**SYNOPSIS REPORT**

**FOR PROJECT-I**

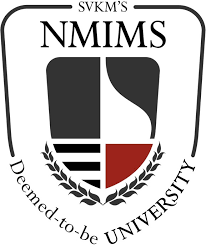
**ON**

**“Football Match Prediction Using Machine Learning”**

**Submitted in Partial Fulfillment of requirements for the Award of**

**Degree of Bachelor of Technology in Computer Engineering**

**Submitted To**



**SVKM’s NMIMS UNIVERSITY**

**Submitted By**

Pranav Kolhe

Dhairya Desai

Abhraneel Das

**Under the guidance of**

Dr. Upendra Verma

**Department of Computer Engineering**

**Mukesh Patel School of Technology Management & Engineering, Shirpur Campus**

**Table Of Contents**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sr. No** | **Title** | **Year** | **Author** | **Algorithm** | **Outcomes** | **Limitations** |
| **1** | Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League | 2020 | Jassim Almulla, Tanvir Alam | KNN, Logistic Regression, Random Forest, XG Boost Classifier, Decision Tree, MLP Classifier, SVM | Out of all the models, Logistic Regression achieved the highest accuracy of 80.1% | Limited Data.  Home/Away factor, grass and ball types, etc not included. |
| **2** | Prediction of Football Matches Results: Decision Forest against Neural Network | 2021 | Azrel Aiman Azeman, Aida Mustapha, Nazim Razali, Aziz Nanthaamornphong, Mohd Helmy Abd Wahab | Multiclass Neural Network, multiclass Decision Forest | Decision Forest had highest accuracy of 88.95%. | Random Forest would have given more accurate result for the dataset used in this paper. |
| **3** | Clustering and Evolutionary System Analysis of Data Mining Algorithms in the field of Football | 2022 | Dongdong Wang | Data clustering through local metric learning,  FKM clustering method | Winning rate percentages ranging from 70-80% were found for all 6 sportsmen | Economic benefits were to be analyzed |
| **4** | Comparative Analysis for Predicting Football Match Outcomes based on Poisson Models | 2022 | Syasya Nadhilah Maozad, Siti Noor Asyikin Mohd Razali, Aida Mustapha, Aziz Nanthaamornphong, Mohd Helmy Abd Wahab, Nazim Razali | Dixon-Coles model, Rue-Salvesen adjustment model, Poisson Regression model | DixonColes model because the value of log-likelihood is the highest meanwhile the value of AIC is the lowest. It has an accuracy of 82% | The models may differ in result if were used to predict expected goal score. |
| **5** | Prediction of Football Players Performance using Machine Learning and Deep Learning Algorithms | 2021 | S Manish, Vandana Bhagat, RM Pramila | Performance Prediction, Deep Learning Algorithm, Machine Learning algorithm, Linear Regression, Neural networks. | The mean squared error for the defence model is 21.3 although the R squared value is 0.90. The R squared values from the  XGBoost algorithm is 0.90. | The multiple linear regression results are better than other regression model for the experimental data. |
| **6** | A Deep Learning Framework for Football Match Prediction | 2022 | S.K. Nivetha, M. Geetha, R.C. Suganthe, R Manoj Prabhakaran, S Madhuvanan, A Mohamed Sameer | LSTM (Long Short-Term Memory), RNN (Recurrent Neural Networks), Deep Learning. | The LSTM model gives the accuracy of  86%. | This model has room for improvement. One approach to do so is to use a larger number of datasets. They can include  individual player statistics. |

**Introduction**

One of the most popular sports in the world is Football. Millions of people watch Football every day. Big clubs like Manchester City, Barcelona, Real Madrid, etc. have a huge fan following. With such a huge fan following comes huge expectations. Every fan wants its team to win. This makes the Predictions of Football matches even more popular, be it the manager, staff, pundits or even the fans themselves. These predictions are made using a random, instinctive method by observing a few matches. In recent times, a lot of Research is been done in Football Predictions. Many Machine Learning and Deep Learning Algorithms are used to predict which team is going to win the match. It is been estimated that Football is played over 200 countries and 1.4 billion Football fans are supporting their teams globally. European Football alone exceeds 28 billion EUROS market scale. This forces the football team management to select proper strategy for each match. This is done with a team of analysts analyzing the data of their players, as well as their opponents.[1]

Football is one of the most far reaching and famous game, along these lines anticipating the consequences of a football match represents a fascinating test. Forecast can likewise help administrators and clubs in settling on the best choices to win associations and competitions. In view of the huge financial stakes engaged with wagering, there is an interest for high prescient exactness in this industry. Especially, LSTM model of Recurrent Neural Network is utilized in our exploration. LSTMs are a slightly modified version of RNNs that use multiplications and adds to make tiny changes to the data[5]. In modern days the margin of error for football game is low, therefore the ultimate aim of the game is to win the match. The performance of the players in the match affects the results of the game. Due to this it is very important to evaluate the player and know his weakness. Manual evaluation tends to generate many errors and take more time. In the current research the statistical model is proposed to predict the stats of the football player based on previous session data by considering various aspects of the game. Through literature reviews it is observed that machine learning and deep learning algorithms can be used predict the performance of football player. [6]

**Literature Review**

**Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League**[1]

This paper compares various ML algorithms to predict the match result in the upcoming season. Football players’ performance is analyzed of the Qatar Stars League (QSL). Classification Framework was formulated to distinguish between the winning team and the losing team. Effectiveness of various parameters like shots on target, distance covered, successful passes, etc. weas checked. This paper worked with various algorithms, from which the Logistic Regression based model was considered to be the best performing model with an accuracy of 80.1%. Data off last 5 years was considered.

Authors claim that this paper would help the players, coaching staff and the team management to focus on specific performance metrics that may lead to winning a match in the Qatar Stars League. This would be made possible as this paper also highlights the key performance metrics from players playing in QSL.

Although the model was successful for QSL, the authors seem to work on bigger, better datasets that have a wider range of players in them. Also, the authors aim to incorporate the drawn matches into the development of machine learning models to predict win, loss and drawn matches.

**Prediction of Football Matches Results: Decision Forest against Neural Network**[4]

This paper presents a comparison between Multiclass Decision Forest algorithm and Multiclass Neural Network Algorithm on comparing a dataset of 2005-2006 Season of the Premier League. Previous work on this particular field was analyzed and then the predictions of outcomes of a Football match were made. These comparisons were carried out using the Azure Machine Learning Studio. Decision Forest had an accuracy of 88.95% and proved to be a better approach than Neural network, that provided 71.58% accuracy.

Authors feel that Random Forest would be more accurate than Decision Trees for this particular dataset as it consists of tabular and categorical values. In future, the authors plan to use this model on different football leagues across the world.

**Clustering and Evolutionary System Analysis of Data Mining Algorithms in the field of Football**[5]

Firstly, this paper introduces clustering mining and expounds the development of data mining. Then, the application of clustering mining technology is studied, the system structure of information clustering mining in the field of football is designed, and the relevant information of football players in a football club is extracted for testing. Finally, the test results show that the system can mine specific information from a large number of messy information data, and can effectively analyze the data value and present it to football trainers.

**Comparative Analysis for Predicting Football Match Outcomes based on Poisson Models**[6]

This study has three objectives which are to predict the football matches outcome using Poisson models, to predict football matches outcomes using Poisson Regression model and lastly is to evaluate the results using prediction accuracy, log-likelihood, AIC value and R-squared value.

The method being used in this research is Poisson Models (default model), the Dixon-Coles model, the Rue-Salvesen adjustment model, as well as the Poisson Regression model. Poisson distribution known as a discrete distribution which is measuring the probability of a given time event over the exact time period

The results concluded that the best prediction model in this study among is the DixonColes model because the value of log-likelihood is the highest meanwhile the value of AIC is the lowest.

**Prediction of Football Players Performance using Machine Learning and Deep Learning Algorithms**[3]

In football the errors committed by players affects the overall result of the match. The important aspect of winning the game is the player’s performance. In this paper we are predicting the player’s performance by using Deep Learning and Machine Learning with the help of the previous years’ data. The known strength and weaknesses of the player becomes helpful to improve the gameplay.

The current study proposed statistical techniques, to predict the future stats of the players based on the current data and develop a model that will consider all the aspects of the game and evaluate the player based on their performance. In the study 4 regression models namely Multiple Regression, Neural Network, XGBoost Regression and Support Vector Regression are implemented and performance analysis has been done. From the results it can be observed that Multiple Regression has outperformed than any other models and predicted the players performance more accurately compared to other models.

Authors claim that this paper would help the players, coaching staff and the team management to focus on the players’ specific performance metrics as the modern-day football is very precise in decision making. Authors also mentioned a framework a MCDM (Multi-Criteria Decision Making) and collecting data from various fitness tests to give a performance rating to the players.

**A Deep Learning Framework for Football Match Prediction**[2]

This paper proposed a profound neural organization-based model to naturally anticipate the result of a football match, explicitly the English Premier League, in this review. In this paper we are using the deep learning algorithm RNN with LSTM model. Here, we did the proposed system which is dataset collection, pre-processing, training the model, model validation and prediction of the model. Then we also created the confusion matrix and the classification report which contains precision, recall, F1 score, support, macro and weighted average.

Experiments are performed in English premier league datasets using the LSTM model of the recurrent neural networks. In which the LSTM model gives the accuracy of 86%. It is observed, when the epochs are in the range of 0 to 8, the accuracy grows gradually but in the range of 50 % to 60 %, and when the epochs are more than 8, the accuracy increases suddenly up to 80 % to 90%.

**Problem Statement**

Given a larger dataset with in-depth information on the players and the team, predict the outcomes of a Football match using various Machine Learning Algorithms.

**Expected Outcome**

Using multiple Machine Learning Algorithms, the model with the highest accuracy for win-loss-draw results would be considered in our project as the best predictive model for Football Match Predictions.

**References**

[1] J. Almulla and T. Alam, “Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League,” *IEEE Access*, vol. 8, pp. 213695–213705, 2020, doi: 10.1109/ACCESS.2020.3038601.

[2] S. K. Nivetha, M. Geetha, R. C. Suganthe, R. M. Prabakaran, S. Madhuvanan, and A. M. Sameer, “A Deep Learning Framework for Football Match Prediction,” *2022 International Conference on Computer Communication and Informatics, ICCCI 2022*, 2022, doi: 10.1109/ICCCI54379.2022.9740760.

[3] S. Manish., V. Bhagat, and R. M. Pramila, “Prediction of football players performance using machine learning and deep learning algorithms,” *2021 2nd International Conference for Emerging Technology, INCET 2021*, May 2021, doi: 10.1109/INCET51464.2021.9456424.

[4] A. A. Azeman, A. Mustapha, N. Razali, A. Nanthaamomphong, and M. H. Abd Wahab, “Prediction of football matches results: Decision forest against neural networks,” *ECTI-CON 2021 - 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology: Smart Electrical System and Technology, Proceedings*, pp. 1032–1035, May 2021, doi: 10.1109/ECTI-CON51831.2021.9454789.

[5] D. Wang, “Clustering and Evolutionary System Analysis of Data Mining Algorithms in the Field of Football,” *2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, IPEC 2022*, pp. 1341–1344, 2022, doi: 10.1109/IPEC54454.2022.9777487.

[6] S. N. Maozad, S. Noor Asyikin Mohd Razali, A. Mustapha, A. Nanthaamornphong, M. H. Abdul Wahab, and N. Razali, “Comparative Analysis for Predicting Football Match Outcomes based on Poisson Models,” pp. 1–4, Jun. 2022, doi: 10.1109/ECTI-CON54298.2022.9795385.