Pro Kabaddi Winner Prediction using Logistic Regression

Mr.Shrinivas Biradar[1], Kavya M[2], Mahima K N[3], Manasa Roopini N[4], Meghana M[5]

Assistant Professor, Department of CSE[1]

Students, Department of CSE[2,3,4,5]

SJC Institute of Technology Chikkaballapura, India

Abstract: *Pro Kabaddi League (PKL) is an Indian men's professional Kabaddi League. It was launched in 2014. After the IPL, Pro Kabaddi League is the most watched sports league in India. In our model, prediction of winning team in Pro Kabaddi League will be implemented by using Logistic regression algorithm. Logistic regression is a supervised learning algorithm used for predicting the categorical dependent variable using a given set of independent variables. Kabaddi Prediction is one of the expanding areas in good predictive accuracy as it involves huge money in betting. The objective is to predict the full time result of the kabaddi match, which decides the winning team. For this prediction model we are collecting data from the past seasons of Pro Kabaddi League.*

Keywords: Pro kabaddi league (PKL), prediction model, win and loss,Sports,Logistic regression.

I. INTRODUCTION

Pro Kabaddi League (PKL) was started in 2014 with eight teams from different cities. This league is formally backed by the International Kabaddi Federation, the Asian Kabaddi Federation and the Amateur Kabaddi Federation of India[6]. The Pro Kabaddi League highlighted the new, modern, international and competitive face sports of kabaddi game throughout the world. Kabaddi is one of the traditional of India, which requires a combination of Agility, Speed, Power, Strength and Co-ordination. Although it is an outdoor sport that is being played on clay court, but now a days with its great success it is being played inindoors on synthetic surface. Kabaddi is a contact team sport, where 14 players (7 on each side) take part at time. Time span of the match is 40 minutes divided equally in two halves of 20 minutes each along with a rest period of 5 minutes between halves[5]. In today's high-tech world, technology in sport plays a vital role for not only recording data in many ways but also analyse it and present it in such a format (graph or tables) that the coaches could easily interpret and make some conclusion for the individual or team performance. These match statistics or data vary from sport to sport and is the one that viewers want to know above all. Many researches were done to predict the future events on the basis of present or past data. These researches had proved that the future performance of an individual or team could be predicted through the analysis of certain variables, which are found to be the basis for total performance[5]. Logistic regression is one of the statistical methods that can be one of the solutions of many prediction models in sports. The main purpose of this statistical technique is to predict the outcome (binary or multinomial) on the basis of predictor variables selected. Many prediction models have been developed by researchers in predicting the match outcome in several sports. The purpose of this study is to develop a prediction model to estimate the outcome of Pro Kabaddi League on the basis of previous seasons data

II. LITERATURE SURVEY

A literature survey or a literature review in a project report shows the various analyses andresearch made in the field of interest and the results already published, taking into account the various parameters of the project and the extent of the project. Literature survey is mainly carried out in order to analyze the background of the current project which helps to find out flaws in the existing system & guides on which unsolved problems we can work out. So, thefollowing topics not only illustrate the background of the project but also uncover the problems and flaws which motivated to propose solutions and work on this project.Before building our application, the following system is taken into consideration:

Fatima Rodrigues, Angelo Pintob. [1] Prediction of football match results with Machine Learning – (2022),In this article, they develop statistics of previous matches and attributes of players from both machine learning methods that take multiple teams as inputs to predict the outcome of football matches. Several prediction models were tested, with the experimental results showing encouraging performance in terms of the profit margin of football bets.

Dhonge, Shraddha Dhole, Nikita Wavre, MandarPardakhe, AmitNagarale . [2] IPL cricket score and winning prediction using machine learning techniques – (2021), Inthese paper the model has been proposed that has two methods the first one is prediction ofscore and the second one is team winning prediction. In these the score prediction includeslinear regression, lasso regression and ridge regression whereas in winning prediction SVC classifier, decision tree classifier and random forest classifier are used. The model used the supervised machine learning algorithm to predict the winning. Random Forest Classifier used for good accuracy and the stable accuracy so that desired predicted output is accurate.

Patrick LoolaBokonda , Khadija Ouazzani-Touhami,NissrineSouissi .[3] Predictive analysis using machine learning : Review of trends and methods – (2020), The purpose of this study is to provide researchers, companies or anyoneby best suited algorithm having highest accuracy.wishing to perform predictive analysis with clues that will enable them to choose the best ML method(s) according to its field of application. This study highlighted the most used methods in various fields ofapplication: DT and ANN in education, LR, RF and DT in building, DT in botany, RF and ANN in social science and RF in medicine.

AmritashishBagchi, Shiny Raizada, AniketMhatre and Anthony Augustine. [4] Forecasting the winner of Pro Kabaddi League matches — (2019),A total data of 272 matches were recorded, out of which 32 matches were draw and therefore not included in the study. For the purpose of this study only the first half data was used and in statisticaltechnique Binary Logistic regression was used to predict the outcome of a match (Win/Loss). The result indicates that the developed Logistic regression model was significant.

SudhanshuAkarshe ,RohitKhade , Nikhil Bankar , PrashantKhedkar , Prof. PrashantAhire. [5] Cricket Score Prediction using Machine Learning Algorithms – (2019), In thispaper, they have built a prediction system that takes in data of matches played in past and makes a prediction of future match events such as final score and results in a victory or loss.They intend to use more features such as pitch condition, weather condition, outcome of toss,individual player performance with respect to match venue. The system finally presentquantitative results displayed

VinodNimbalkar, DurveshMundhe, Prof. Dr. SuhasiniVijayKumar. [6] Data Analysis and Visualization on Pro-Kabaddi League – (2018), Data analysis comprises of inspecting data followed by cleaning, transforming and modeling it to emphasize on crucial pieces of information to help in decision making. The paper focuses on Analysis and Visualization of "Pro-Kabaddi League" data. Researchers have created a Deterministic Model using KNN and Logistic Regression algorithms to find the accuracy. The visualization represents different aspects of the game in the graphical manner to aid understanding of data.

Dev Karan Singh, SarthakAgarwal, Sanjeev Gupta, Manisha Singh, UtkarshSaxena.[7] Prediction of match winners of IPL using machine learning algorithms –(2018),Prediction of winner of tournament or match has also a concern area of fans or followers. Onthe other hand technology is evolving at high rate. Machine learning algorithms are alwaysthe first choice of researchers to predict something after training a model. So in this paper weare predicting the winners of Indian Premier League matches using different supervised learning techniques.

III. RELATED WORK

Predicting outcomes in Pro Kabaddi has been the subject of interest for researchers and sports enthusiasts. Various approaches have been explored in this domain. Statistical models have been developed, incorporating factors such as team performance, player statistics, historical data, and match-specific variables. These models aim to assess the

likelihood of a team winning a particular match or the overall tournament. Machine learning algorithms, such as decision trees, random forests, support vector machines (SVM), and artificial neural networks, have been utilized to analyze historical data and relevant features to train predictive models[7]. Some researchers have also delved into sentiment analysis, leveraging social media data to gauge audience sentiment and predict potential winners based on public opinion. Hybrid approaches that combine multiple techniques, including statistical models, machine learning algorithms, and sentiment analysis, have also been employed to enhance predictive accuracy. It is worth noting that the field of sports prediction is dynamic, and staying updated with the latest research, academic databases, sports analytics platforms, and conferences is crucial to accessing the most recent studies and developments in predicting Pro Kabaddi.

IV. PROPOSED METHODOLOGY

The dependent variable chosen for the study was the Match Outcome (Win/Loss). The predictor variables considered were Raid Points, Tackle Points, All Out Points, and Extra Points[5]. These variables were selected based on their potential relevance to predicting match outcomes in the Pro Kabaddi League .For the analysis, we have used past seasondata. Statistical techniques, specifically binary logistic regression, were employed to develop the prediction model. In the proposed system, we consider previous seasons data and implemented the prediction model using Logistic regression. The main aim of the model is to consider full time result of the kabaddi matches to predict the winning team.

Fig 1 ARCHITECTURE OF PROPOSED SYSTEM

Fig: Architecture of Proposed System
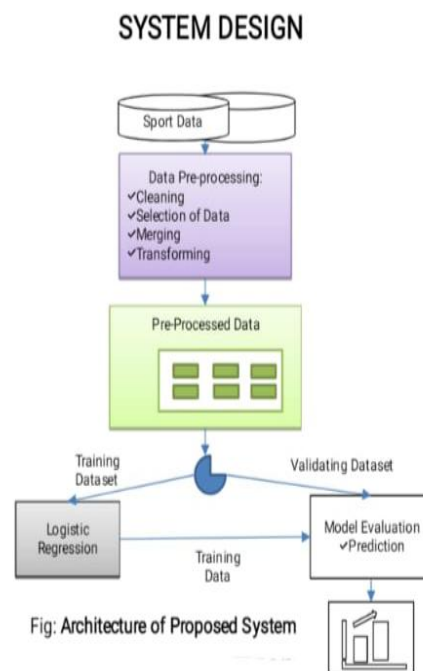
V. WORKFLOW OF PROPOSED SYSTEM

- **Data Collection**: Gather historical data from the Pro Kabaddi League matches, including match outcomes (win/loss), and the predictor variables such as Raid Points, Tackle Points, All Out Points, and Extra Points. Ensure that the data covers a significant number of matches and is representative of different teams and seasons.

- **Data Preprocessing**: Clean the collected data by handling missing values, outliers, and inconsistencies. Perform data transformations, if necessary, such as normalization or standardization, to ensure the variables are on a comparable scale.

- **Exploratory Data Analysis**: Conduct exploratory data analysis to understand the distribution, relationships, and potential patterns in the data. Calculate descriptive statistics, create visualizations, and examine correlations between the predictor variables and the match outcomes.

- **Split the Data**: Divide the dataset into training and testing subsets. The training set will be used to build the prediction model, while the testing set will be used to evaluate its performance.

- **Model Selection**: Choose an appropriate predictive modeling technique for binary outcomes, such as logistic regression, decision trees, random forests, or support vector machines. Consider the specific characteristics and requirements of the Pro Kabaddi League data to determine the most suitable model.

- **Model Development**: Build the prediction model using the training dataset. Implement the selected algorithm and train it on the predictor variables and the corresponding match outcomes.

- **Model Evaluation:** Evaluate the performance of the developed model using the testing dataset. Calculate metrics such as accuracy, precision, recall, and F1 score to assess the model's ability to correctly predict match outcomes. Compare the model's performance against baseline or random models to determine its effectiveness.

- **Model Improvement**: Iterate and refine the model by considering additional factors that may influence match outcomes. Explore feature engineering techniques to create new variables or combinations of variables that could improve the model's predictive power.

- **Documentation**: Summarize the findings, including the methodology, data preprocessing steps, model selection, evaluation metrics, and results. Provide a clear explanation of how the predictor variables contribute to predicting match outcomes.

- **Discussion and Future Work**: Discuss the limitations and potential improvements of the proposed system. Consider factors such as sample size, data quality, and the inclusion of other variables that may enhance the model's predictive accuracy. Highlight potential areas for future research and development.

VI . RESULTS

In Additionally, the correlation matrix was utilized to examine the hypothesis of high multicollinearity among the variables. Multicollinearity occurs when predictor variables in a regression model are highly correlated with each other, which can affect the stability and interpretation of the model. The correlation matrix allows for assessing the strength and direction of the relationships between variables, helping to identify potential multicollinearityissues.By utilizing descriptive statistics and the correlation matrix, researchers can gain insights into the data's characteristics, evaluate the presence of multicollinearity, and make informed decisions regarding the logistic regression model. These analyses help ensure that the assumptions and requirements specific to logistic regression are addressed appropriately.

VII. CONCLUSION AND FUTURE WORK

Kabaddi is a highly strategic game where making conclusions on the outcome of the matches is challenging.So we have made an effort in prediction ofpro kabaddi league winner to help different stakeholders to understand the odds of winning or losing for upcoming matches. The implementation of logistic regression on pro kabaddi data set has an accuracy of 78%.

In the future, there are several enhancements that can be made to Kabaddi prediction models, in which incorporating more features such as player statistics, team rankings, and weather conditions could improve the accuracy of the predictions. Additionally, using more advanced machine learning algorithms, such as deep learning and ensemble methods, could further improve the performance of the model

REFERENCES

[1]. Fátima Rodrigues, Ângelo Pinto 1877-0509 © 2022 "Prediction of football match results with Machine Learning" The Authors. Published by Elsevier B.V.

[2]. [Nikhil Dhonge*1, Shraddha Dhole*2, Nikita Wavre*3, MandarPardakhe*4, AmitNagarale*5 "IPL CRICKET SCORE AND WINNING PREDICTION USING MACHINE LEARNING TECHNIQUES" Volume:03/Issue:05/May2021 Impact Factor- 5.354 www.irjmets.com e-ISSN: 2582-5208

[3]. YashAjgaonkar,Prof. AnaghaPatil,KunalBhoya,Jenil Shah "Prediction of Winning Team using Machine Learning" ISSN: 2278-0181 Published by, www.ijert.org NTASU - 2020 Conference Proceeding

[4]. LoolaBokonda,NissrineSouissi,OuazzaniTouhami Khadija, "Predictive analysis using machine learning: Review of trends and methods" November 2020 DOI: 10.1109/ISAECT50560.2020.9523703

[5]. AmritashishBagchi, Shiny Raizada, AniketMhatre and Anthony Augusthine, "Forecasting the winner of pro kabaddi league matches"2456-0057 IJPNPE 2019; 4(1): 383- 386 © 2019 IJPNPE

[6]. RohithKhade,NikhilBankar,PrashantKhedkar,Prof. PrashantAhire," Cricket Score Prediction using Machine Learning Algorithms GRD Journal for Engineering | Volume 5 | Issue 1 | December 2019 ISSN: 2455-5703

[7]. VinodNimbalkar, DurveshMundhe, Prof. Dr. SuhasiniVijayKumar,"Data Analysis and Visualization on Pro-Kabaddi League" © 2018 IJCRT | Volume 6, Issue 2 April 2018 | ISSN: 2320-2882

[8]. RenatorAmorimTorres "Prediction of nba games based on machine learning methods". University of Wisconsin, Madison,2013

[9]. "A. Tripathi, J. Vanker, B. Vaje, V. Varekar, Cricket Score Prediction system using clustering algorithm

[10]. JassimAlmulla and TanvirAlam, College of science and Engineering, Hamad Bin Khalifa University, Doha,Qatar "Machine learning models reveal key performance metrics of football players to win matches in Qatar stars league" , Published in November 17,2020