# INTERNATIONAL JOURNAL
# FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

# Survey on Game Prediction System

Nikith Joshua Dsouza[1], S. Gautham[2], Samrudh H N[3], Shaman Bhat B U[4], Nagarathna N[5]

[1, 2, 3, 4]*Undergraduate student,* [5]*Professor, Department of Computer Science & Engineering, B.M.S College of Engineering*

*Abstract: Predicting the outcome of a sporting event is very challenging, even for sports experts. Predicting sports games is gaining popularity with the advancements in machine learning algorithms, thus making it an exciting challenge to predict the outcome of the sporting match. The outcome of any match depends on various factors, including players involved, their skills, injuries, and various other factors. Currently, there are a plethora of websites that provide various results based on multiple parameters that are less accurate and ambiguous. In addition, possibilities for improvements in this system have been explored in order to design a prediction system that increases the accuracy of results. This paper speaks of using Machine Learning (ML) approaches such as Logistic regression, Support Vector Machines (SVM), Random forest regressor, and Decision trees to predict the result of a game.*

*Keywords: Machine Learning(ML), Logistic Regression, SVM, Random forest regressor, Decision trees.*

## I. INTRODUCTION

Predicting the results of a sports game is gaining popularity in the world, which in turn has given rise to online sports betting. Prominent sports such as Football, Hockey, Cricket, and Basketball have led to increasing demand for software that is accurate in predicting the outcome of the game. Multiple algorithms have been developed in recent times to achieve this goal. But, developers are faced with various problems, such as finding the right datasets and using the right parameters in the algorithms. This paper surveys a system that predicts the overall outcome of different matches in a particular game and to find different algorithms that can be used to predict the results of these games. By using the correct dataset and the most accurate prediction algorithms, we can generate the win probability of a given team with high accuracy[1,4].

## II. PREDICTION OF VARIOUS TYPES OF GAMES

Prediction is used in sports like basketball, football, cricket, and various other sports. Here we did a study on two sports:

### A. Prediction on Basketball

In the paper by Dragan Miljković, he proposed a way to predict the outcomes of the National Basketball Association (N.B.A.) games in a season. This is considered as a classification problem, where the result falls in either one of these categories: Host team wins, or the visiting team wins. For prediction, 141 attributes are used. For each of the teams, these attributes are divided into two categories. The first one consists of attributes of a game like the number of 3-pointers and field points. The second one consists of attributes of the current league, like the total number of wins and losses, and streak.

To calculate the spread, multivariate linear regression was used, and the point difference was used as a target variable. All the other attributes were used as explanatory variables. Feature selection and normalization were used in order to improve the performance of the classification system. Miljković carried out experiments with a variety of different classification techniques such as decision trees, k-nearest neighbors, Naive Bayes, and SVM. Naive Bayes classifier, combined with normalization and feature selection, gave the best and most accurate results. Miljkovic achieved an accuracy of around 67%, which shows that the predicted outcome of more than two-thirds of the games was successful from a total of 778 matches in that period[1].

### B. Prediction on Football

S. Mohammad Arabzad proposes an NN-based perspective to predict the outcome of the last week of the 13th Iran Pro League. For the prediction to be accurate, a set of essential criteria that affect the match results are required. To obtain the necessary output, an extensive database of match results of the past ten years was collected. Using data mining over the acquired data allows the N.N. Model to learn and therefore predict the result of the future matches. The methodology can be visualized in four steps, as shown in Fig. 1[3].

1) *Selecting Prediction Parameters:* In order to predict the match results, six match-based parameters were selected as prediction parameters. These parameters consisted of past performance in earlier seasons, performance in previous encounters between the teams, win streaks, managerial changes, home advantages, and selected strategy for the game[3].
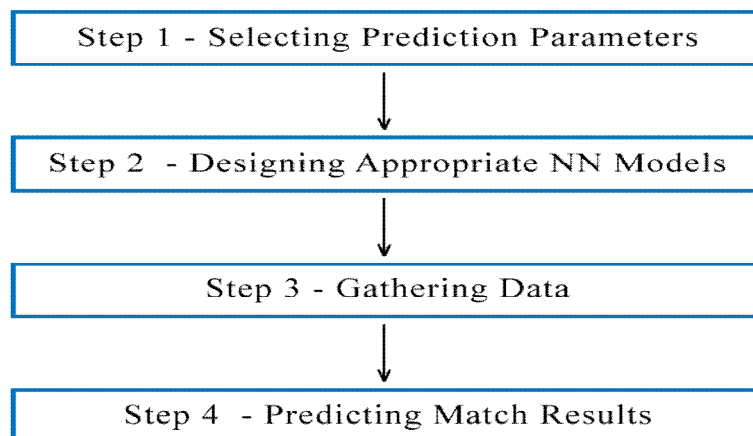
Fig 1. Four-step methodology

2)  *Designing Appropriate N.N. Model:* The most applicable N.N. Model, Multi-Layered Perceptron (M.L.P.), shown in Fig. 2, was chosen as the most appropriate model to predict the results of football matches. Many input parameters and one output are made the structure of the M.L.P. model[3].
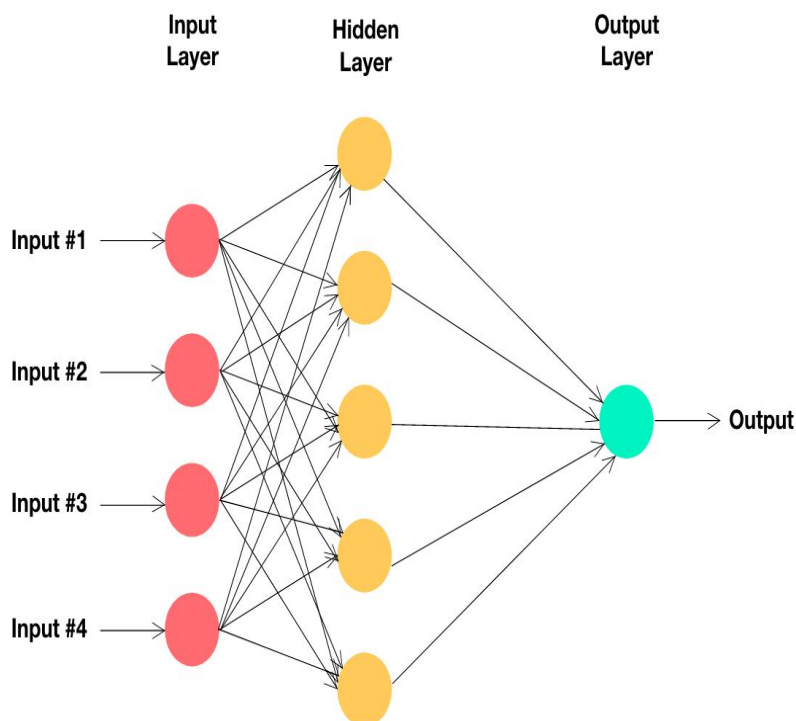


Fig 2. Multi-Layered Perceptron

3)  *Gathering Data:* The football datasets consist of the last ten years of match results of the Iran Pro League. This dataset will be used as inputs to the N.N. The dataset must first be cleaned. The null values, unwanted data, and corrupted data must first be removed. The ten years of Iran Pro League data must then be combined into a single dataset that can be used by the model[3].

4)  *Predicting the Match Results:* The final outcome of this model is to get the win probability of each team. By using various regression models and the most appropriate algorithms, we try to find the highest accuracy that can be achieved[3]. These are the steps that are generally followed while predicting the outcome of the game. The following section talks about the parameters that play a role in prediction.

## III. PARAMETERS THAT INFLUENCE PREDICTION

The following parameters are said to be influential in predicting the result of a game. Each of these parameters explains how they affect the outcome of a particular game. These parameters change with respect to a sport.

### A. Matches Won, lost, Ended in Draw

Miljkovic showed the win:loss:draw ratios of basketball teams to be significant features in predicting National Basketball League (N.B.A.) basketball matches. The same ratio can be used to predict football matches as well. Every team's wins and losses are essential parameters in finding their winning probability. The results of sports games are determined by the number of points/goals scored by each team in a game[5,6,8].

### B. Goals/Points Scored

Baio proposed a Gamma distribution mixture model for Italian Serie A match prediction, relying on the number of goals scored and conceded by the two teams in a match. Miljkovic used a number of field points scored and acknowledged, and the number of 3-pointers scored and conceded for the prediction of basketball match outcome prediction. With possible automatic retrieval of win/lose/draw percentages and an average number of scored and conceded goals, these features can be added to the candidate feature set[8,9].

### C. Goals/points Conceded

The amount of goals/points conceded shows how good the defense of a team is. By taking the goals conceded and total shots conceded, we can find the Goals Scored Against Ratio. Miljkovic considered the number of field points and 3-pointers that were conceded and scored for predicting a basketball game[8,9].

### D. Performance in Earlier Encounters

It was found that the previous games played between the same teams have a relatively high correlation in a three games tally in LaLiga in Aranda-Corral papers. In their prediction model, they decided to make three categories: the number of draws, wins, and losses for either one of the teams in all the seasons that have passed. Hucaljuk and Rakipovic have also done the same thing by considering the results of the previous games between two teams[10,11].

### E. Streaks

It was shown in Heuer and Rubner's paper that a streak of losing does have a negative impact on a team's chance of winning; however, the vice-versa is not true. Goddard, in his paper, has shown otherwise, that losing streak increases the probability of winning, and vice-versa. Constantinou, in his paper, stated that form is one of the important models that are created by experts[12,13,14].

### F. Managerial/Coaching Changes

Different studies state different effects of managerial tactics on the results of a game. In a paper written by Koning and Bal-duck, the effect of these changes depends on the new manager and team characteristics[15,16].

### G. Game Strategies

In Palomino's paper, he used a game theory approach to highlight the effect of playing strategies on the outcome of a game. As this strategy can even change during the course of a game several times, it's not recommended to use this as a feature[17].

### H. Home Advantage

Every football team has its home stadium. Teams playing games in their own home stadium, in front of their fans, can massively boost their morale.

The away team could also be intimidated and tired when traveling to a different stadium. This is proven in the paper written by Palomino and Carmichael and Thomas. However, it is also shown that this advantage differs with different leagues in consideration like the Turkish Super League (T.S.L.) and English Premier League (EPL)[2,17,18,19].

*I. Injuries & Suspensions*

Drawer & Fuller assessed how the injuries of top key players in a squad affect the way a team performs a match. In the assessment, the team's quality is measured as the total quality of all the players in a team of the five seasons of EPL. This assessment concluded that a team's performance in a game is significantly affected by injuries to a top player in that team[20,21,22]. These are the parameters that are chosen widely for predicting a game. The parameters to select depend heavily on the sport being considered.

## IV. MODELS FOR PREDICTION

The section talks about the various models, be it Machine learning models or Neural Network models that were studied and used in prediction

*A. Dimensionality Reduction*

According to Niek Tax and Yme Joustra, based on the experiments performed on the effect of various dimensionality reduction techniques on the accuracy of prediction, the following techniques were chosen: Principal Component Analysis(P.C.A.), Sequential Forward Selection[7].

*B. Classification Techniques*

Various new techniques are being introduced to help us in analyzing, much more accurately, this huge amount of data to gain useful insights. According to Niek Tax and Yme Joustra, Bayesian Network is viewed as a common classification technique that is used with such huge data but is seen to have its downside as well. In fact, in work done by Hucaljuk & Trevathan, they recommend that it is not the best performing model. Other classification models include SVM, Decision trees, fuzzy system, and logistic regression[23].

1) *Support Vector Machines (SVM):* Maral Haghighat implies that, in general, SVM uses the concept of non-linear mapping of the given training set with a high value of dimensionality. The SVM algorithm searches for an optimal separating hyperplane, which is used to act as a decision boundary between the two classes. An SVM will find the optimal hyperplane by employing the vectors, which is nothing but the training dataset and margins, which are defined by vectors. The SVM algorithm is believed to have high accuracy due to its high capability in building non-linear, complex decision boundaries, but training the data using SVM takes more time compared to other algorithms[23,24].

2) *Decision Trees:* In the paper written by Kalpdrum and Niravkumar, a decision tree is similar to a flowchart but in a tree-like manner, where each node is an attribute of the considered dataset, and each branch is the outcome of it. It follows the ID3 algorithm, which is based on information gain. The highest accuracy that they were able to come up with this model is 80.46%, with 90% train data and 10% test data.[23].

3) *Fuzzy System:* Fuzzy logic is a newly developed technology that requires advanced and complicated mathematical analyses and contributes to the development of systems. While in traditional binary sets, variables can take either one or zero, fuzzy logic variables may have a truth value, which ranges from one to zero. Fuzzy systems can accurately describe irrational and indefinite phenomena. They work based on continuous membership function rules (IF-THEN rules) stored in a knowledge base. A fuzzy system converts human knowledge into a definite mathematical formula[4,23].

4) *Logistic Regression:* The most popular tool when it comes to classification problems is Logistic Regression. Similar to linear regression, logistic regression depends on a linear combination of features; these features are mapped to a value between one and zero by the logistic function. Therefore, the dependent variables must have a continuous value, which is also a function of the odds of the events. Logistic regression involves two stages. The first stage is estimating the odds of characteristics for each group, and the second stage is determining the cut-off points and categorizing the features accordingly. The coefficients are estimated by maximum likelihood estimation[23].

Many other algorithms, like KNeighbours classification, S.V.C., linear regression, are also considered in researches, which proves to be less accurate. The choice of the algorithm also depends on the sport taken into consideration.

## V. CONCLUSIONS

Sports have gained increasing popularity in recent years. Many teams spend a huge amount of money to gain better results in a match. To understand key statistics, which would help to improvise and strategize for further matches. The principle undertaking of game prediction is to help users to predict the outcome of a particular game. Sports results prediction proves to be a vital application in the field of sports as coaches can use sports prediction to prepare their teams for any game. Traditionally, this is done using

statistical and mathematical models, but ML models are gaining popularity as it is proving to be more accurate. Despite saying that, some of these models are not accurate enough because of various reasons like team inconsistency, strategic differences, sudden player injuries, and fatigue level. Also, implementing twitter sentiment analysis about the game, and the players play an important role in determining various factors which are useful to provide accurate results through various algorithms. With more information about each match statistics in the dataset, the model can be made more accurate to predict the outcome of matches.

## REFERENCES

[1] Dragan Miljković, Ljubiša Gajić, Aleksandar Kovačević, Zora Konjović "The Use of Data Mining for Basketball Matches Outcomes Prediction"

[2] Igiri, Chinwe Peace; Nwachukwu, Enoch Okechukwu" An Improved Prediction System for Football a Match Result"

[3] S. Mohammad Arabzad1*, M.E. Tayebi Araghi1, S. Sadi-Nezhad1,Nooshin Ghofrani" Football Match Results Prediction Using Artificial Neural Networks; The Case of Iran Pro League"

[4] B., Kim, J., Choe, C., Eom, H. and McKay, R.I. (2008) "A compound framework for sports results prediction: A football case study", Knowledge-Based Systems, Vol. 21, No. 7, pp. 551-562.

[5] International Journal of forecasting, Goddard, J. "Regression models for forecasting goals and match results in association football" (2005) Vol. 21, No.2.

[6] Goddard, J. and Asimakopoulos, I. "Forecasting football results and the efficiency of fixed-odds betting" (2004) Journal of Forecasting, Vol. 23, No.1.

[7] Tax, Niek & Joustra, Yme. (2015). "Predicting the Dutch Football Competition Using Public Data: A Machine Learning Approach." 10.13140/RG.2.1.1383.4729.

[8] D. Miljkovic, L. Gajic, A. Kovacevic, and Z. Konjovic, "The use of data mining for basketball matches outcomes prediction," in Proceedings of the 8th International Symposium on Intelligent Systems and Informatics. IEEE, 2010, pp. 309–312.

[9] G. Baio and M. Blangiardo, "Bayesian hierarchical model for the prediction of football results," Journal of Applied Statistics, vol. 37, no. 2, pp. 253–264, 2010.

[10] G. A. Aranda-Corral, J. Borrego-D'ıaz, and J. Gal' an-Paez, "Complex concept lattices for simulating human prediction in sport, Journal of Systems Science and Complexity, vol. 26, no. 1,pp. 117–136, 2013.

[11] J. Hucaljuk and A. Rakipovic, "Predicting football scores using machine learning techniques," in MIPRO, 2011 Proceedings of the 34th International Convention. IEEE, 2011, pp. 1623–1627.

[12] A. Heuer and O. Rubner, "Fitness, chance, and myths: an objective view on soccer results," The European Physical JournalB, vol. 67, no. 3, pp. 445–458, 2009.

[13] J. Goddard, "Who wins the football?" Significance, vol. 3, no. 1,pp. 16–19, 2006.

[14] A. C. Constantinou, N. E. Fenton, and M. Neil, "pi-football: A Bayesian network model for forecasting association football match outcomes," Knowledge-Based Systems, 2012.

[15] R. H. Koning, "An econometric evaluation of the effect of firing a coach on team performance," Applied Economics, vol. 35,no. 5, pp. 555–564, 2003.

[16] A. L. Balduck, A. Prinzie, and M. Buelens, "The effectiveness of coach turnover and the effect on home team advantage, team quality and team ranking." Journal of Applied Statistics, vol. 37, no. 4, pp. 679–689, 2010.

[17] F. Palomino, L. Rigotti, and A. Rustichini, "Skill, strategy and passion: An empirical analysis of soccer." in Proceedings of 8thWorld Congress of the Econometric Society, 2000, pp. 11–16.

[18] F. Carmichael and D. Thomas, "Home-field effect and team performance evidence from English premiership football." Journal of Sports Economics, vol. 6, no. 3, pp. 264–281, 2005.

[19] A. Seckin, "Home advantage in association football: Evidence from Turkish super league," in Proceedings of the EcoMoD2006Conference, 2006.

[20] Samba, Stefan. (2019). Football Result Prediction by Deep Learning Algorithms. 10.13140/RG.2.2.25014.45122.

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

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

[23] R. G. Sonkamble, T. A. Dasare, S. S. Patil, S. B. Patil, I. C. Pawar, A. M. Phadtare, Sports Result Prediction System, International Journal of Research in Engineering, Science and Management Volume-2, Issue-1, January-2019F. Carmichael and D. Thomas, "Home-field effect and team performance evidence from English premiership football."Journal of Sports Economics, vol. 6, no. 3, pp. 264–281, 2005.

[24] Ben Ulmer, Matthew Fernandez, Predicting Soccer Match Results in the English Premier League.

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  ⓦ (24*7 Support on Whatsapp)