REVIEW ON CRICKET ANALYSIS AND PREDICTION USING MACHINE LEARNING APPROACH

[1]Patel Divya Shantilal, [2]Shete Rutuja Arun, [3]Diwani Jay Kantilal, [4]Asst.Prof. Pawar Supriya Haridas

[1]Student, Dept. Of Computer Engineering, Jaihind College of Engineering, Kuran, India
[2]Student, Dept. Of Computer Engineering, Jaihind College of Engineering, Kuran, India
[3]Student, Dept. Of Computer Engineering, Jaihind College of Engineering, Kuran, India
[4]Asst.Professor, Dept. Of Computer Engineering, Jaihind College of Engineering, Kuran, India

Abstract: Many sports may need a tactical assessment, movement analysis, video, and statistical database and modeling, and coach and player data presentations. Large amounts of data were collected in the world of cricket. The data requires to be analyzed for useful insights. Cricket analytic is an ideal market for acquiring innovative players. The Cricket Analysis and Prediction using Machine learning will make use of different machine learning techniques. Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. Predictions like who would be the winner, what would be a score, or which player has the same skills as other players will be made using the Machine learning model. The proposed system will be developed using the Python programming language. Django framework will be used. The Ml learning model will be built and trained using Sklearn and Tensorflow.

***Key Words* - Machine Learning, Supervised learning, Analytic, Cricket.**

I. INTRODUCTION

Sports Analytics is the collection of relevant, historical, and statistical data that can provide a competitive advantage to a team or individual. This data is created using sporting statistics. As far as their significance, these are mostly used in basketball and baseball. The one difference between statistical models used by analysts and advanced statisticmodels, however, is that the latter use probability as their main tool. Examples of statistical models include win shares, win shares of players, win shares of teams, median wins, and so on. These are statistical models that use the methods of mathematics and statistics to describe the output of aset of numbers such as votes in elections or the number of batters in a baseball lineup. These models use heuristics to provide a solution to problems, such as the forecasting of the number of players in a lineup. Cricket analytic is an ideal market for acquiring innovative players. The teams are trying to integrate the latest tools into their infrastructure to improve the performance of their players. Cricket Analytics helps customers understand and predict the performance of cricket teams and players.

In this paper, we propose a system that will help users to find multiple types of insights from cricket data. The proposed system will make use of different machine learning techniques. Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. Predictions like who would be the winner, what would be a score, or which player has the same skills as other players will be made using the Machine learning model. The proposed system will be developed using the Python programming language. Django framework will be used. The Ml learning model will be built and trained using Sklearn and Tensorflow.

II. LITERATURE SURVEY

In [1] the system for cricket analytics and prediction uses machine learning. The paper proposes the use of Lasso Regression for training the regression model. The author of [2] has compared Logistic Regression, Support Vector Machines, K-Nearest Neighbor, and Multi-Layer Perceptron Classifier. The results show that the Multi-Layer Perceptron Classifier performs better than other models used.

III. PROPOSED SYSTEM

In proposed Modified Cricket analytics and prediction using Machine Learning, the area of focus is one game format instead of all. The proposed system will make use of the IPL dataset. The dataset has details of matches and ball by ball deliveries from 2008 to 2020. The dataset will first pass through the cleaning process to use the dataset. The cleaning process is carried out by handling missing values and replacing the same team name variants. The clean data will then be stored in a CSV file. The data will be cleaned using NumPy and pandas library available in python. The CSV file will be used to show descriptive statistics and the training machine learning model. The Modified Cricket analytics and prediction using Machine Learning will present statistics on the Tournament level, Team level, and Players Level. The graphs will be generated using the Plotly express library. The dataset will be used to train four machine learning models. Models will be built to predict the following outcomes.

1.   Which team will win the toss?
2.   Will the team choose to bat or to field if they win the toss?
3.   Which team will win the match?
4.   What would be the score of the first innings of the match?

Each machine learning problem can be broadly categorized into two categories -a classification problem and a regression problem. The above problems can be categorized as

1.   Which team will win the toss?   ---Classification
2.   Will the team choose to bat or to field if they win the toss?   -----Classification
3.   Which team will win the match? ---Classification
4.   What would be the score for the first innings of the match? -----Regression

Before training each model the data will go through pre-processing required for that particular model. The pre-processing includes the following steps
1. Encoding data
2. Separation of features and label
3. Split training and testing set
4. Scaling data

The proposed Modified Cricket analytics and prediction using Machine Learning will make use of the following architectures for training purposes.

1.   Which team will win the toss?   ----------------------------- MLP Classifier
2.   Will the team choose to bat or to field if they win the toss? ----------Random Forest Classifier
3.   Which team will win the match?   -------------------------------------SVC
4.   What would be the score of the first innings of the match? ------------Random Forest Regression

IV. CONCLUSION

The proposed system can prove beneficial in the cricket analytic area. It will also help to monitor the performance of a player. Clustering the same typeof players will become easier.

V. References

[1] Harshit Barot, Arya Kothari, Pramod Bide, Bhavya Ahir, Romit Kankaria," Analysis and Prediction for the Indian Premier League",2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020

[2] P. Sri Harsha Vardhan Goud, Y. Mohana Roopa, B. Padmaja," Player Performance Analysis in Sports: with Fusion of Machine Learning and Wearable Technology" Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4

[3] Manuka Maduranga Hatharasinghe, Guhanathan Poravi"Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Links"2019 5th International Conference for Convergence in Technology (I2CT) Pune, India. Mar 29-31, 2019

[4] Amal Kaluarachchi, Aparna S. Varde," CricAI: A Classification Based Tool to Predict the Outcome in ODI Cricket"978-1-4244-8551-2/10/$26.00 ©2010 IEEE

[5] Deepak Saraswat, Vijai Dev, Preetvanti Singh" Analyzing the performance of the Indian Cricket Team using Weighted Association Rule Mining"2018 International Conference on Computing, Power and Communication Technologies (GUCON) Galgotias University, Greater Noida, UP, India. Sep 28-29, 2018

[6] P.Vijayalakshmi, M.Vandhana, K.Vasanthapriya, P.B.Thushikka, Prediction of Flood by Rainfall using MLP Classifier of Neural Network Model, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 05 | May 20202395-0056