CRICKET SCORE REDICTION

Dr. Sanjeeva Polepaka
Associate Professor (CSE), G.R.I.E.T, Bachupally,
Hyderabad, TS, India
`email`

Ajith Varma
Research Scholar (CSDS), G.R.I.E.T, Bachupally,
Hyderabad, TS, India
ajithvarma1020@gmail.com

Abhinav Sai Ratan
Research Scholar (CSDS), G.R.I.E.T, Bachupally,
Hyderabad, TS, India
abhinav091202@gmail.com

Valaparla Sathvik
Research Scholar (CSDS), G.R.I.E.T, Bachupally,
Hyderabad, TS, India
sathvik2405@gmail.com

Mohammed shoiab
Research Scholar (CSDS), G.R.I.E.T, Bachupally,
Hyderabad, TS, India
mohammedshoiab1107@gmail.com

ABSTRACT

Cricket is a popular sport that involves a high degree of variability in terms of game conditions and player performance. The ability to accurately predict cricket scores could provide valuable insights for coaches, analysts, and fans, as well as offer opportunities for sports betting and fantasy games. This project explores the use of machine learning techniques to predict cricket scores based on a variety of contextual and historical factors. The publicly available cricket dataset is used to build and evaluate several regression models that predict the total runs scored by a team in a limited-overs cricket match. This analysis includes feature engineering to extract and transform relevant input variables, model selection to compare and choose among different regression algorithms, and performance evaluation to assess the accuracy and robustness of the models. This project also conducts sensitivity analysis to identify the most influential predictors and explore the potential biases and limitations of the models. The results indicate that machine learning techniques can effectively predict cricket scores and provide valuable insights into the factors that contribute to team performance. The findings have implications for cricket teams, coaches, and analysts who seek to improve their game strategies and player selection, as well as for sports betting and fantasy game platforms that seek to provide more accurate and engaging experiences for users.

Keywords: Cricket, Machine Learning, Fantasy games, Analysis, Performance

I. INTRODUCTION

1.1 Introduction

Cricket is a popular sport that involves a high degree of variability in terms of game conditions and player performance. The ability to accurately predict cricket scores could provide valuable insights for coaches, analysts, and fans, as well as offer opportunities for sports betting and fantasy games. This project also conducts sensitivity analysis to identify the most influential predictors and explore the potential biases and limitations of the models. The results indicate that machine learning techniques can effectively predict cricket scores and provide valuable insights into the factors that contribute to team performance. The findings have implications for cricket teams, coaches, and analysts who seek to improve their game strategies and player selection, as well as for sports betting and fantasy game platforms that seek to provide more accurate and engaging experiences for users.

1.2 History

Cricket score prediction has a fascinating history, evolving alongside the sport itself. Early on, cricket enthusiasts relied on their intuition and expertise to predict match outcomes. As technology advanced, statistical analysis and mathematical models emerged, enabling more accurate predictions. The advent of computer algorithms and machine learning further revolutionized score prediction. Today, sophisticated models incorporate a wide range of factors like team performance, player statistics, pitch conditions, weather, and more. These models continuously learn and refine their predictions, enhancing their accuracy over time. Cricket score prediction has become an integral part of the sport, generating excitement and anticipation among fans worldwide.

1.3 Working Principle

The working principle of cricket score prediction involves analyzing a multitude of variables to estimate the likely outcome of a match. Statistical models and machine learning algorithms process historical data, including team and player performance, venue conditions, weather patterns, and more. These models identify patterns, trends, and correlations within the data to generate predictions. Advanced algorithms continuously learn and adapt based on new information and real-time updates during matches. The models consider various factors' relative importance, assign weights to them, and generate a probability-based prediction for the final score or match result. These predictions provide valuable insights and engage fans, enhancing their experience of the game.

1.4 Applications

1. Betting
2. Fantasy sports
3. Match analysis
4. Broadcasting
5. Fan engagement
6. Player performance evaluation
7. Team strategy

1.5 Advantages

1. Enhanced fan experience- Exciting predictions for cricket scores.
2. Informed betting decisions- Reliable predictions for better bets.
3. Fantasy sports strategy- Optimize your fantasy team with accurate predictions.
4. Strategic team decisions- Make informed choices based on predictions.
5. Statistical analysis- Valuable insights for player performance evaluation.

1.6 Challenges and Pitfalls

Cricket score prediction faces several challenges and pitfalls. Firstly, the unpredictability of the game makes accurate predictions difficult, as cricket is influenced by various factors like weather, pitch conditions, player injuries, and unforeseen events. Limited availability of comprehensive and reliable data can also hinder prediction accuracy. Additionally, the emergence of match-fixing and manipulation poses ethical concerns and undermines the integrity of predictions. The complexity of cricket, with its multiple formats and constantly evolving strategies, further adds to the challenge. Lastly, overreliance on predictive models without considering the human element of the game can lead to erroneous predictions. Continuous research, data refinement, and adapting models are essential to overcome these challenges and improve accuracy.

II. EXISTING METHODS

2.1 Overview

Existing approaches for cricket score prediction encompass a range of methodologies. Statistical analysis involves analyzing historical data to identify patterns and trends. Machine learning techniques, such as regression models, decision trees, and neural networks, are employed to learn from past data and generate predictions. Some approaches incorporate complex algorithms that consider multiple factors like player performance, team dynamics, venue conditions, and weather. Data-driven models are continuously refined and updated with real-time information during matches. Additionally, ensemble methods combine the predictions of multiple models to enhance accuracy. These approaches aim to leverage statistical insights and advanced algorithms to provide valuable predictions and enhance the understanding and enjoyment of the game.

2.2 Summary of Existing Approaches

Table 1. Existing Approaches

Reference number	Description	Algorithm	Accuracy
[1]	Score and Winning Prediction in Cricket through Data Mining Oct 8-10, 2015	1.Linear Regression algorithm 2.Naive Bayes classifier	Accuracy of LR is 20% greater than CRR method

[2]	Money Ball - Data Mining on Cricket Dataset 2019	1.Naive Bayes classifier 2.Support Vector Machine 3.K-Nearest Neighbor method 4.Random Forest method	Linear regression 80.76 Ridge regression 80.69 Lasso regression 81.00
[3]	Cricket Match Outcome Prediction Using Tweets and Prediction of the Man of the Match using Social Network Analysis: Case Study Using IPL Data 2018	1.Naive Bayes classifier 2.Support Vector Machine 3.K-Nearest Neighbor method 4.Random Forest method 5.Logistic Regression	1.Sentiment analysis has an accuracy of up to 85%. 2.Tweet based and Mixed model has 89% accuracy. 3.Natural attributes-based model has accuracy up to 83%.
[4]	Cricket Squad Analysis using multiple Random Forest Regression 2019	1. Linear Regression 2. SVR 3. Decision Tree 4. Random Forest	Accuracy can be increased by predicting matrices of the player against the player.
[5]	Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms 2018	1. Linear Regression 2. SVM with linear kernel 3.SVM with polynomial kernel Linear Regression	Accuracy (with 90 % training data): SVC: 43 Decision Tree: 61 Random Forest: 76

2.3 Models Used In Proposed Method

2.3.1 Linear Regression

With the purpose of establishing a linear relationship between a dependent variable and one or more independent variables, linear regression is a well-liked and frequently applied statistical modelling technique. It is a supervised learning algorithm belongs to the moreithm that inclusive regression analysis subset. After the model has been fitted, predictions can be made using fresh, unforeseen data by entering the values of the independent variables. Based on the calculated coefficients and the input values, the model determines the predicted value of the dependent

variable.Relationships between many variables can be studied using linear regression. It may deal with several independent variables, enabling the investigation of more intricate interactions and taking into account various variables that have an impact on the dependent variable.

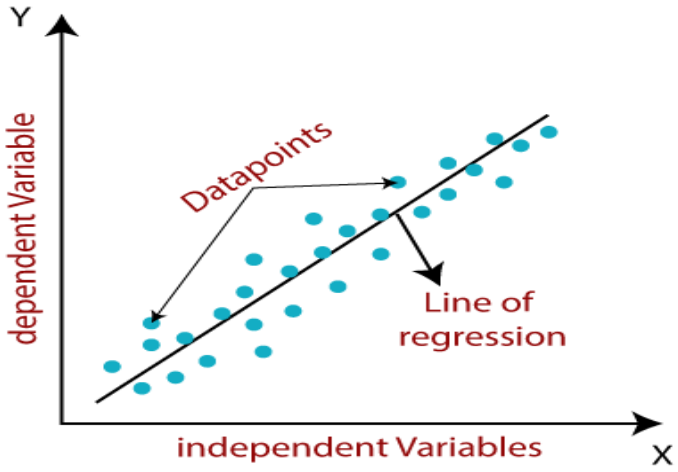

Figure 1. Linear Regression

2.3.2 Logistic Regression

Logistic regression is intended to forecast the likelihood that an event will occur or not, as opposed to linear regression, which tries to predict continuous numeric values. When the dependent variable is binary or dichotomous—that is, when it can only have one of two potential values—such as "yes" or "no," "success" or "failure," or 0 and 1—it is especially helpful. A common statistical modelling method for forecasting probability or binary outcomes is logistic regression. Even though its main objective is classification rather than regression, it is a supervised learning approach that is included in the regression analysis category.

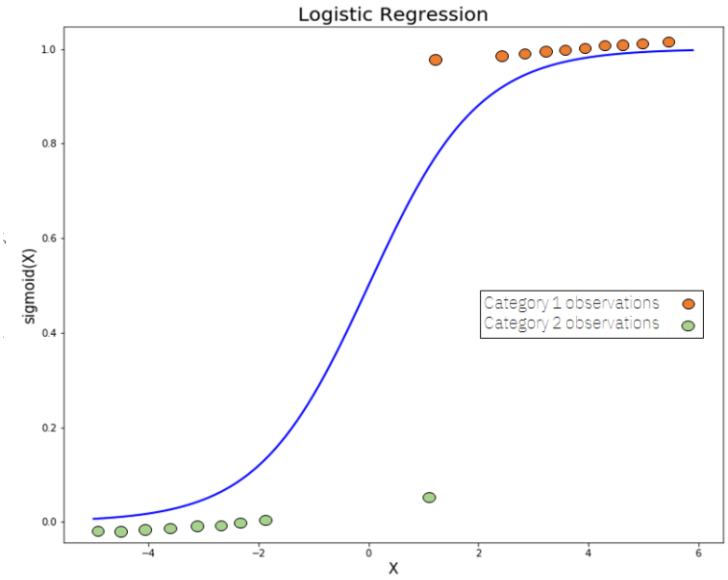

Figure 2. Basic Logistic Regression

2.3.3 Lasso Regression

The objective of conventional linear regression is to reduce the squared sum of errors between the predicted and actual values of the dependent variable. But in lasso regression, an extra penalty component is introduced to the objective function, which is the result of adding the absolute values of the regression coefficients and a regularisation parameter, lambda. By reducing part of the regression coefficients to zero, lasso regression has the main benefit of performing automatic feature selection. By raising the value of, less significant variables suffer more penalties, leading to a sparser model that only retains the most pertinent predictors. Because of this characteristic, lasso regression can be used to handle datasets with a lot of characteristics or when there is a possibility that the predictors may be multicollinear.

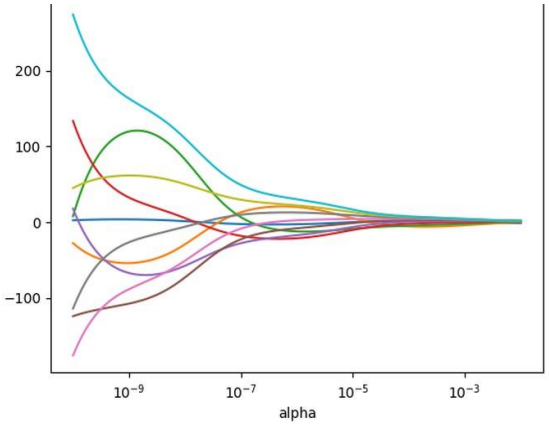

Figure 3. Lasso Regression

2.3.4 Decision Trees

The decision tree algorithm starts with the entire dataset and splits it into smaller groups iteratively based on the values of several attributes. The information benefit is maximised or the impurity measure (such as Gini impurity or entropy) is minimised at each stage of the partitioning procedure. A minimal number of samples are taken from each leaf if a stopping condition is met, such as when a maximum depth is reached. Applying the decision rules along the way, you start at the root node of a decision tree and work your way to the leaf node. The leaf node gives the input instance the anticipated class or value. Because the final model is clear and easy to understand, decision trees have a variety of advantages, one of which is their interpretability.

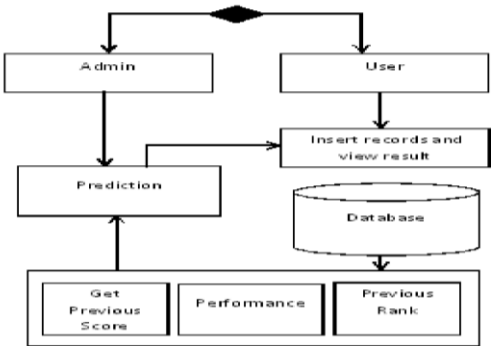

Figure 4. Decision Trees

2.3.5 Random Forest

The strength of several decision trees is combined by the machine learning algorithm known as Random Forest to produce a reliable and precise predictive model. It fits into the ensemble learning area, where various models are merged to enhance overall performance.
The essential tenet of Random Forest is that the individual flaws of each decision tree can be reduced by mixing many trees, each of which has been trained on a distinct sample of the data. The term "Random Forest" refers to the process of randomly choosing the subsets of data and features that will be utilized to construct each decision tree.

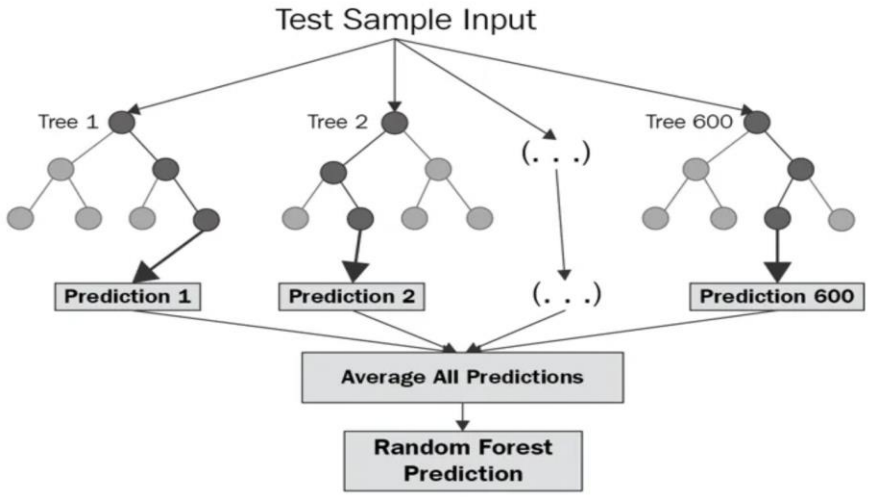

Figure 5. Random Forest

III. PROBLEM STATEMENT AND OBJECTIVES

3.1 Problem Statement

In the profession of cricket score prediction, a variety of strategies are used to forecast the innings score of a cricket match. Numerous systems and prediction computations are used to forecast the outcomes of ODI and T20 cricket matches. When predicting the results of cricket matches, the CRR technique is frequently used. The total number of overs in an inning is multiplied by the number of runs scored in an over in the CRR technique. This method excludes all other factors and only considers runs scored in an over. By accounting for many variables, we are striving to improve the accuracy of the current systems and the predictions. Our objective is to forecast a live game's result.

3.2 Objective

The objective of cricket score prediction is to provide accurate and reliable estimates of the final score or match outcome. By leveraging statistical analysis, historical data, and advanced algorithms, the goal is to enhance fan engagement and excitement during matches. Predictions aid bettors in making informed decisions, fantasy sports players in strategizing their teams, and teams in analyzing opponents and devising strategies. Additionally, score prediction aims to offer valuable statistical insights for player performance evaluation and assist broadcasters in providing informative and engaging commentary. Ultimately, the objective is to leverage predictive models to enhance the overall cricket experience and deepen the understanding of the game.

IV. PROPOSED METHOD

4.1 Project Description

The project aims to develop a cricket score prediction system that utilizes statistical analysis, machine learning, and historical data to generate accurate predictions for match outcomes. The system will consider various factors such as team performance, player statistics, pitch conditions, weather, and other relevant variables. Through the implementation of advanced algorithms and continuous learning, the system will refine its predictions over time. The project will focus on creating an intuitive user interface for easy access to predictions and integrate real-time updates during matches. The ultimate goal is to provide cricket fans, bettors, and fantasy sports players with valuable insights and enhance their overall experience of the game.

4.2 Architecture Diagram

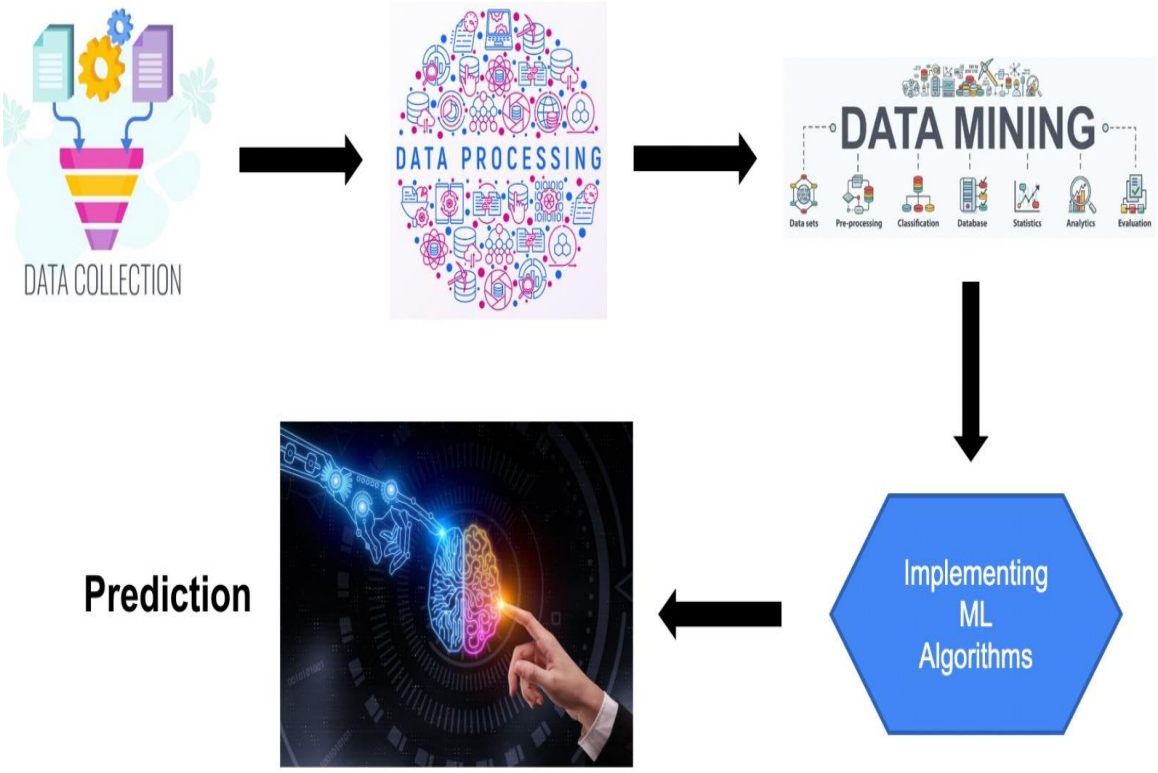

Figure 6. Architecture Diagram

1. **Data Collection** We will obtain the data-set from the Kaggle data-sets. We'll collect the data in CSV format. The following step will involve cleaning the data that was obtained from the website.
2. **Data Cleaning** As part of the data cleaning process, we wish to eliminate all extra columns, such as the match id, location, bowler and batsman names, as well as the striker and non-striker scores. These columns will be omitted as estimation won't need them. Some teams no longer engage in the league, in accordance with IPL data. The IPL does not feature the Deccan Chargers, Gujarat Lions, Pune Warriors India, or Rising Pune Supergiant. As a result, we need to exclude those teams from the data collection and only take into consideration those who are reliable teams. After five overs, we will evaluate the data. The date column, which is available

in the data collection in string format, is the subject of a variety of manipulations that we would like to carry out.

3. **Data Preprocessing** After data has been cleaned, we will need to prepare our data. The process of data preparation will include the one-shot encoding. One hot encoding covers a lot of the actualization portion. We'll need to redesign the columns in our data collection during the preprocessing phase of the data. Adjusting the columns is essential since we need to appropriately arrange our columns in particular series. After gathering the data, we will separate it so that IPL matches played prior to 2016 will be utilized for the model's training and matches played after 2016 will be used as test data.
4. **Algorithms** For the forecast, we'll use the Random Forest Regression, the Linear Regression model, and the Lasso Regression model. The model with the greatest precision will be used to make the prognosis. The model that we are going to use for the prediction will be explained during action chapter.
5. **Prediction** The model will process the data before gathering user inputs. After gathering user inputs and comparing them to historical data, we will be able to estimate a score range, or from lower bound to upper bound. The model architecture of the CFP system is shown in the diagram above.

The data set was encoded in one step. Data must be organized logically for one-hot encoding to occur. We encode data into numbers because many machine learning algorithms cannot function on categorical input. We have columns for the batting and bowling teams in our data gathering. However, our model must be able to understand the user's input when we provide it. The batting and bowling columns both contain several squads. We rely on one-hot encoding because we don't want to provide the batting team and bowling team's input in string format. The model receives the encoded data frame as input. After the user submits the form, the model makes an estimation based on the data. Here we are using many algorithms for perfection like linear regression, decision tree, lasso regression, logistic regression etc.

4.3 Modules and Description

Statistical Models Statistical models use historical data estimate the outcome of a sports event. These models can be used to analyse the output of the play, the scoreline, and the likelihood of different events occurring during the game. Some common statistical models used for sports score prediction include regression analysis, time series analysis, and machine learning algorithms. Data scientists can identify correlations between variables and produce predictions by applying statistical models to raw data to produce accessible visualisations. Common data sets for statistical analysis include census data, public health data, and data from social media. By doing this, you will fully understand every idea from every subject. To find ideas from a set of data, we use statistical models. We can perform modelling on a relatively small sample of data to attempt and understand the core nature of the data. Each statistical model has built-in weaknesses or errors. They are used to approach reality. There are times when the model's underlying assumptions are simply excessively rigid and inaccurate. The traditional instance of using one or more variables to analyse how every explanation variable influences the independent variable is regression.

Scikit-Learn The well-known Python machine learning framework Scikit-Learn can be used to predict cricket results. It offers multiple methods, including decision trees, and random forests, for predicting cricket match results.is the most effective and trustworthy machine learning library for Python. It provides a wide range of efficient techniques for mathematical modelling and machine learning, including regression, classification, clustering, and dimensionality reduction, through a Python programming interface. This library was mainly developed in Python and is based on NumPy and Matplotlib. Some of the most widely recognized model that Sklearn provides are as follows- The Supervised Learning Algorithms Scikit-learn contains almost all popular supervised learning techniques, such as Linear Regression, Support Vector Machine (SVM), Decision Tree, and others. On the other hand, it also encompasses all of the widely used unsupervised learning techniques, such as unsupervised neural networks, factor analysis, PCA, and clustering.

Tensor Flow Tensor Flow is an open-source machine learning library developed by Google. It can be used for cricket score prediction by building a deep neural network model that takes in features such as batting averages, bowling averages, and team rankings. Data sets that are established as computational nodes in a graph-like structure are used by the TensorFlow software. When the edges integrating the nodes in a structure represent multidimensional vectors or matrices, tensors are produced. TensorFlow programmed use a data flow architecture that functions with standardized intermediate outcomes of the calculations, making them especially well-suited to applications involving very large-scale parallel processing, with neural networks providing as a prevalent instance. The framework includes sets of both high-level and low-level APIs. Google recommends employing the high-level ones as often as possible to accelerate the development of data pipelines and application programming.

XG Boost It is an optimized distributed gradient boosting library that can be used for cricket score prediction. It is designed to handle large data-sets and can be used to build a model that takes in various features to guess the output of a game. Understanding XGBoost requires a thorough understanding of the mathematical foundations of decision trees, gradient boosting and supervised machine learning. In supervised machine learning, a model is trained using algorithms for identifying patterns throughout an assortment of characteristics and labels. The model is then applied for predicting the labels on the features of a new dataset. The computational capacity for boosted tree algorithms is increased to its limit by the gradient boosting tool XGBoost. It is exceptionally precise and adaptable. Its main purpose in development was to make machine learning models more effective and efficient. For Python and R, the first XGBoost implementations were created. As a result of its widespread usage, XGBoost now includes package implementations for languages including Scala, Julia, java, Perl, and others. When implementing XGBoost, usability, speed, and performance on big data sets were all given the greatest importance. It doesn't require parameter optimization or modifications; therefore, it may be used right away after downloading with no additional settings.

Simulation Models Simulation models use computer programs to simulate the outcome of a game. These models take into account various factors such as the players' skill levels, the teams' form, and the weather conditions to predict the outcome of a game. Simulator models are frequently known. To put it another way, we are aware of how to take input data, perform a computation, and find the output. Unknown are the inputs. We don't really know the values of the inputs because they are random variables (at least some of them). A probability distribution is created from expert

estimations or it is fitted to the input using historical data. By selecting input values at random and repeatedly calculating the output, the aim is to discover a range of results. A computer program requires a playground to experiment with concepts and learn from its failures and successes. Such a setting could exist in the real world or online. There are no restrictions on simulation models; they are practically entirely free and can be built up in a controlled manner.

V. RESULTS AND DISCUSSIONS

5.1 Description about Dataset

The datasets for our cricket score prediction is from kaggle.com. The Kaggle dataset itself contains multiple datasets that are required for the cricket score prediction using machine learning. Here are some of the datasets that perform the analysis:

1. Dataset Players
2. Dataset with Venues
3. Dataset with Player's performance
4. Dataset with Average score at venue
5. Dataset with Player's scores at individual venues

These Datasets comprise all the Cricket Matches in a particular interval of time. It consists of these files

1. Original Dataset.csv Raw Dataset File which I scraped using Pandas
2. CategorialDataset.csv Categorial Features suitable for models like MLP Classifier & DT Classifier
3. Continuous Dataset.csv Purely for Experimental Purposes
4. Labelled Dataset.csv Suitable for Support Vector Machines

5.2 Experimental Results

1. The data set contains data from the previous five years to forecast scores. The data used for victory predictions spans a period of seventeen years.
2. We split the data into training and testing portions at a ratio of 70 to 30. 70% of the data is used for training, while 30% is tested.
3. Training - Training can be carried out using training data obtained from the data set. The system will learn from the data about the pattern and different relationships with the aid of this training.
4. Testing - Data testing is done to determine if the training phase was successful or not. The testing data is used to test the data after the training to see whether the machine learning algorithm's prediction or computation was correct or incorrect.
5. Supervised machine learning is a technique used to train an algorithm to carry out the same job on a variety of data sets to uncover patterns and relationships. To train the system, supervised learning provides data with examples and results.
6. Naive Bayes: They are very ascendable and call for a set of parameters that are linearly spaced out from the number of variables in a learning problem. Most of the time, maximum-likelihood coaching involves linearly evaluating a closed-form phrase.
7. To produce a better forecast, linear regression repeats the same task repeatedly. Future values are predicted using this model.
8. Score model: This model displays a numeric number that has been calculated and predicted using a variety of algorithms.
9. Evaluate model - This model is used to evaluate whether the prediction is right or wrong.

Out[6]:

	Unnamed: 0	match_id	venue	innings	ball	batting_team	bowling_team	striker	non_striker	bowler	runs_off_bat	extras	wicket_type	player_dism
0	0	335982	M Chinnaswamy Stadium	1	0.1	Kolkata Knight Riders	Royal Challengers Bangalore	SC Ganguly	BB McCullum	P Kumar	0.0	1.0		
1	1	335982	M Chinnaswamy Stadium	1	0.2	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	SC Ganguly	P Kumar	0.0	0.0		
2	2	335982	M Chinnaswamy Stadium	1	0.3	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	SC Ganguly	P Kumar	0.0	1.0		
3	3	335982	M Chinnaswamy Stadium	1	0.4	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	SC Ganguly	P Kumar	0.0	0.0		
4	4	335982	M Chinnaswamy Stadium	1	0.5	Kolkata Knight Riders	Royal Challengers Bangalore	BB McCullum	SC Ganguly	P Kumar	0.0	0.0		

Figure 7. This is an experimental result of a match between Royal Challengers Bangalore and Kolkata Knight Riders.

Out[3]:

	Team	Player	Tournament	Matches	Batting Innings	Not Out	Runds Scored	Highest Score	Batting Average	Balls Faced	...	Runs Conceded	Wickets Taken	Best Bowling Figures	Bowling Average	Bowling Economy Rate	Bowling Strike Rate
0	Delhi Daredevils	CH Morris	IPL 2016	12	7	4	195	82*	65.00	109	...	308	13	2/30	23.69	7.00	20.3
1	Delhi Daredevils	CH Morris	IPL 2017	9	9	4	154	52*	30.80	94	...	240	12	4/26	20.00	7.74	15.5
2	Delhi Daredevils	CH Morris	IPL 2018	4	4	3	46	27*	46.00	26	...	143	3	2/41	47.66	10.21	28.0
3	Delhi Daredevils	JP Duminy	IPL 2016	10	8	3	191	49*	38.20	156	...	55	2	1/4	27.50	7.85	21.0
4	Delhi Daredevils	Q de Kock	IPL 2016	13	13	1	445	108	37.08	327	...	-	-	-	-	-	-

Figure 8. This is the experimental result of performances of Delhi Daredevils players.

Out[16]:

	venue	innings	ball	batting_team	bowling_team	striker	non_striker	bowler	run	wickets	...	Runs Conceded	Wickets Taken	Best Bowling Figures	Bowling Average	Bowling Economy Rate	Bowlin Strik Rat
0	15	1	0.1	7	13	186	30	201	1.0	0.0	...	1	1	1	1	1	
1	15	1	0.2	7	13	30	184	201	0.0	0.0	...	0	0	0	0	0	
2	15	1	0.2	7	13	30	184	201	0.0	0.0	...	0	0	0	0	0	
3	15	1	0.2	7	13	30	184	201	0.0	0.0	...	0	0	0	0	0	
4	15	1	0.3	7	13	30	184	201	1.0	0.0	...	0	0	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
5186	31	1	6.5	14	10	127	47	220	1.0	1.0	...	0	0	0	0	0	
5187	31	1	6.5	14	10	127	47	220	1.0	1.0	...	0	0	0	0	0	
5188	31	1	6.6	14	10	48	126	220	1.0	1.0	...	0	0	0	0	0	
5189	31	1	6.6	14	10	48	126	220	1.0	1.0	...	0	0	0	0	0	
5190	31	1	6.6	14	10	48	126	220	1.0	1.0	...	0	0	0	0	0	

Figure 9. Output

5.3 Significance of the Proposed Method

The proposed methods are statistical models that include regression analysis, time series analysis and machine learning algorithms. The second one is scikit learn it is one of the libraries used in python. The next one is tensor flow which is a machine learning library. The next one is XG boost which is a boosting library. And the last one is Simulation Models which use computer programs to simulate the outcomes of a game. Below are some advantages of cricket score prediction.

1. We can predict the outcomes of a match even before a match starts.
2. It helps us in changing the game plans and strategies for the team coaches and supporting staff.
3. Model that can be updated along with the instant modifications and changes.
4. Takes all important features into account the players playing in each match.
5. Good accuracy arrived through both prediction models.
6. It may prove helpful for numerous stakeholders to use machine learning to study cricket matches while taking, player performance, archaeological game data, ecological criteria, preceding game conditions, and other features into account.

VI. CONCLUSION AND FUTURE ENHANCEMENTS

6.1 Summary

Currently, there is an algorithm that can figure out the team's final score based on the present run rate. It doesn't take into consideration the players' performance or other factors. There are two models in this system. The first model uses the present scenario to anticipate the score a team will receive after the inning. The second model makes victory percentage predictions for both sides prior to the contest even beginning. The research is carried out using historical data in this

case. Data preprocessing, visualization of data, data preparation, feature selection, and the implementation of various machine learning algorithms are just a few of the data science subfields that will converge.

1. To improve the general attraction to the Premier League.
2. To predict the cricket score.
3. Effective prediction technique.
4. Essential for making strategic decisions.

Cricket score prediction seeks to benefit many stakeholders by offering insightful information, supporting their decision-making, and improving their overall comprehension and enjoyment of the game. It's vital to remember that cricket score prediction should be considered entertainment or an analytical tool rather than a trustworthy way to anticipate how a match will turn out. The fluid nature of the game and unforeseen occurrences that can happen during a match cannot be taken into account by predictions, even though they can offer insights and probabilities based on historical data and trends. Because of these drawbacks and uncertainties, it is advised to proceed with caution while making cricket score predictions. Instead of relying exclusively on predictions, it is always preferable to appreciate the sport and its unpredictable nature.

6.2 Importance

Cricket score prediction holds significant importance in various aspects of the sport. Firstly, it adds excitement and engagement for fans, making the viewing experience more thrilling. Predictions also play a crucial role in betting, helping bettors make informed decisions and increasing the potential for successful bets. In fantasy sports, accurate score predictions aid players in strategizing their teams, enhancing their chances of success. For teams and coaches, score predictions offer valuable insights into opponents, enabling strategic decision-making and planning. Broadcasters benefit from predictions by providing informative commentary and analysis. Overall, cricket score prediction contributes to the sport's popularity, fan involvement, strategic planning, and informed decision-making.

6.3 Adopted Method

One commonly adopted method for cricket score prediction involves a combination of statistical analysis and machine learning techniques. Initially, historical data is collected, including information on team performance, player statistics, pitch conditions, weather, and other relevant factors. Statistical analysis is applied to identify patterns, trends, and correlations within the data. Machine learning algorithms, such as regression models, decision trees, and neural networks, are then employed to learn from the historical data and generate predictions. These models consider multiple variables and assign weights based on their relative importance.

To improve accuracy, ensemble methods are often utilized, which combine the predictions of multiple models. This helps in mitigating the potential shortcomings of individual models and increases overall prediction reliability. Continuous learning and adaptation are crucial, with models updated and refined as new data becomes available. Real-time updates during matches, including live scores and player performance, are incorporated to adjust predictions on-the-fly. The adopted method emphasizes the integration of statistical insights, advanced algorithms, and continuous learning to generate reliable predictions for cricket scores.

6.4 Results

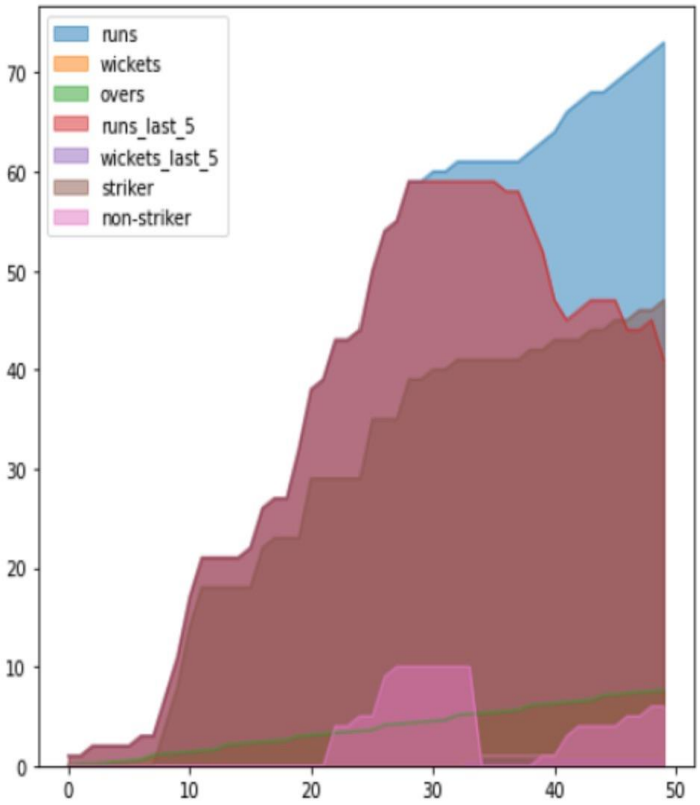

Figure 10. Result Graph-1

```
Out[38]: 9.314617116412084
```

Figure 11. Result Output-1

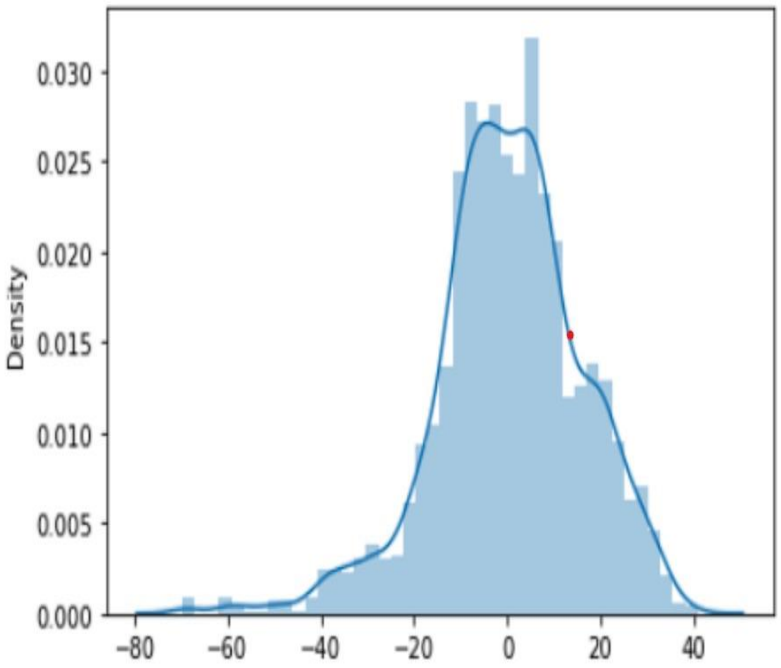

Figure 12. Result Graph-2

```
Out[39]: 11.856571876802397
```

Figure 13. Result Output-2

6.5 Future Scope

The future scope for cricket score prediction is promising, with several areas offering opportunities for further development.

1. Advanced Machine Learning Techniques- Incorporating more advanced machine learning algorithms, such as deep learning and reinforcement learning, can enhance prediction accuracy by capturing complex patterns and dynamics within the game.
2. Real-time Data Integration- Leveraging real-time data feeds during matches, including player performance, ball-by-ball updates, and social media sentiment analysis, can improve the timeliness and accuracy of predictions.
3. Integration with IoT and Wearable Technology- Integrating data from IoT devices and wearable technology worn by players can provide additional insights, such as player fitness levels, biometric data, and real-time player tracking, leading to more accurate predictions.

4. Sentiment Analysis- Incorporating sentiment analysis of fan opinions, social media trends, and expert commentary can provide a deeper understanding of the game's emotional and psychological aspects, enhancing prediction models.
5. Predictive Analytics for Team Management-Utilizing score predictions for team management, such as identifying optimal batting orders, setting fielding strategies, and making in-match decisions, can provide a competitive edge to teams.
6. Personalized Fan Experience- Tailoring score predictions to individual fan preferences and providing interactive platforms for fan engagement can further enhance the overall cricket experience.

6.6 Future Enhancements

1. We can use a model to predict the possibility of chasing in the future. In other words, the algorithm may be able to predict whether a team would be effective in achieving the goal.
2. The model utilized in this project can be made more accurate. The prediction can take into account variables like the venue, the playing surface, and the opposition team.
3. Additional factors such as batsmen partnerships and pitch conditions, can be introduced to further improve the model's accuracy.

REFERENCES

[1] Available at https://www.semanticscholar.org/paper/CRICKET-MATCH-OUTCOME-PREDICTION-USING-MACHINE-Naik-Pawar/4fc5b8dec6ee0281a63c75f93e4a040464e52c2a

[2] Available at https://ijirt.org/master/publishedpaper/IJIRT157821_PAPER.pdf

[3] Prasad Thorat, Vighnesh Buddhivant, Yash Sahane; Review Paper on Cricket Score Prediction; April 2021

[4] Tejinder Singh, Vishal Singla, Parteek Bhatia; - Score and Winning Prediction in Cricket through Data Mining Oct 8-10,2015

[5] D. Thenmozhi, P. Mirunalini, S. M. Jaisakthi, Srivatsan Vasudevan , Veeramani Kannan V, SagubarSadiq S; Moneyball - Data Mining on Cricket Dataset; 2019

[6] A.N.Wickramasinghe, Roshan D.Yapa; Cricket Match Outcome Prediction Using Tweets and Prediction of the Man of the Match using Social Network Analysis: Case Study Using IPL Data; 2018

[7] Nigel Rodrigues1, Nelson Sequeira2, Stephen Rodrigues3, Varsha Shrivastava4; Cricket Squad Analysis using multiple Random Forest Regression;2019

[8] Animal Islam Anik, Sakif yeaser, A.G.M. Emam Hussain, Amitabha Chakraborty; Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms;2018

[9] Siyamalan Manivannan, Mogan Kausik; Convolutional Neural Network and Feature Encoding for Predicting the Outcome of Cricket Matches;2019

[10] Manuka Madranga Hatharasinghe, Guhanathan Poravi Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Link;2019

[11] Jalaz Kumar, Rajeev Kumar, Pushpender Kumar; Outcome Prediction of ODI Cricket Matches using Decision Trees and MLP.

[12] Pallavi Tekade, Kunal Markad, Aniket Amage and Bhagwat Natekar (2020). “CRICKET MATCH OUTCOME PREDICTION USING MACHINE LEARNING” .

[13] Prof. R. R. Kamble, Nidhi Koul, Kaustubh Adhav, Akshay Dixit and Rutuja Pakhare (2021). “Cricket Score Prediction Using Machine Learning” .
[14] Rohit Khade, Nikhil Bankar, Prashant Khedkar and Prof. Prashant Ahire (2019). “Cricket Score Prediction using Machine Learning Algorithms” .
[15] Omkar Mozar, Soham More, Shubham Nagare and Prof. Nileema Pathak (2022). “Cricket Score and Winning Prediction” .
[16] Dhonge, N., Dhole, S., Wavre, N., Pardakhe, M., & Nagarale, A (2021). “IPL Cricket Score and Winning Prediction Using Machine Learning Techniques”.
[17] Kumash Kapadia, Hussein Abdel-Jaber, Fadi Thabtah, Wael Hadi (2019). “Sport analytics for cricket game results using machine learning: An experimental study”.
[18] Daniel Mago Vistro, Faizan Rasheed, Leo Gertrude David (2019). “The Cricket Winner Prediction With Application Of Machine Learning And Data Analytics”.
[19] Prasad Thorat, Vighnesh Buddhivant, Yash Sahane (2021). “CRICKET SCORE PREDICTION”.
[20] Apurva Lawate, Nomesh Katare, Salil Hoskeri, Santosh Takle, Prof. Supriya. B. Jadhav (2021).