-nn), and others. The choice of any technique depends on the application domain as well as the feature sets. The priority of a system developer or designer in most cases is to obtain a high prediction accuracy and playing position cannot be relegated in soccer matches, as placing a player in the best position possible is an advantage to the team [20], [21], [22], [78].

1.1.Developmental Trends of Soccer Game

According to [23], the game of soccer can be described as a game involving two teams, each at playing from one end of the field towards the opponent's side. Each team is made up of eleven (11) players with some number of reserve players, each playing/defending from opposite side of the field with a goal post at both ends. It can also be referred to as a 'round-leather' game consisting of two teams vying for victory over each other. A team is said to have scored the other when she succeeded in kicking the ball across the goal line of the other opponent. It can be regarded as a game that does not only requires technicality but skills and has produced the likes of the legend Pele of Brazil to the great Abedi Pele of Ghana, the magical Diego Armando Maradona of Argentina to the skilful Austin Okocha of Nigeria [23].
Soccer is the most played sport around the world. According to FIFA's most recent Big Count survey, 265 million people actively play soccer around the world [24], [25]. This figure accounts for about four percent of the world's population and does not include the number of people who play actively without organized competition. There are an estimated 108 professional soccer leagues located in 71 countries around the world. Some of the best players in the world came together to compete in England. England has a system of leagues bound together with promotion and relegation, with one league reigning at the top of the English Premier League [26]. The English Premier League consists of twenty clubs. At the end of every season the bottom three teams are relegated to the first

division and three teams are promoted from the first division to the premier league. In a soccer match, one move can determine the result of the game, making it very challenging to predict the outcome [27]. While these statistics can help determine the result, only one result matters in the end-whether the match results in a win, lose or draw.
English league is the most watched soccer league all over the world having more than 2 billion fans [29], broadcast in 212 territories to 643 million homes and a potential TV audience of 4.7 billion people. In the 2014/2015 season, the average Premier League match attendance exceeded 36,000, second highest of any professional football league behind the Bundesliga's 43,500. Most stadium occupancies are near capacity. The Premier League ranks third in the UEFA coefficients of leagues based on performances in European competitions over the past five seasons. [29] found that League One soccer teams were 80% less likely to win playing away than playing at home. Similarly, team and opposition quality have been found to have an important influence on performance [30], [31], [32], [33], [34], [35], [36], [81].
Team quality has often been categorized using the previous season’s final league position with teams then categorized as strong, weak, top 3, bottom 3, etc. and has been shown to influence match difficulty in rugby union [37], [38]. However, [39] suggested that this method could be considered arbitrary or unfair as teams could, for example, miss being classified as a strong team by just a few points, despite having been in the top three for the majority of the season.

1.2.Different Approaches to Predicting Outcome of Soccer Game

1.2.1. Predicting football scores via Poisson regression model

Poisson regression model was proposed in order to forecast football match outcomes. The proposed methodology was applied to two national competitions: the 2012/2013 English Premier League and the 2015 Brazilian Football League [40], [41], [42]. [40] proposed a Poisson regression model considering control variables, which consist of the rating for each team and the match venue given by the Federation International of Football Association (FIFA). The authors used their results and other results about the quality of forecasts to simulate the 1998 FIFA World Cup. The number of goals scored by each team in a match is assumed to follow Poisson distribution, whose average reflects the strength of the attack, defense and the home team advantage. Inferences about all unknown quantities involved are made using a Bayesian approach. Probabilities were calculated for a win, draw and loss for each match using a simulation procedure. Besides, also using simulation, the probability of a team qualifying for continental tournaments, being crowned champion or relegated to the second division was obtained.
[42] proposed a Bayesian approach to predict the outcomes of matches using specialists' opinions and FIFA rankings to build a Power prior. Using simulations, the authors calculated the probabilities of wins, draws, losses and odds of the teams being ranked in the group stage are obtained. [43] developed a Bayesian methodology for the Poisson-gamma model in which the priors are chosen considering historical and recent information. The authors calculated the probabilities of win, draw and loss for the 2010 FIFA World Cup games. [44] considered the Poisson distribution for the number of goals scored by England, Ireland, Scotland and Wales in the British International Championship (1883-1980).

1.2.2. Predicting Premiership Football Match Results using CRISP model

[45] used CRISP-DM (CRoss-Industry Standard Process for Data Mining) methodology is implemented to create and discover a variety of predictive classification models in an attempt to accurately and objectively predict the outcome of English Premier League football matches using freely available online data that is based only on previous Premier League match results from 2002 to 2012. Using an iterative approach, numerous data visualization, pre-processing and

transformation techniques have been applied and evaluated to determine the most effective predictive model, which has been assessed by its ability to correctly forecast the match outcome of the final 50 matches in the 2011/2012 Premier League season as per the initial business and data mining objectives. The result of the study showed that a number of models outperform human expert match predictions.

1.2.2. Predicting the championship results by the diverse models

Normally we expect a better classification score by the fuzzy-inference model due to the stochastic nature of many of the outcomes. However, the genetic programming approach is proved capable of achieving a high rate of classification as compared with the other models. An explanation for this may be that the genetic programming model was trained by using as fitness measure directly a pick-winner value. It is still characterized though, by comparatively heavy requirements in time and computing resources. Nevertheless, the solution proposed by a genetic programming procedure is more understandable for humans than a neural network's configuration and comparable to that of the fuzzy system. The predicting ability of models for picking winners of serial games, is based on the example of the Ninth Ukrainian Football Championship (2000-2001). Source information for the prediction is the set of results of the eight previous championships and the results of the first five tours in the current championship. The task is the prediction of the results of the remaining sixth, seventh, and twenty-sixth tours. The results of the prediction show that both the neural network and the fuzzy system were also proved capable of capturing most of the underlying trends and predicting with a high rate of success very close to the reality final rank and accumulated points.

1.2.3. Predicting soccer outcome with machine learning based on weather condition

Massive amounts of research have been doing on predicting soccer matches using machine learning algorithms. Unfortunately, there are no prior researches used weather condition as features. In the research by [46], three different classification algorithms were study for predicting the outcomes of soccer matches by using temperature difference, rain precipitation, and several other historical match statistics as features. The dataset consists of statistic information of soccer matches in La Liga and Segunda division from season 2013-2014 to 2016-2017 and weather information in every host city. The results show that the SVM model has better accuracy score for predicting the full-time result compare to KNN and RF with 45.32% for temperature difference below 5° and 49.51% for temperature difference above 5°. For over/under 2.5 goals, SVM also has better accuracy with 53.07% for rain precipitation below 5 mm and 56% for rain precipitation above 5 mm.

1.2.4. Football Match Statistics Prediction using Artificial Neural Networks

The predictions of the outcomes of football (American soccer) matches are widely done using the if-else case-based Football Result Expert System (FRES). The proposed prediction technique by [77] uses a neural network approach to predict the results of football matches. The neural network detects patterns from a number of factors affecting the outcome of a match making use of historical cases of training. The approach described in this paper takes advantage of the fact that neural networks are good at recognizing patterns and mapping these patterns to outputs. The approach proposed in this paper uses a number of factors, some which may even not be available to a standard human decision maker as inputs to the neural network.

1.3. Related Work

[47] found that self-efficacy was also positively correlated with the performance of soccer players participating in the Amputee World Cup. They indicated that psychological skills, such as activation and relaxation ability increased the self-efficacy levels of participants, resulting in improved performance.

The work of [7] was innovative in two ways; by analyzing at the exact technical requirements of each position and also due to the fact that it used qualitative data within a quantitative system. The study was aimed to analyze every individual's technical ability that competes in the European Football Championships of 2004. This measure was based on a subjectively drawn continuum that analyses a player's technical movement throughout the game. First approaches to account for possible dependencies between the scores by using adjusted Poisson models are proposed by. Alternatively, the bivariate Poisson distribution allows to explicitly model (positive) dependence within the Poisson framework.

[48] evaluated whether players in different positional roles have a different physical and physiologic profile. For the purpose of this study, physiologic measurements were taken of 270 soccer players during the precompetitive period of 2005/06 and the precompetitive period of 2006/07. According to the positional roles, players were categorized as defenders (n = 80), midfielders (n = 80), attackers (n = 80), and goalkeepers (n = 30).

[49] worked on reliability and validity of match performance analysis in soccer, using statistical analysis. The main aim of this research work was to develop, test the reality of, and apply a new method for team match performance analysis.
[50] implemented structural equation modeling to examine the linkages between financial performance, sporting performance and stock market performance for English football clubs over the period from 1995 to 2007. The results indicate that there is a strong correlation between financial and sporting latent constructs. Additionally, the study indicates that the sports managers seek to achieve a minimum level of profit and maximize sporting performance. This situation remains even when the club is owned by a group of investors. On the other hand, the confirmatory factor analysis and regression analysis show that financial and sporting factor scores are statistically correlated with stock returns, but not with risk.

[76] identified specific performance indicators that discriminate the top clubs from the others based on significantly different pitch action performance in the Spanish Soccer League. All 380 games corresponding to the 2008/2009 season were analyzed in this study. The studied variables were divided into three groups related to goals scored (goals for, goals against, total shots, shots on goal, shooting accuracy, shots for a goal), offense (assists, crosses, off sides committed, fouls received, corners, ball possession) and defense (crosses against, off sides received, fouls committed, corners against, yellow cards, red cards).

Extensive research has focused upon quantifying the physical demands of soccer competition [51], [52]. Despite constituting only 12% of distance travelled, high-speed locomotion has received significant attention when analyzing competitive performance [53], [54].

[55] established a comprehensive overview of the demands of futsal by a time-motion analysis on eight Australian National Team players and ten state League Team players over four futsal matches. The study analyzed six locomotor activity categories, focusing on total distance covered, total frequency of activities effort distance and effort duration.

[56] compared match performance in professional soccer players across two major European championships including Spanish La Liga and English FA Premier League (FAPL). Data were collected using a computerized match analysis system. A total of 5938 analyses were recorded during the 2006/2007 season. The players were classified into six positional roles: central defenders, full backs, central defensive midfielders, wide midfielders, central attacking midfielders, and forwards. The match performance variables analyzed included physical activity like total distance covered, distances covered at high-intensities both with and without possession of the ball; technical actions like heading and ground duels, passing, time in possession, and ball touches. Comparison of the total distance covered by FAPL and La Liga players showed no difference across individual playing positions but FAPL players generally covered greater distances in sprinting.

In [57], the study's focus was two folds, which are to characterize repeated high-intensity movement activity profiles of a professional soccer team in official match-play; and to inform and verify the construct validity of tests commonly used to determine repeated-sprint ability in soccer by investigating the relationship between the results from a test of repeated-sprint ability and repeated high-intensity performance in competition. High-intensity running performance was measured in 20 players using computerized time-motion analysis.

[58] investigated sport specific cognitive traits, while few studies have focused on general cognitive traits. The study examined if measures of general executive functions can predict the success of a soccer player. The present study used standardized neuropsychological assessment tools assessing players' general executive functions including on-line multi-processing such as creativity, response inhibition, and cognitive flexibility. In a first cross-sectional part of the study we compared the results between High Division players (HD), Lower Division players (LD) and a standardized norm group. The result shows that both HD and LD players had significantly better measures of executive functions in comparison to the norm group for both men and women. Moreover, the HD players outperformed the LD players in these tests. In the second prospective part of the study, a partial correlation test showed a significant correlation between the result from the executive test and the numbers of goals and assists the players had scored two seasons later. The results from this study strongly suggest that results in cognitive function tests predict the success of ball sport players.

[59] developed a model that provides an overview and comparison of predictive capabilities of several methods for ranking association football teams. The main benchmark used is the official FIFA ranking for national teams. The ranking points of teams are turned into predictions that are next evaluated based on their accuracy. This enables us to determine which ranking method is more accurate. The best performing algorithm is a version of the famous Elo rating system that originates from chess player ratings, but several other methods (and method versions) provide better predictive performance than the official ranking method.

[60] proposed a paper about the financial performance of football clubs in Kenya, considering the case of Kenyan Premier League. The study had one objective to achieve which was to analyze the financial performance of football clubs in Kenya. The research design was a descriptive study. Data was collected using financial statements and other sources of financial data of the selected sixteen Kenyan premier league clubs. Descriptive, correlation coefficients and regression analysis were used to analyze the financial performance. The findings were presented in tables. It was also clear that there was a significant relationship between the financial performances of the football clubs in the premier League the financial performance variables represented by R2 value of 0.935 which translates to 93.5%. The study confirmed that the profitability of the Kenyan Premier League clubs is majorly affected by return on assets and the liquidity of football clubs in the league. The analysis

further suggest that clubs should have strong financial performance policies in using its total assets effectively and finding ways of sustaining high liquidity ratios to guarantee meeting of their obligations in the unforeseeable future. The study recommends that football clubs need to ensure effective utilization of clubs' assets to generate returns.

[78] developed a model to investigate the psychological skills of African youth soccer players in different playing positions. The role of psychological skills and overall team performance was also determined. The sample consisted of male soccer players (N=152) between the ages of 14 and 18 years from ten African countries competing in the 2010 Copa Coca-Cola soccer tournament. A cross-sectional survey design was used to determine the players' psychological skills by means of the Bull's Mental Skills Questionnaire and the Athletic Coping Skills Inventory-28 (ACSI-28). Results yielded insignificant differences between the subscale scores of the players in different playing positions. Concentration was the only psychological variable associated with performance. The middle four-ranked teams outscored the most successful and least successful teams in relaxation. Findings from this study could not confirm the widely acclaimed research assumption that psychological skill demands differ among players in different playing positions, nor the positive correlation between psychological skills and team success. Future research should investigate the perceptions and extent of psychological skills training among African youth soccer players, as well as the efficiency of psychological skills interventions aimed at enhancing overall team performance.

[61] proposed Soccer Match Result prediction system. The research work was carried out 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 whereby 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.

[23] adopted object-oriented score-time-based model. The Unified Modelling Language (UML) was used. An Object Oriented approach was adopted for the implementation of the proposed model with Java programming technique. The paper has 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).

[26] presented an Adaptive Neuro-Fuzzy Inference System (ANFIS) approach for match prediction which has served as the focal aim of their research using seven premier league teams and nine linguistic values. Matric Laboratory (MATLAB) 7.0 served as the tool of implementation highlighting various views. The ANFIS training was successful completed at epoch 2 and having an error of 1.41237e-006. The model was further used to predict the outcome of 7 matches with a successful rate of 71%.

[62] soccer research has traditionally focused on technical and tactical aspects of team play, but few studies have analyzed motor skills in individual actions, such as goal scoring. The objective of the study was to investigate how Lionel Messi, one of the world's top soccer players, uses his motor skills and laterality in individual attacking actions resulting in a goal. It analyzed 103 goals scored by Messi between over a decade in three competitions: La Liga (n=74), Copa del Rey(n=8), and the UEFA Champions League(n=21). The study used an ad hoc observation instrument (OSMOS soccer player) comprising 10 criteria and 50 categories; polar coordinate analysis, a powerful data reduction technique, revealed significant associations for body part and orientation, foot contact zone, turn direction, and locomotion. No significant associations were observed for pitch area or interaction with opponents. The analysis confirms significant associations between different aspects of motor skill use by Messi immediately before scoring, namely use of lower limbs, foot contact zones, turn direction, use of wings, and orientation of body to move toward the goal. Studies of motor skills in soccer could shed light on the qualities that make certain players unique.

[63] developed conceptual model that describes the relationships between cohesion and performance including the antecedents and consequences of the variables. The aim of the study is to examine the relationship between cohesion and performance for soccer teams during a full competitive season as well as the direction of the relationship. The study suggests a new model for the relationship between cohesion and performance. In total, 173 Greek soccer players (M= 21.91) completed the Group Environment Questionnaire [63] in all the measurements from the beginning of the preparation to the end of the competitive season. The results showed that cohesion and performance are two variables that affect each other in soccer, with a stronger direction from cohesion to performance. However, cohesion affects performance either positively or negatively throughout the season. All statistical analyses were performed using the SPSS package. The techniques employed were descriptive statistics for all measurements of cohesion, correlations between the variables and hierarchical multiple regression analyses to identify the relationship between cohesion and performance.

[64] proposed Functional Movement Screen Scores and Physical Performance among Youth Elite Soccer Players. The study had two main objectives which to determine if differences in Functional Movement Screen (FMS) scores exist between two levels of competition and also to analyze the association between FMS individual and overall scores and physical performance variables of lower-limb power (jumps), repeated sprint ability and shot speed. Twenty-two under 16 (U16) and twenty-six under 19 (U19) national competitive soccer players participated in this study. All participants were evaluated according to anthropometrics, FMS, jump performance, instep kick speed and anaerobic performance. There were no significant differences in the individual FMS scores between competitive levels. There were significant negative correlations between hurdle step (right) and Running-based Anaerobic Sprint Test (RAST) power average ($\rho = -0.293$; $p = 0.043$) and RAST fatigue index (RAST Fat Index) ($\rho = -0.340$; $p = 0.018$). The hurdle step (left) had a significant negative correlation to squat jump (SJ) ($\rho = -0.369$; $p = 0.012$). Rotary stability had a significant negative correlation to RAST fatigue index (Right: $\rho = -0.311$; $p = 0.032$. Left: $\rho = -0.400$; $p = 0.005$). The results suggest that individual FMS scores may be better discriminants of performance than FMS total score and established minimal association between FMS scores and physical variables. Based on that, FMS may be suitable for the purposes of determining physical function but not for discriminating physical performance.

[65] developed a model to assess the technical skills of both male and female players according to their position. Twenty-Seven female ($M = 12.52 \pm .51$) and thirty seven male soccer players ($M = 12.46 \pm .51$) who were members of amateur youth leagues participated in the study. Players were classified according to their playing position [72] into the following groups: central defenders (CD),

fullbacks (FB), midfielders (M), wingers (W), forwards (F). Shooting, short passing, long passing, dribbling, and dribbling after passing abilities were aspects assessed in the research. Although there were not significant differences between players of various positions the central defenders as well as the female players performed significantly lower scores in most of the technical skills.

In [66] it has been shown by the help of gradient boosting techniques that (at least for their setting of EURO data) no additional modeling of the covariance structure is necessary: for suitably designed linear predictors, which are based on highly informative covariates, two (conditionally) independent poisson distributions are adequate. The aim of the study was to pursue a different approach and investigate an alternative tool for the prediction of the outcomes of football matches, namely random (decision) forests – an ensemble learning method for classification, regression.

[67] developed the identity in technical demands of different playing positions controlling the effects of situational variables in the UEFA Champions League from a long-term perspective. Data of 18 technical performance-related match actions and events achieved by 1,990 out-field players (classified into five positional roles) in 1,000 matches played in the UEFA Champions League from season 2009/2010 to 2016/2017 were collected. Generalized mixed linear modeling was employed taking the value of each of the 18 technical performance-related variables as the dependent variable, and taking the classified 5 positions as the predictor variable. Uncertainty in the true effects of the predictors was evaluated using non-clinical magnitude-based inference. Results showed that the differences between central defenders and forwards were biggest while central defenders and full backs presented the smallest difference. The performances of midfielders in variables related to passing and organizing were worse than expected and wide midfielders showed relative better performances than central midfielders in passing and organizing. Meanwhile, defenders, especially central defenders, achieved good performance in variables related to passing and organizing. Forwards played an important role in the aspects of goal scoring and organizing, they also participated in the initial defending process.

[68] and [74] analyze performance differences of football players two years prior and the year after signing a new contract (the following year) while taking playing position, nationality, player's role, team ability, and age into account. The sample was comprised of 249 players (n = 109 defenders, n = 113 midfielders; and n = 27 forwards) from four of the major European Leagues (Bundesliga, English FA Premier League, League 1, and La Liga) during the seasons 2008 to 2015. The dependent variables studied were: shooting accuracy, defense (the sum of defensive actions, tackles, blocks, and interceptions), yellow cards, red cards, passing accuracy, tackle success, and minutes played per match. Two-step cluster analysis allowed classifying the sample into three groups of defenders (national important, foreign important and less important players) and four groups of midfielders and forwards (national important, foreign important, national less important, and foreign less important players). Magnitude Based Inference (MBI) was used to test the differences between player’s performances during the years of analyses. The main results (very likely and most likely effects) showed better performance in the year prior to signing a new contract than the previous year for foreign important defenders (decreased number of red cards), national important midfielders (increased number of minutes played), foreign important forwards (increased minutes played and defense), and national important forwards (increased minutes played). In addition, performance was lower the year after signing the contract compared to the previous one for less important defenders (decreasing defense), national less important midfielders (decreased minutes played), and foreign less important forwards (decreased defense) [75].

[69] proposed the problem of evaluating the performance of soccer players in attracting the interest of many companies and the scientific community, thanks to the availability of massive data

capturing all the events generated during a match (e.g., tackles, passes, shots, etc.). Unfortunately, there is no consolidated and widely accepted metric for measuring performance quality in all of its facets. The paper designed and implemented PlayeRank, a data-driven framework that offers a principled multi-dimensional and role-aware evaluation of the performance of soccer players. A framework was built by deploying a massive dataset of soccer-logs and consisting of millions of match events pertaining to four seasons of 18 prominent soccer competitions. By comparing PlayeRank to known algorithms for performance evaluation in soccer, and by exploiting a dataset of players' evaluations made by professional soccer scouts, it was shown that PlayeRank significantly outperforms the competitors.

[70] examined the extent to which fuzzy logic can be used to predict soccer match results. Fuzzy Logic Gaussian Membership function (Gmf) technique was adopted for the computation of membership grade for each input parameter of twenty football teams in English Premier League of 2017/2018 session. It was implemented using MATLAB. From the result produced in the model experiment, an average testing error of 0.075 with improved prediction accuracy of 89.27% was achieved.
As stated by [71], the ability to provide information about individual's technical performance and the profiles of such players can significantly modify playing behaviour and promote successful performance. Information about technical performance is also much preferable to cursory comments made by coaches following competition [79]. The recording of events in some coded form can help such coach observations to be formed, especially by defining each skill performed as successful or unsuccessful [79]. If an accurate analysis of the technical attributes of each player's position could be established, then the results could significantly influence team selections and coaching sessions [7], and tactical decisions that are to be made by coaching teams, before and during matches. As vividly and clearly seen from all presently existing literatures, none as consider using only an information of a section of the playing position especially that of the back 4 to determine the outcome of soccer matches.

Match prediction model has become increasingly popular in the last few years and many different approaches of prediction model have been proposed. Prediction in football has become a fascinating research problem, because there are many factors that can determine match outcome, such as teamwork, skills, weather, and home advantage among many others. The challenge is also faced by the experts in football as well because it is very difficult to anticipate the actual results of football matches. As much as match win is highly desirable, sharing a point is also embraced especially when playing away or when a more tactical team come visiting. One compartment of the team that assures game draw is the defensive line. Literature survey shows that no much emphasis was laid on this in most of the predictive analytic research works done. This paper therefore consider it expedite to forecast match draw by estimating the strength of teams' back '4'. The study deploys the use of Support Vector Machine (SVM) to actualize its goal.

2.0. Methodology

This study adopts machine learning approach using Support Vector Machine. For successful execution of this study, the following steps as shown in Figure 1 were followed:

i. **Time-series data:** The time-series data compose a sequence of data that is collected at regular intervals over a period of time. In this case, it is a set of data built from football match history. Players' performance and manager indices were gathered from the 2015-2018 season of the English Premier League.
ii. **Data pre-processing:** Two data pre-processing features are introduced to the system; namely, "replace missing value" and "normalization". "Replace missing value" is used to replace missing

values since SVM does not support missing value. A precise imputed missing value data-cleansing operator is used to execute this operation.

iii. **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.

iv. **Model building using SVM:** This model, is a non-probabilistic binary linear classifier used to train the data sets for the model.

v. **Predictive model:** This model describes the use of Machine Learning Approach using Support Vector machine in predictive model.

From the study methodology, the various stages to be involved in prediction are shown in the conceptual model represented in Figure 2.

Figure 1: Study Methodology

2.2. System Architecture

The Architecture of the model as clearly shown in Figure 3, consist of the following under listed components/parameters:

i. Data Collection
ii. Data Preprocessing
iii. Data Splitting
iv. Testing Model
v. Model Building Classification (SVM)
vi. Parameter Optimization
vii. Prediction
viii. Evaluate Model

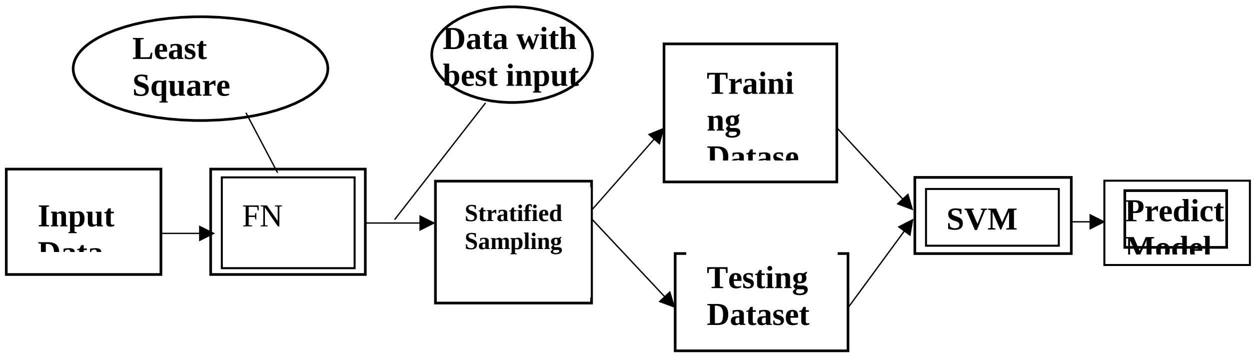


Figure 2: Conceptual Model

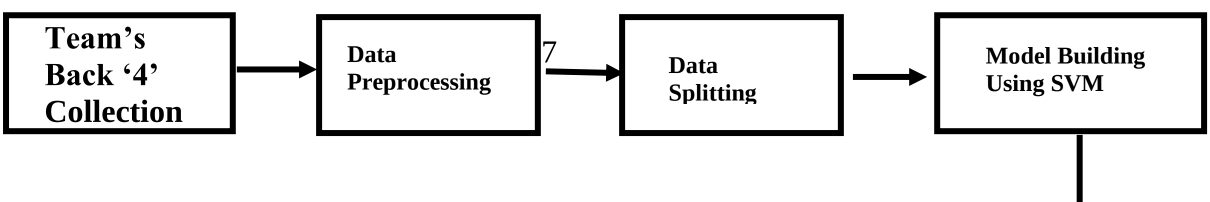


Figure 3: System Architecture

Table 1: Data Source

S/N	Data Source	Source Address	Date Accessed
1.	English premier league	https://www.premierleague.com	25/11/2019 - 18/12/2019
2.	Sofa score	www.Sofascore.com	20/12/2019 – 27/12/2019
3.	Fifa Index	https//www.fifaindex.com	29/11/2019 - 10/01/2020

2.3. Data Collection

Team's Back '4'performance data of English premier league clubs were used. The sample of the study comprised data from all football clubs participating in the English Premier League. Then data was collected from different sources https://www.premierleague.com, https://www.sofascore.com, https://www.fifaindex.com as seen in Table 1. Manually by inputting data 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 English Premier League as shown in Table 2. The data were periodically extracted between 25th November 2019 and 10th January 2020. Figure 4 shows the flow of the data extraction process.

The predictive attributes are enumerated below:

i. **Attacking**: It is the movement of the players and the ball for the team who has possession making a forceful attempt to score or gain an advantage.
ii. **Technicality:** It 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:** This 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:** It requires players to know themselves, work with each other, and understand the very best ways in which they can combine.
v. **Defending:** This is moving quickly to the player with the ball and preventing goal.

vi. **Marking:** This is defensive strategy where defenders are assigned a specific opposition player to mark rather than covering an area of the pitch.

Table 2: Screenshot of Sample of the Dataset

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Sample Id	**Teams**	**Back '4' Names**	**Attacking**	**Creativity**	**Defending**	**Tactical**	**Marking**	**Slide Tackle**	**Jumping**	**Stamina**	**Stand Tackle**	**Technical**	**mean**	**Label**
1	Man Utd	Victor Lindelof	37	43	72	63	82	79	83	80	76	50	66.5	1
2	Man Utd	Luke Shaw	46	59	69	55	80	79	82	78	74	55	67.7	1
3	Man Utd	Chris Smalling	35	40	87	84	80	81	82	76	71	47	68.3	1
4	Man Utd	Metteo Darmian	40	37	48	42	78	80	82	71	77	69	62.4	1
5	Man Utd	Phil Jones	36	46	81	80	75	80	78	81	64	51	67.2	1
6	Man Utd	Ashley young	51	66	60	67	72	74	76	51	71	57	64.5	1
7	Man Utd	Marcus Rojo	30	38	43	30	72	79	78	81	64	40	55.5	0
8	Man Utd	Dalay Blind	52	72	68	65	83	81	84	77	75	58	71.5	1
9	Everton	Leighton Braines	42	46	60	41	75	76	77	67	55	58	59.7	0
10	Everton	Micheal Keane	77	51	23	49	78	80	82	71	77	65	65.3	1
11	Everton	Mason Holgate	43	45	67	59	75	77	78	75	67	49	63.5	1
12	Everton	Seamus Coleman	46	55	63	54	78	80	82	71	77	57	66.3	1
13	Eveton	Kurt Zouma	41	43	72	68	75	79	79	88	74	51	67	1
14	Everton	Lucas Digne	58	77	71	65	78	84	83	92	85	61	75.4	1
15	Burnmouth	Adam Smith	42	47	61	53	76	76	77	74	64	50	62	1
16	Burnmouth	Steve Cook	40	45	78	74	77	77	78	68	75	48	66	1
17	Burnmouth	Nathan Ake	43	45	80	76	82	81	82	81	82	50	70.2	1
18	Burnmouth	Charlie Daniels	43	52	49	41	72	75	76	68	75	46	59.7	0
19	Hudderfield	Chris Lowes	50	57	41	34	73	71	70	79	85	49	60.9	1
20	Hudderfield	Terence Kongolo	39	42	73	61	68	73	76	72	68	42	61.4	1
21	Hudderfield	Christopher Schind	34	42	75	69	74	73	76	72	68	48	63.1	1

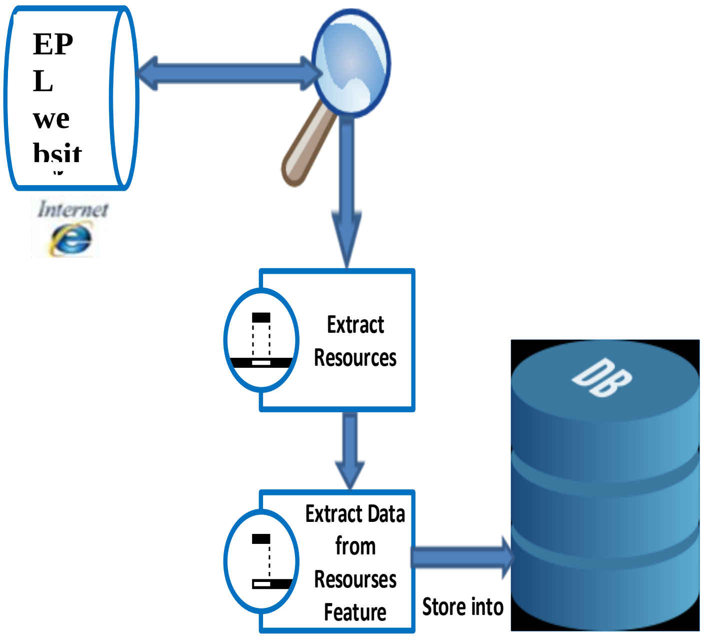


Figure 4: Data Extraction Process

vii. **Side Tackle:** This is a tackle in association football. It is completed with one leg extended to push the ball away from the opposing player. Sliding tackles can often be sources of controversy, particularly when players being tackled fall down over the tackler’s foot.

viii. **Stand Tackle:** This tackle is used to regain possession of the ball for your team and are sometimes seen as critical particularly in defensive positions when the opponents are in an offensively advantageous position.

ix. **Stamina:** This is the ability to sustain prolonged physical or mental effort.

x. **Jumping:** It is a pass in which a player leaps into the air and throws the ball to a teammate before returning to the ground.

2.4. Data pre-processing

Two data pre-processing features are introduced to the system; namely, “replace missing value” and “normalization”. “Replace missing value” is used to replace missing values since Support Vector Machine does not support missing value. A precise inputted missing value data-cleansing operator is used to execute this operation. It is a nested operator that always takes in data sets and returns a model. This operator calculatedly guesses missing values by learning models for each attribute (excluding the label) and applying those models to the data sets. Normalization is also applied to rescale feature values to fit in a precise range. Nominal-to-numerical operators have been used to transform non-numerical values to numeric as seen in Table 3.

Table 3: Data Pre-Processing Extraction Process

Attacking	Creativity	Defending	Tactical	Marking	Slide Tackle	Jumping	Stamina	Stand Tackle	Technical	mean	Label
37	43	72	63	82	79	83	80	76	50	66.5	1
46	59	69	55	80	79	82	78	74	55	67.7	1
35	40	87	84	80	81	82	76	71	47	68.3	1
40	37	48	42	78	80	82	71	77	69	62.4	1
36	46	81	80	75	80	78	81	64	51	67.2	1
51	66	60	67	72	74	76	51	71	57	64.5	1
30	38	43	30	72	79	78	81	64	40	55.5	0
52	72	68	65	83	81	84	77	75	58	71.5	1
42	46	60	41	75	76	77	67	55	58	59.7	0
77	51	23	49	78	80	82	71	77	65	65.3	1
43	45	67	59	75	77	78	75	67	49	63.5	1
46	55	63	54	78	80	82	71	77	57	66.3	1
41	43	72	68	75	79	79	88	74	51	67	1
58	77	71	65	78	84	83	92	85	61	75.4	1
42	47	61	53	76	76	77	74	64	50	62	1
40	45	78	74	77	77	78	68	75	48	66	1
43	45	80	76	82	81	82	81	82	50	70.2	1
43	52	49	41	72	75	76	68	75	46	59.7	0
50	57	41	34	73	71	70	79	85	49	60.9	1
39	42	73	61	68	73	76	72	68	42	61.4	1
34	42	75	69	74	73	76	72	68	48	63.1	1

Class Label: to get the relative attribute values we hand-code each response (1 and 0) and to get the class label for each sample we take the mean of each row (1, 2…10)

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

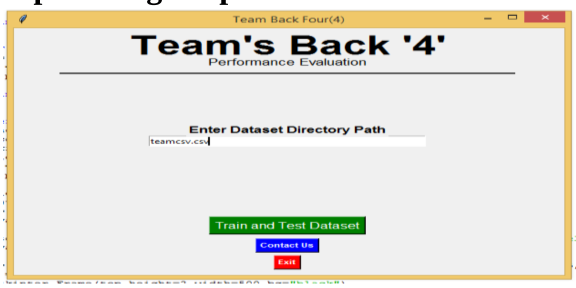
After finishing the building of our new set of crucial attributes, we split the data into training and testing data. 70% of the data were used for training while 30% were used for the testing with an epoch of 105.

2.6. Test Model: 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%.

2.7. Model Building Classification (SVM): This model, is a non-probabilistic binary linear classifier used to train the data sets for the model. A detailed description of the parameters used for this study is in the discussion of the results.

2.8. 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.
2.9. Prediction: The Level shall be predicted using support Vector Machine.
2.10. Evaluate Model: SVM model is used to evaluate the testing data using the classification accuracy evaluation measure.

Input Design/Input Interface



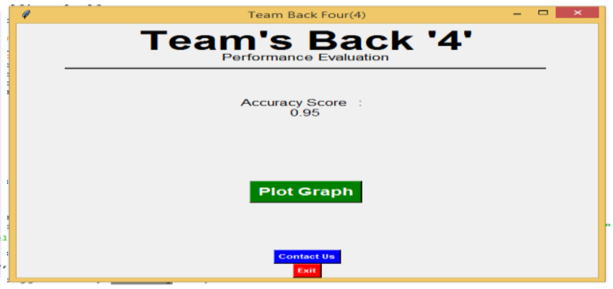
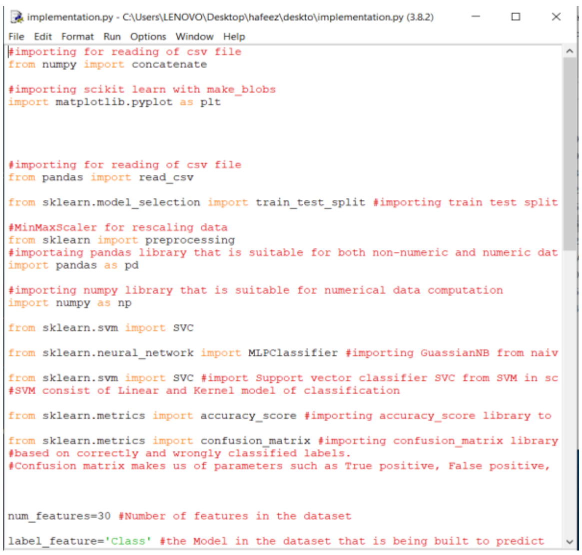


Figure 5: Team's Back '4'Home Page and Input Interface Figure 6: Team's Back '4'Accuracy Score Interface

Output Design/ Output Interface: Model Parameters



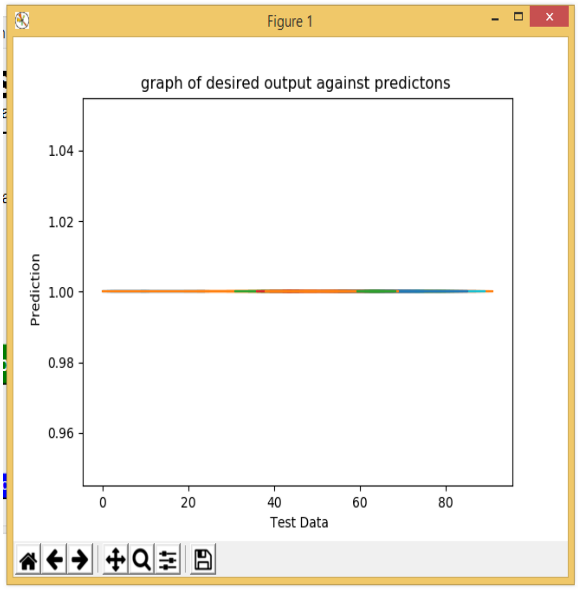


Figure 7: Graphical Interface

The model parameters are outlined below.

a. Number of input variable(s) = 10, Number of output variable(s) = 1
b. Approach Used = Support Vector Machine
c. Maximum Number of Linguistic variable(s) = 2 (Draw and Win/Loss)
d. Programming Language of Implementation: Python

3. Results and Discussion

The research work done by [49] concluded that an assessment of opponent interaction is critical to evaluate the effectiveness of playing tactics, and hence improve the validity of team match performance analysis. This research work was unable to cover area of EPL team performance relating to spending.

The implementation phase of this study shows the prediction of a football match Draw by Assessing Team's Back '4', which is carried out using machine learning algorithm SVM and JetBrains PyCharm, Python 3.8.1 IDLE. The implementation data was pre-processed from https://www.premierleague.com, https://www.sofascore.com, https://www.fifaindex.com designed with the credibility criteria and further prepared using Microsoft Excel Spreadsheet.

3.1. Dataset overview

The dataset contains Team's Back '4'players Attributes in English Premier League depending on specific criteria whose results can be either, 0= win/Loss, 1= Draw. Table 4 shows the classification sets/labels for the data.

Table 4: Classification set for the data

Support Vector Machine	**Classification Set**
Draw	1
Win/Loss	0

Table 5: Collected Data

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Teams	**Back_Names**	**Attacking**	**Creativity**	**Defending**	**Tactical**	**Marking**	**Slide Tackle**	**Stand Tackle**	**Jumping**	**Stamina**	**Technical**	**mean**	**Label**
Man Utd	Victor Lindelof	37	43	72	63	82	79	83	80	76	50	66.5	1
Man Utd	Luke Shaw	46	59	69	55	80	79	82	78	74	55	67.7	1
Man Utd	Chris Smalling	35	40	87	84	80	81	82	76	71	47	68.3	1
Man Utd	Metteo Darmian	40	37	48	42	78	80	82	71	77	69	62.4	1
Man Utd	Phil Jones	36	46	81	80	75	80	78	81	64	51	67.2	1
Man Utd	Ashley young	51	66	60	67	72	74	76	51	71	57	64.5	1
Man Utd	Marcus Rojo	30	38	43	30	72	79	78	81	64	40	55.5	0
Man Utd	Dalay Blind	52	72	68	65	83	81	84	77	75	58	71.5	1
Everton	Leighton Braines	42	46	60	41	75	76	77	67	55	58	59.7	0
Everton	Micheal Keane	77	51	23	49	78	80	82	71	77	65	65.3	1
Everton	Mason Holgate	43	45	67	59	75	77	78	75	67	49	63.5	1
Everton	Seamus Coleman	46	55	63	54	78	80	82	71	77	57	66.3	1
Eveton	Kurt Zouma	41	43	72	68	75	79	79	88	74	51	67	1
Everton	Lucas Digne	58	77	71	65	78	84	83	92	85	61	75.4	1
Burnmouth	Adam Smith	42	47	61	53	76	76	77	74	64	50	62	1
Burnmouth	Steve Cook	40	45	78	74	77	77	78	68	75	48	66	1
Burnmouth	Nathan Ake	43	45	80	76	82	81	82	81	82	50	70.2	1
Burnmouth	Charlie Daniels	43	52	49	41	72	75	76	68	75	46	59.7	0
Hudderfield	Chris Lowes	50	57	41	34	73	71	70	79	85	49	60.9	1
Hudderfield	Terence Kongolo	39	42	73	61	68	73	76	72	68	42	61.4	1
Hudderfield	Christopher Schind	34	42	75	69	74	73	76	72	68	48	63.1	1

3.2. Confusion Matrix: Each row of the confusion matrix represents the instances of an actual class and each column represents the instances of a predicted class.

```
 SVC score:  0.95
confusion_matrix:
 [[ 0  1]
 [ 0 19]]
```

Figure 8: Confusion Matrix

3.3. Data Training

Having Two (2) input variables (Draw= 1, Win/Loss= 0). The rule base was constructed to control the output variable, using = IF (Label >= 60, 1,0). 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%.

```
[[13.  94.  37.  ... 50.  66.5  1. ]
 [13.  59.  46.  ... 55.  67.7  1. ]
 [13.  19.  35.  ... 47.  68.3  1. ]
 ...
```

Figure 9: Training data

3.4. Data Testing

The data used for the testing stage is 30%, where the result obtained using Support Vector Machine approach gives 0.95% with an average error rate of 0.05, which shows that SVM gives a better and more accurate result.

```
 [ 3.  11.  39.  ... 44.  63.3  1. ]
 [ 3.  43.  41.  ... 67.  70.8  1. ]
 [ 3.  66.  44.  ... 55.  66.4  1. ]]
(99, 13)
(99,)
(20, 13)
(79, 13)
(20,)
(79,)
```

Figure 10: Testing data

Output

```
dtype: int64
           Teams  Back_Names  Attacking  ...  Technical
count  99.000000   99.000000  99.000000  ...  99.000000
mean   12.050505   49.000000  41.545455  ...  50.121212
std     7.383619   28.722813   7.438555  ...   7.137530
min     0.000000    0.000000  30.000000  ...  37.000000
25%     6.000000   24.500000  36.500000  ...  45.000000
50%    12.000000   49.000000  40.000000  ...  50.000000
75%    18.000000   73.500000  45.000000  ...  55.000000
max    25.000000   98.000000  77.000000  ...  69.000000

[8 rows x 14 columns]
[[13.  94.  37.  ... 50.  66.5  1. ]
 [13.  59.  46.  ... 55.  67.7  1. ]
 [13.  19.  35.  ... 47.  68.3  1. ]
 ...
 [ 3.  11.  39.  ... 44.  63.3  1. ]
 [ 3.  43.  41.  ... 67.  70.8  1. ]
 [ 3.  66.  44.  ... 55.  66.4  1. ]]
(99, 13)
(99,)
(20, 13)
(79, 13)
(20,)
(79,)


 SVC score:  0.95
confusion_matrix:
 [[ 0  1]
 [ 0 19]]
```

Figure 11: Result of the Dataset

It could be vividly seen that most match result if they are not draw, are very close to draw. Looking at Figure 10, the analysis of the 2018/2019 season results shows that the difference in goals per match is always very low and most of the time not more than two. About 81.57% of the matches account for a goal difference of not more than two goals. From the diagram in Figure 15, it is clearly seen that the charts are more clouded at the values close to zero.

The result of the study's evaluation as displayed in Figure 11 shows the prediction of different matches of 2018/2019 season in English Premier League, which justifies the abilities of team's back '4'home and away defenders, the data used for the training stage is 75%, the data used for the testing stage is 25%, where the result obtained using Support Vector Machine approach as seen in Figure 17 gives 0.768% with an average error rate of 0.23, which shows that SVM gives a better and more accurate result.

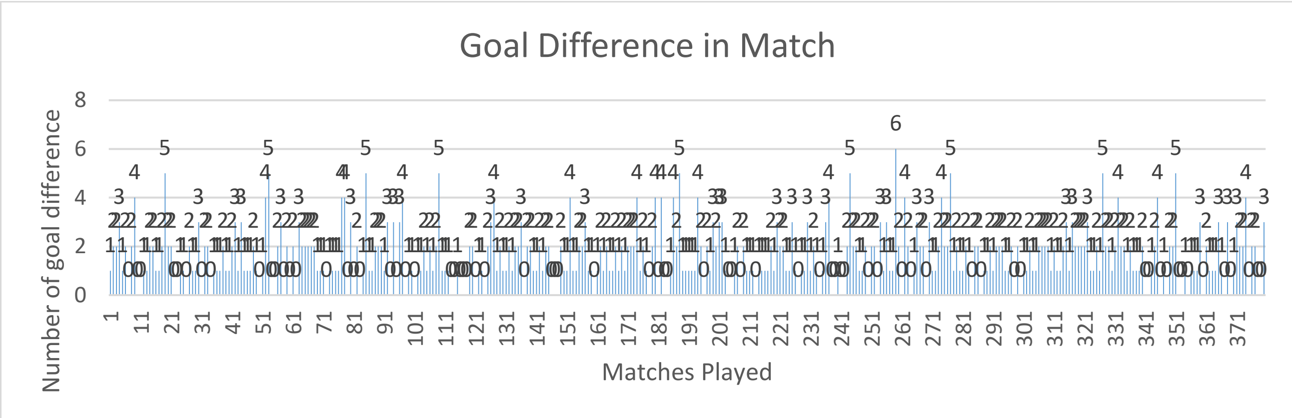


Figure 12: Goal difference per match

Table 3: Goal Margins of matches

Results	Number	Percentage %
Win with One goal margin	130	34.21
Win with Two goals margin	109	28.68
Draws	71	18.68
Wins with more than Two goals Margin	70	18.42
Total Matches	380	100.00

Table 4: Prediction of 2018/2019 EPL Season

	Unnamed: 0	HomeTeam	AwayTeam	BWD	VCD	PSD	FTR	HomeDef	AwayDef
0	0	Man United	Leicester	4.00	4.00	3.93	0	0.65450	0.597000
1	1	Bournemouth	Cardiff	3.40	3.60	3.63	0	0.64475	0.613250
2	2	Fulham	Crystal Palace	3.30	3.40	3.46	0	0.60600	0.635000
3	3	Huddersfield	Chelsea	3.90	4.00	4.02	0	0.60875	0.693750
4	4	Newcastle	Tottenham	3.50	3.40	3.57	0	0.63125	0.695000
...	...	...	...	...	...	...	...	...	...
375	375	Liverpool	Wolves	5.75	5.75	5.77	0	0.70375	0.642667
376	376	Man United	Cardiff	6.25	6.25	6.33	0	0.65450	0.613250
377	377	Southampton	Huddersfield	4.75	4.80	4.83	1	0.61900	0.608750
378	378	Tottenham	Everton	3.50	3.50	3.64	1	0.69500	0.662000
379	379	Watford	West Ham	3.70	3.75	3.85	0	0.64600	0.632750

380 rows × 9 columns

```
In [219]: print(grid.best_estimator_)

SVC(C=0.001, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=0.0001, kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)

In [223]: grid_predictions = grid.predict(X_val)
          print(confusion_matrix(y_val, grid_predictions))
          print(classification_report(y_val, grid_predictions))

[[73  0]
 [22  0]]
              precision    recall  f1-score   support

           0       0.77      1.00      0.87        73
           1       0.00      0.00      0.00        22

    accuracy                           0.77        95
   macro avg       0.38      0.50      0.43        95
weighted avg       0.59      0.77      0.67        95

In [225]: accuracy_score(y_val, grid_predictions)
Out[225]: 0.7684210526315789
```

Figure 13: Snap shot of prediction accuracy

4. **Conclusion**

The research work which is machine learning prediction using Support Vector Machine approach to Access Team's Back '4'for Predicting Match Draw.

The system helps and enable coaches, team owners to determine whether spending high amount on team do bring out good performance or not, and the system would also help the football fans to easily predict their season's performance based on their spending.

The work adopted the following steps for a successful completion of the work

i. Online Data Collection of Team's Back '4'in English Premier League was collected for successful prediction
ii. Support Vector Machine approach was implemented in predicting the dataset to ensure accurate outcome
iii. Then a model was developed for the various entities and actions included in the system
iv. After a model has been developed, the system was implemented using Python Programming Language

After successful completion of the research work, the work has been able to achieve the main aim of the work which is Assessing Team's Back '4'for predicting Match Draw. Also, the proposed system would also help the football fans to easily predict their season's performance based on their spending.

Based on the successful completion of the work and achievements made on the research work, we recommend the system be adopted and used by Coaches, Team's and football fans to predict Draw in a football match outcome.

Also, it is recommended that the proposed system would also help and enable team owners to determine whether spending high amount on team back '4'do bring out good performance or not. So far, our study has worked on the English Premier League, it is expected that this will fit to any other league having the same structure and organization. The limitations to this study should be considered when interpreting the study results. The limitation of this work is that, this can only be applicable for teams that uses back '4', as some couches uses different formations and also the fact that coaches change formation during a game to meet up the situation on-going on the pitch. Also, the evaluation was done in a bid to get a draw from a match, and not necessarily a win, as the efficiency of the back '4'does not guarantee a win. Further work can be done on the midfield and attacking role too in predicting game.

References

[1] M. J., Moroney, "Facts from figures". 3rd edition, Pengium, London, 1956
[2] C., Reep, and B., Benjamin, "Skill and chance in association football." Journal of the Royal Statistical Society. Series A (General), Vol. 131, No. 4, pp. 581-585, 1968

[3] I., Hill, “Association Football and Statistical Inference.” Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 23, No. 2, pp. 203-208, 1974
[4] T., Reilly, and V., Thomas, “A motion analysis of work-rate in different positional roles in professional football match play”. Journal of Human Movement Studies Vol. 2, pp. 87-97, 1976
[5] K., Yamanaka, D. Y., Liang, and M., Hughes, “An analysis of the playing patterns of the Japan national team in the 1994 World Cup qualifying match for Asia”. In T. Reilly, J. Bangsbo, M. Hughes (Eds.). Science and Football III. London: E&FN Spon; pp. 221-228, 1997
[6] D., Partridge, R., Mosher, I., Franks, “A computer assisted analysis of technical performance-a comparison of the 1990 world cup and intercollegiate soccer.” In T. Reilly, J. Clarys, A. Stibble (Eds.). Science and Football II. London: E&FN Spon; 1993, pp. 221-231, 1993
[7] M, Hughes, and G., Probert, “A technical analysis of elite male soccer players by position and success.” In H. Dancs, M. Hughes, P. O’donoghue. (Eds.). Notational Analysis of Sport -VII, Cardiff: UWIC pp. 76-91, 2006.
[8] J. J., Williams, “An investigation into operational definitions used within performance analysis.” Presentation at the International Workshop in Performance Analysis, Lincoln. 2009
[9] M., Lewis, “The art of winning an unfair game.” Managerial and Decision Economics Vol. 25, pp. 550-552, 2003
[10] M. J., Maher, (1982): “Modelling association football scores” Statistica Neerlandica, Vol. 36, pp. 109–118, 1982
[11] K. S., Courneya, and A. V., Carron, “The home advantage in sport competitions: A literature review.” Journal of Sport and Exercise Psychology, Vol. 14, No. 1, pp. 13-27, 1992
[12] L., Knorr-Held, “Dynamic Rating of Sports Teams.” Second Edition. Collaborative Research Center 386, Discussion Paper 98, 1999.
[13] A., Groll, and J., Abedieh, “Spain retains its title and sets a new record - generalized linear mixed models on European football championships,” Journal of Quantitative Analysis in Sports, Vol. 9, pp. 51–66, 2013
[14] A., Groll, G., Schauberger, and G., Tutz, “Prediction of major international soccer tournaments based on team-specific regularized Poisson regression: an application to the FIFA World Cup 2014,” Journal of Quantitative Analysis in Sports, Vol. 11, pp. 97–115, 2015
[15] M. J., Dixon, and S. G., Coles, “Modelling association football scores and inefficiencies in the football betting market.” Journal of the Royal Statistical Society: Series C (Applied Statistics), Vol. 46, pp. 265–280, 1997
[16] H., Rue, and O., Salvesen, “Prediction and Retrospective Analysis of Soccer Matches in a League.” Journal of the Royal Statistical Society: Series D (The Statistician). Volume 49 pp. 399 – 418, 2000
[17] Ankit, “Math behind Support Vector Machine(SVM)” https://mc.ai/math-behind-support-vector-machinesvm, 2019
[18] A., Gelman, J. B., Carlin, H. S., Stern, and D. B., Rubin, “Bayesian Data Analysis.” Chapman and Hall, London, 1995
[19] L., Breiman, (2001). “Random Forests.” Machine Learning Vol. 45, pp. 5–32, 2001
[20] A., Gioldasis, E., Bekris, and I., Gissis, “Playing position: anthropometric and fitness demands in youth soccer.” Sport Science Review Vol. 23 Issue 3-4, pp. 151-168, 2014
[21] M. D., Hughes, “Notational Analysis of Soccer.” In T. Reilly (Ed.). Science and Soccer II. London: E. and F.N. Spon; pp. 245-264, 2004
[22] M., Hughes, T., Caudrelier, N., James, A., Redwood-Brown, I., Donnelly, A., Kirkbride, and C., Duschesne, “Moneyball and soccer - an analysis of the key performance indicators of elite male soccer players by position.” J. Hum. Sport Exerc. Volume 7, No. 2, pp. 402-412, 2012
[23] O. O., Ajayi, “Implementing a Score-Time based Model for Handling Deadlocks in Football Group Matches” International Journal of Computer Applications (0975 – 8887), 2015
[24] H., Avipshu, “The Aesthetics, Poetics and Rhetoric of Soccer, Soccer and Society”, DOI:10.1080/14660970.2019.1694235, 2019
[25] FIFA Magazine. N.p., 2016. Web. 14 Nov., 2016.
[26] F., Amadin, and J. C., Obi., “English Premier League (EPL) Soccer Matches Prediction using An Adaptive Neuro-Fuzzy Inference System (ANFIS).” Transactions on Machine Learning and Artificial Intelligence.”10.14738/tmlai.32.1027, 2015
[27] R., Desmond, L., Colin, F., Shad, M., Alan, “Developing World-Class Soccer Players: An Example of the Academy Physical Development Program From an English Premier League Team.”, 2017
[28] T. Jonathan, “2014-15 English Premier League and MLS TV”, 2014
[29] K., Harrop, and A., Nevill, “Performance indicators that predict success in an English professional League One soccer team.” International Journal of Performance Analysis in Sport, Vol. 14, No. 3, pp. 907-920, 2014
[30] J., Castellano, and D., Casamichana, “What are the differences between first and second divisions of Spanish football teams?” International Journal of Performance Analysis in Sport, Vol. 15, No. 1, pp. 135-146, 2015
[31] P. D., Jones, N., James, and S. D., Mellalieu, “Possession as a performance indicator in soccer.” International Journal of Performance Analysis in Sport, Vol. 4, No. 1, pp. 98-102, 2004
[32] C., Lago, “Consequences of a busy soccer match schedule on team performance: empirical evidence from Spain.” Int Sport Med J. Vol. 10, No. 2, pp. 86-92, 2009

[33] C., Lago-Peñas, and A., Dellal, "Ball possession strategies in elite soccer according to the evolution of the match-score: The influence of situational variables." Journal of Human Kinetics Vol. 25, No. 1, pp. 93-100, 2010
[34] C., Lago-Penas, and J., Lago-Ballesteros, "Game location and team quality effects on performance profiles in professional soccer." Journal of Sports Science & Medicine Vol. 10, pp. 465, 2011
[35] J. B., Taylor, S. D., Mellalieu, N., James, and D. A., Shearer, "The influence of match location, quality of opposition, and match status on technical performance in professional association football." Journal of Sports Sciences, Vol. 26, No. 9, pp. 885-895, 2008
[36] M., Vogelbein, S., Nopp, and A., Hökelmann, "Defensive transition in soccer–are prompt possession regains a measure of success? A quantitative analysis of German Futball-Bundesliga 2010/2011." Journal of Sports Sciences, Vol. 32, No. 11, pp. 1076-1083, 2014
[37] S., Robertson, and D., Joyce, "Informing in-season tactical periodization in team sport: Development of a match difficulty index for Super Rugby." Journal of Sports Sciences Vol. 33, No. 1, pp. 99–107, 2015
[38] D., Higham, W., Hopkins, D., Pyne, and J., Anson, "Performance indicators related to points scoring and winning in international rugby sevens." Journal of Sports Science and Medicine, Vol. 13, No. 2, pp. 358–364, 2014
[39] C., Carling, C., Wright, L., Nelson, and P., Bradley, "Comment on 'Performance analysis in football: A critical review and implementation for future research." Journal of Sports Sciences, Vol. 32, No. 1, pp. 2-7, 2014
[40] D., Dyte, and S. R., Clarke, "A ratings-based Poisson model for World Cup soccer simulation", Journal of the Operational Research Society, Vol. 51, pp. 993–998, 2000
[41] P., Volf, "A random point process model for the score in sport match", IMA Journal of Management Mathematics, Vol. 20, pp. 121-131, 2009
[42] A. K.., Suzuki, L. E. B., Salasar, J. G., Leite and F. Louzada-Neto, "A Bayesian approach for predicting match outcomes: the 2006 (Association) Football World Cup." Journal of the Operational Research Society 61: 1530–1539, 2010
[43] L. S., Bastos, and J. M. C., daRosa, "Predicting probabilities for the 2010 FIFA World Cup games using a Poisson-Gamma model." Journal of Applied Statistics, DOI:10.1080/02664763.2013.788619, 2013
[44] J. B., Keller, "A characterization of the Poisson distribution and the probability of winning a game", The American Statistician Vol. 48, pp. 294–298, 1994
[45] M., Hofmann, and E., Kirwan, "Predicting Premiership Football Match Results (Machine vs. Men – can predictive models outperform human experts?)" Conference: 12th international Information Technology and Telecommunication Conference, At Athlone, Ireland, 2013

[46] D. A., Palinggi, "Soccer outcome with machine learning based on weather condition." 2019
[47] J., Lowther, A., Lane, and H., Lane, "Self-efficacy and psychological skills during the Amputee Soccer World Cup". Atheletic Insight, Vol. 4, 2002
[48] G., Sporis, I., Jukic, S. M., Ostojic, and D., Milanovic, "Fitness profiling in soccer: physical and physiologic characteristics of elite players." J Strength Cond Res Vol. 23, No. 7, pp. 1947–1953, 2009
[49] A., Tenga, D., Kanstad, L. T., Ronglan, and R., Bahr, "Developing a New Method for Team Match Performance Analysis in Professional Soccer and Testing its Reliability", International Journal of Performance Analysis in Sport, Vol. 9, pp. 1, 8-25, 2009
[50] S., António, C., Eduardo and C., Jorge, "Sporting, financial and stock market performance in English football: an empirical analysis of structural relationships." ISEG, Rua Miguel Lupi, 20, 1249-078 Lisbon, Portugal, 2010
[51] J., Bangsbo, M., Mohr, and P., Krustrup, "Physical and metabolic demands of training and match play in the elite football player." Journal Sport of Science Vol. 24, No. 7, pp. 665-674, 2006
[52] A., Molinos, "Evaluation of movement and physiological demands of full-back and center-back soccer players using global positioning systems." J Hum Sport Exerc, Vol. 8, No. 4, pp. 1015-1028, 2013
[53] P. S., Bradley, W., Sheldon, B., Wooster, P., Olsen, P., Boanas, and P., Krustrup, "High-intensity running in English FA Premier League soccer matches." Journal of Sport Science, Vol. 27, No. 2, pp. 159-168, 2009
[54] Di-Salvo, B., Ramon G., Carlos, G., Christian, P., Fabio, and B., Norbert, "Sprinting analysis of elite soccer players during European Champions League and UEFA Cup matches Valter." Journal of sports sciences Vol. 28, No. 14, 1489-1494, 2010
[55] S. N., Dogramaci, M. L., Watsford, and A. J., Murphy, "Time-motion analysis of international and national level futsal." J Strength Cond Res 25(3): 646-651, 2011
[56] D., Alexandre, C., Karim, P. W., Del, A., Said, K., Dominique, B., Ricardo, N. B., Gian, and C., Christopher, "Comparison of physical and technical performance in European soccer match play: FA Premier League and La Liga", European Journal of Sport Science, Vol. 11, No. 1, pp. 51-59, 2011
[57] C., Carling, F., LeGall, and G., Dupont, "Analysis of repeated high-intensity running performance in professional soccer," Journal of Sports Science, Vol. 30, No. 4, pp. 325-336, 2012
[58] T., Vestberg, R., Gustafson, L., Maurex, M., Ingvar, and P., Petrovic, "Executive Functions Predict the Success of Top-Soccer Players." PLoS ONE 7(4): e34731. doi:10.1371/ journal. pone.0034731, 2012

[59] J., Lasek, Z., Szlávik, and S., Bhulai, (2013) “The predictive power of ranking systems in association football”, International Journal of Applied Pattern Recognition Vol. 1, No. 1, pp. 27–46.
[60] L. K., Mmbaya, “Financial performance of football clubs in Kenya case of Kenyan premier league.” Master of Business Administration, University of Nairobi. Master’s Thesis, 2013
[61] O. O., Ajayi, and Alonge “Soccer Match Result prediction system.” International Science Index Journal, Vol. 8, No. 11, pp. 1173 – 1178, 2014
[62] C., Marta, B., Daniel, C., Oleguer, M. A., Teresa, C., Albert. and H., Raúl (2016), “Goal Scoring in Soccer: A Polar Coordinate Analysis of Motor Skills Used by Lionel Messi,” doi: 10.3389/fpsyg.2016.00806, 2016
[63] A. V., Carron, W. N., Widmeyer, and L. R., Brawley, “The development of an instrument to assess cohesion in sport teams: The group environment questionnaire.” Journal of Sport Psychology, Vol. 7, pp. 244-266, 1985
[64] S., Bruno, M. C., Filipe, C., Miguel, and B., Pedro, “Functional Movement Screen Scores and Physical Performance among Youth Elite Soccer Players method, Sports Vol. 5, No. 16, doi:10.3390/sports5010016, 2017
[65] G., Aristotelis, S., Athanasios and C., Orestis “Technical skills according to playing position of male and female soccer players,” International Journal of Science Culture and Sport Vol. 5, No. 4, pp. 2148-1148, 2017
[66] A., Groll, T., Kneib, A., Mayr, and G., Schauberger, “On the dependency of soccer scores – A sparse bivariate Poisson model for the UEFA European Football Championship 2016,” Statistical Modelling, 2018
[67] Y., Qing, J., Hong, L., Hongyou and M. Á., Gómez, Technical demands of different playing positions in the UEFA Champions League, International Journal of Performance Analysis in Sport, 2018
[68] M., Gómez, J., Álvaro, M. I., Barriopedro, “Behaviour patterns of finishing plays in female and male soccer. Kronos,” Vol. 8, No. 15, pp. 15-24, 2009
[69] P., Luca, C., Paolo, F., Paolo, M., Emanuele, P., Dino, and G., Fosca, “PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach.” ACM Trans. Intell. Syst. Technol. Vol. 10, No. 59, pp. 1-27, 2019
[70] T. G., Omomule, A. J., Ibinuolapo, and 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).
[71] T., McGarry, and I. M., Franks, “In Search of Invariant Athletic Behaviour in Competitive Sport Systems: An Example from Championship Squash Match-Play.” Journal of Sports Sciences, Vol. 14, pp. 445-456, 1996
[72] W., Abbott, G., Brickley, and N. J., Smeeton, “Physical demands of playing position within English Premier League academy soccer.” Journal of Human Sport and Exercise, 13(2): 285-295, 2018
[73] K., Dimitris, and N., Ioannis, “Analysis of sports data by using bivariate Poisson models,” Journal of the Royal Statistical Society: Series D The Statistician Vol. 52, pp. 381–393, 2003
[74] M. A., Gomez, C., Lago, M. T., Gomez, and P., Furley, “Analysis of elite soccer players’ performance before and after signing a new contract.” PLoS ONE Vol. 14, No. 1, 2019
[75] M. F., Ian, and G., David, “A systematic approach to analysing sports performance.” Journal of Sports Sciences Vol. 4, No. 1, pp. 49-59, 1986.
[76] L., Joaquin, and L., Carlos, “Performance in Team Sports: Identifying the Keys to Success in Soccer.” Journal of Human Kinetics Vol. 25, No. 3, pp. 85-91, 2010
[77] A., Tsakonas, “Soft Computing-Based Result Prediction of Football Games,” Soft Computing-Based Result Prediction of Football Games // Proc. Of the First International Conference on Inductive Modelling ICIM'2002, 2002
[78] J., Jooste, B. J. M., Steyn, and B. L., Vanden, “Psychological skills, playing positions and performance of African youth soccer teams”, South African Journal for Research in Sport, Physical Education and Recreation Vol. 36, No. 1, pp. 85-100, 2014
[79] I. S., Ishak, and R., Alias,“Designing A Strategic Information Systems Planning Methodology For Malaysian Institutes Of Higher Learning (Isp-Ipta).” Issues in Information Systems Vol. 6, pp. 325-331, 2005
[80] B. S., Blanchard, and W. J., Fabrycky, (2010). Systems Engineering and Analysis (5th Ed.), New Jersey: Prentice hall.
[81] R., Nazim, M., Aida, A. Y., Faiz, and A. A., Ruhaya, “IOP Conference Series: Materials Science and Engineering,” Volume 226, International Research and Innovation Summit (IRIS2017) 6-7 May 2017, Meleka, malaysia.