IPL SCORE PREDICTION SYSTEM

Harshada Patel[1], Harshal Patil[2], Bhavik Patil[3], Shubham Patil[4], Nikhil Chintale[5], S H Rajput[6]

Student, Department of Computer Engineering, SSBT COET, Bambhori Jalgoan, Maharashta, India[1-5]

Professor, Department of Computer Engineering, SSBT COET, Bambhori Jalgoan, Maharashta, India[6]

Abstract: Cricket is huge in India, and the Indian Premier League (IPL) is a major cricket tournament that draws players from all over the world. Predicting IPL match outcomes is important for online traders and sponsors. We can do this by looking at various factors like the players' skills, team performance, and match conditions. In our research paper, we proposed using machine learning (ML) algorithms like SVM, Random Forest, Logistic Regression, and K-Nearest Neighbor to predict IPL match outcomes. This research shows that ML techniques can effectively predict IPL match outcomes. As cricket evolves, using advanced technologies like ML for predictions not only makes the game more exciting for fans but also helps teams and stakeholders make better decisions. Our model is a step forward in using ML to understand T Twenty cricket better.

Keywords: Indian Premier League(IPL), Linear Regression,Ridge Regression, Lasso Regression, Machine Learning, IPL Score Prediction.

I. INTRODUCTION

Cricket, a widely-loved outdoor sport played with bat and ball, pits two teams of 11 players against each other. Its popularity spans the globe, with its three main formats—Tests, Twenty-twenty (T20), and One Day Internationals (ODI)—holding prominent spots among the world's most beloved sports.

The Indian Premier League (IPL), established in 2008 and overseen by the Board of Control for Cricket in India (BCCI), is a pinnacle of cricketing excitement. Featuring a 20-20 format, the IPL draws players from regional, national, and international teams and boasts ownership by celebrities, businessmen, and other stakeholders.

With eight teams competing in the 2021 edition—Royal Challengers Bangalore (RCB), Rajasthan Royals (RR), Chennai Super Kings (CSK), Mumbai Indians (MI), Kolkata Knight Riders (KKR), Delhi Capitals (DC), Punjab Kings (PK), and SunRisers Hyderabad (SRH)—the IPL captures the imagination of cricket enthusiasts worldwide.

This paper delves into several intriguing questions surrounding IPL matches: What are the probabilities of winning based on the toss decision to bat or field first? Which bowler records the most dismissals in a match? Does playing at the home ground influence the outcome? Using sophisticated machine learning techniques—such as SVM, Random Forest, and Logistic Regression—we aim to predict IPL match winners based on stadium selections and toss decisions. Through this analysis, we hope to shed light on the intricate dynamics of cricketing contests and contribute to the evolving landscape of sports analytics.

II. LITERATURE SURVEY

Kalpdrum Passi and Niravkumar Pandey conducted a study focusing on the prediction accuracy of two crucial aspects in cricket matches: runs scored by batsmen and the number of wickets taken by bowlers within each team. Their research likely involved analyzing historical data, applying machine learning techniques, and evaluating the performance of predictive models in forecasting these outcomes. The study's findings would provide insights into the effectiveness of such predictive models in cricket analytics and decision-making processes.

P. Wickramasinghe introduces a methodology for predicting batsmen's performance over five years using a hierarchical linear model (HLM). The study entails collecting comprehensive batting data and developing an HLM to model the hierarchical structure of player performance within teams and across seasons. Wickramasinghe validates the model's predictive accuracy and discusses its potential applications in informing team selection and batting strategies based on individual player performance.

Akhil Nimmagadda and colleagues propose a predictive model for cricket matches. They combine Multiple Variable Linear Regression and Logistic Regression to forecast scores. They enhance their model with the Random Forest algorithm for improved accuracy. Their study likely involves collecting cricket match data to include player performance and match conditions. Their goal is to enhance score prediction, contributing to cricket analytics.

III.   DATASET FEATURES

In an IPL score prediction system, datasets are used to analyze historical match data, player statistics, team performance, venue conditions, toss outcomes, and social media sentiments.

Features are engineered from these datasets to train machine learning models. These models are then used to predict match outcomes by considering various factors. The system continuously learns and updates its datasets and models for better accuracy over time.

Score Prediction:-

For conducting our research, we collected data on all the IPL matches played in 2008- 2017. The dataset consists of 76014 numbers of rows. Dataset consists 15 columns over which we applied feature selection techniques and selected 8 features in which 7 are input feature and 1 is our target variable. The attributes selected were bat team, bowl team, overs, runs, wickets, runs in prev 5, wickets in previous 5 for score prediction.

Table 1 : Dataset Attributes and their values

Attributes	Values
Batting Team	Batting Team Name among 8 teams in IPL
Bowling Team	Bowling Team Name among 8 teams in IPL
Overs	Value > 5 Over
Runs	0-300
Wickets	0-10
Run Scored in last 5 overs	0-300
Wickets fall in last 5 overs	0-10
Total Runs	0-300

Feature Selection :-

Feature selection in IPL score prediction involves choosing essential variables like player and team performance metrics, venue factors, toss outcomes, weather conditions, head-to-head statistics, recent form, player availability, match context, and day-night match differences. These factors help build accurate predictive models by capturing key aspects that influence match outcomes

IV.   ARCHITECTURE

The architecture of an IPL score prediction system involves data collection, preprocessing, and feature engineering. Machine learning models are trained on historical data to make predictions. A prediction engine generates real-time predictions based on match parameters. The system integrates with external data sources and provides a user-friendly interface for interaction.

The figure 1 represents the architecture of the model which includes different components like datasets, split data, Training, Testing, Supervised Learning models and Result.
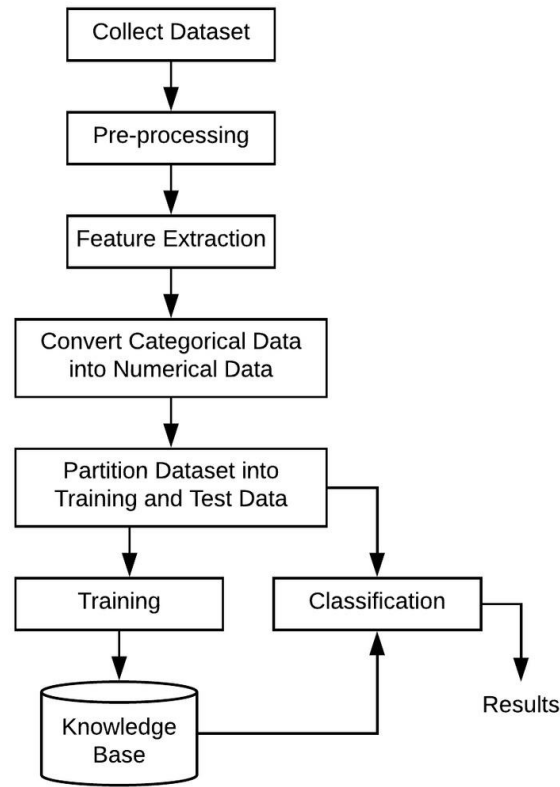
Fig 1. Architecture of the Model.

V. ALGORITHMS

We tried to use three machine learning techniques: regression and classification. Selected algorithms from each technique were trained then.

1.Regression:

Regression analysis is a statistical method used to predict or explain the relationship between a dependent variable (like IPL scores) and one or more independent variables (like player stats or match conditions). It assumes a linear relationship and uses a regression equation to estimate the impact of predictors on the target variable. Common techniques include linear, ridge, and lasso regression, each serving specific purposes like handling multicollinearity or selecting important features. Evaluation metrics like MSE and R-squared assess model performance.

2. Linear Regression

To predict the continuous values, Linear regression is used. Linear regression for IPL score prediction involves gathering historical match data, selecting relevant features like player stats and venue conditions, training a linear regression model on this data to learn the relationship between features and match scores, evaluating the model's performance, and using it to predict scores for upcoming matches.

This process is iterative, with refinements made based on performance evaluations, ultimately leading to the deployment of a predictive model for real-time score prediction. Linear regression is a supervised learning machine learning algorithm. It performs the task to predict dependent variable (y) with respect to other independent variables (x …. Xn). It helps us find out a relationship between the dependent and independent variables.

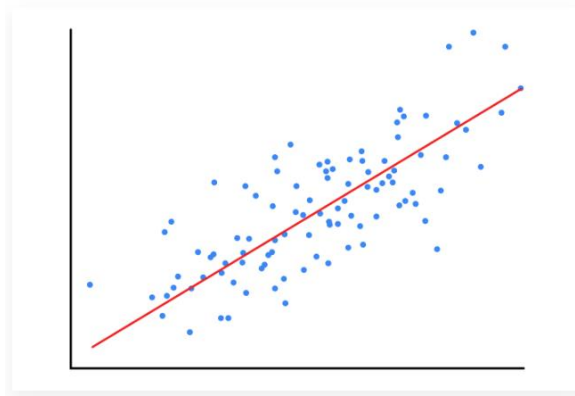The equation is $Y = C + w_1X + …. W_nX_n$

Fig 2. Linear Regression

3. Ridge Regression :

Ridge regression is used in IPL score prediction systems to handle multicollinearity and prevent overfitting. By adding a regularization term to the linear regression equation, it shrinks coefficients towards zero, reducing sensitivity to data fluctuations and improving generalization performance. This makes the model more robust and accurate in predicting match scores.

Fig 3.Ridge Regression

4. Lasso Regression :

Lasso regression in IPL score prediction systems is used for feature selection and regularization. It shrinks less important features to zero, simplifying the model and improving its interpretability. This helps in building more accurate models by focusing on the most relevant predictors for predicting IPL scores**.**
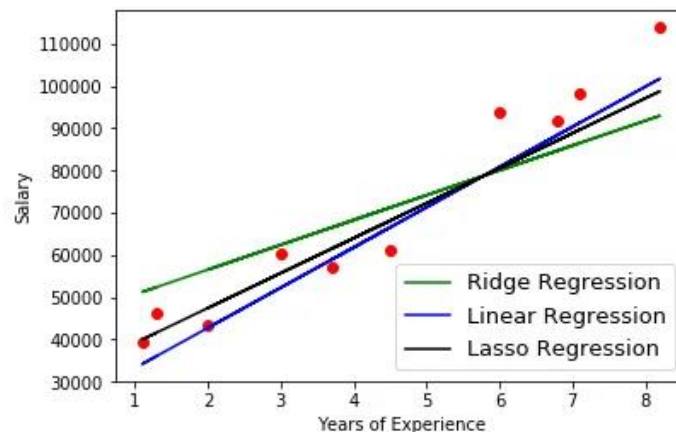

Fig 4.Comparison of different regressions techniques.

VI.   COMPARATIVE ANALYSIS OF ALGORITHM

1. Score Prediction Algorithms:

It is found that the Ridge regression is giving the more accuracy as compared to Linear regression and Lasso regression. For the score prediction the ridge regression gives the highest accuracy result as we see. So, the formula of ridge regression for getting theoretical result is as follows.

The penalty term for ridge regression is $\lambda(\text{slope})^2$, where lambda denotes the degree of deflection from the original curve by restricting the coefficients of predictor variables but never makes them zero.

Therefore the equation for ridge regression is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + \lambda(\text{slope})^2$$

VII.   IMPLEMENTATION OF THE GUI

The Graphical User Interface (GUI) is developed using the Flask Framework, with Python as the backend language. The web application is designed to predict IPL match scores based on data from the last 5 overs of the match.

Upon loading the application, users are presented with the first screen, where they need to enter and select various parameters:

1.   Batting Team : Users select the team that is currently batting.
2.   Bowling Team : Users select the team that is currently bowling.
3.   Cricket Ground : Users choose the venue where the match is being played.
4.   Runs Scored : Users input the total runs scored by the batting team in the match so far.
5.   Overs Bowled : Users specify the total number of overs bowled in the match, ensuring it is greater than 5 to utilize the last 5 overs data.
6.   Runs Scored in the Last 5 Overs : Users enter the total runs scored by the batting team in the last 5 overs.
7.   Wickets Taken in the Last 5 Overs : Users provide the total number of wickets taken by the bowling team in the last 5 overs.

Once users input these parameters, the web application uses machine learning models to predict the expected total score for the batting team based on the provided data. These predictions can assist cricket enthusiasts, analysts, and fans in gaining insights into match dynamics and potential outcomes.
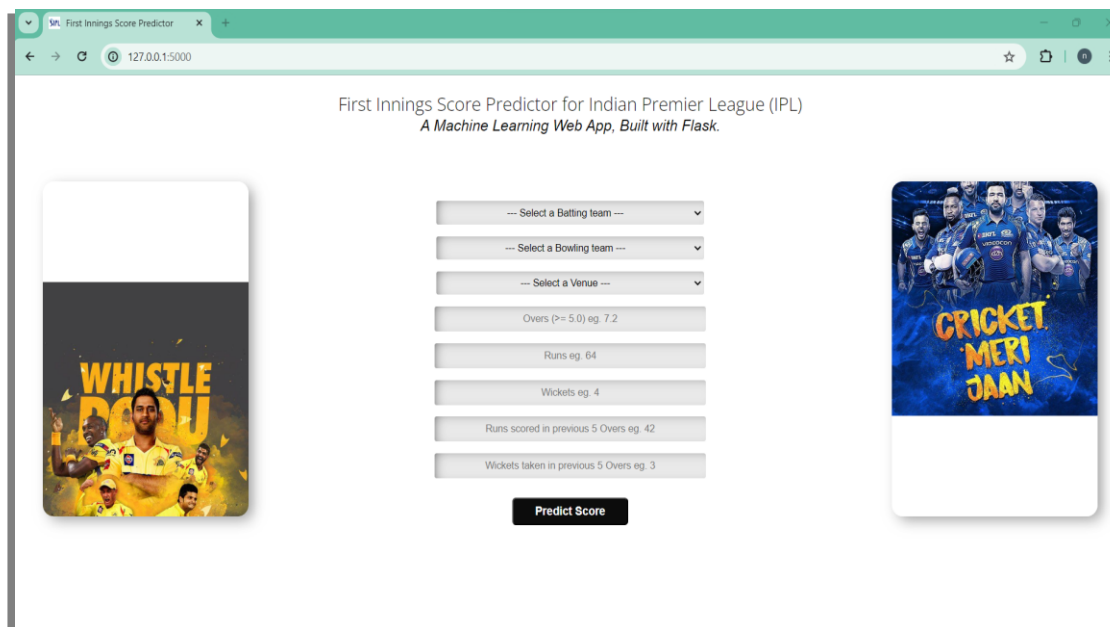

Fig 5.Basic GUI first loaded

Let's enter some random data that fits inside the scope of the project. First let's select the Batting team as Mumbai Indians, and then select the Bowling Team as Chennai Super Kings, then we will select the cricket ground as Eden Gardens Kolkata.

Then enter Runs scored as 104, overs completed as 12.3, wickets down as 2, runs scored in the last five over as 55 and wickets gone in the last five overs as 1 wickets. And see the prediction of score.

Fig 6.Input to Score Prediction Model

VIII.   CONCLUSION

This report will give the important information regarding IPL score prediction, that which parameters are required algorithms. it helps in mathematical operation.

Using all the information we have developed a website for that the important work we have to do for the model is comparative analysis of machine learning techniques that is for score prediction the regressions. In Score Prediction analysis accuracy of Ridge Regression is more than Linear and Lasso Regression.

ACKNOWLEDGMENT

REFERENCES

[1]. Mrs. G V Gayathri, Deepika Gonnabattula, Srinivasa Reddy Gajjala, Manideep Kanakam, Sai Swaroop Medisetty, " Prediction of IPL Match Score and Winner Using Machine Learning Algorithms", Journal of Emerging Technologies and Innovative Research (JETIR) ISSN:2349-5162 , vol. 8, Issue 6 , June 2021.

[2]. KALYAN PAVAN LATCHIPATRUNI, PAVAN SIMHADRI, "ANALYSIS OF IPL MATCH RESULTS USING MACHINE LEARNING ALGORITHMS", International Journal for Research in Applied Science Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429 Volume 9 Issue VI Jun 2021.

[3]. "Prediction of IPL Match Score and Winner Using Machine Learning Algorithms", International Journal of Emerging Technologies and Innovative Research (www.jetir.org),ISSN:2349-5162, Vol.8, Issue 6, page no.c437-c444, June-2021.

[4]. "Prediction on Ipl Data Using Machine Learning   Techniques in R   Package" G. Sudhamathy and G. Raja Meenakshi Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, India CTACT JOURNAL ON SOFT COMPUTING.

[5]. A. Basit, M. B. Alvi, F. H. Jaskani, M. Alvi, K. H. Memon   and R. A. Shah, "ICC T20 Cricket World Cup 2020 Winner Prediction Using Machine Learning Techniques," 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 2020, pp. 1-6, doi: 10.1109/INMIC50486.2020.9318077.

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

[7]. Lokhande, Chawan, Pramila, "Prediction of Live Cricket Score and Winning," International Journal of Trend in Research and Development, Volume 5(1), (2018).

[8]. https://en.wikipedia.org/wiki/List_of_Indian_Premier _League_seasons_and_results.

[9]. https://en.wikipedia.org/wiki/Machine_learning.

[10].   https://www.geeksforgeeks.org/ipl-score-prediction-using-deep-learning/

[11].   https://medium.com/@mehulved1503/feature-selection-in-machine-learning-variable-ranking-and-feature-subset-selection- methods-89b2896f2220.