Cric Pro - Fantasy Points Prediction Using Gradient Boosting Algorithms

A Project Report submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science and Engineering

by

Aryan Shukla (112115028) , Aryaman Rajes Engglay (112115026), Mir Tahfiz Hafiz (112115090), Abhay Raj Singh (112115003) and Akshat Jaiswal (112115017),
Under the Supervision of: Dr. Anagha Khiste

Semester: 5th

Name of Department: Department of Computer Science and Engineering

Indian Institute of Information And Technology, Pune

(An Institute of National Importance by an Act of Parliament)

November 2023

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **"Cric Pro - Fantasy Points Prediction Using Gradient Boosting Algorithms."** submitted by **Aryan Shukla** bearing the **MIS No: 112115028**, **Aryaman Rajes Engglay** bearing the **MIS No: 112115026**, **Mir Tahfiz Hafiz** bearing the **MIS No: 112115090**, **Abhay Raj Singh** bearing the **MIS No: 112115003** and **Akshat Jaiswal** bearing the **MIS No: 112115017**, in completion of their project work under the guidance of **Dr. Anagha Khiste** is accepted for the project report submission in partial fulfillment of the requirements for the award of the degree of
Bachelor of Technology in the **Department of Computer Science and Engineering**, Indian Institute of Information Technology, Pune (IIIT Pune), during the academic year **2022-23**.

Dr. Anagha Khiste
Project Guide
Assistant Professor
Department of CSE
IIIT Pune

Dr. Sanjeev Sharma
Head of the Department
Assistant Professor
Department of CSE
IIIT Pune

Project Viva-voce held on: 6TH November 2023

Undertaking for Plagiarism

We **Aryan Shukla** bearing the **MIS No: 112115028**, **Aryaman Rajes Engglay** bearing the **MIS No: 112115026**, **Mir Tahfiz Hafiz** bearing the **MIS No: 112115090**, **Abhay Raj Singh** bearing the **MIS No: 112115003** and **Akshat Jaiswal** bearing the **MIS No: 112115017** solemnly declare that research work presented in the **report** titled **"Cric Pro - Fantasy Points Prediction Using Gradient Boosting Algorithms."** is solely **our** research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete report has been written by **us**. I understand the zero tolerance policy of **Indian Institute of Information Technology Pune** towards plagiarism. Therefore **we** declare that no portion of my **report** has been plagiarized and any material used as reference is properly referred/cited. I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of the degree, the Institute reserves the rights to withdraw/revoke my **B.Tech** degree.

Aryan Shukla (112115028) 6rd November 2023

Aryaman Rajes Engglay (112115026): 6rd November 2023

Mir Tahfiz Hafiz (112115090): 6rd November 2023

Abhay Raj Singh (112115003): 6rd November 2023

Akshat Jaiswal (**112115017)**: 6rd November 2023

Conflict of Interest

Manuscript title: <u>Cric Pro - Fantasy Points Prediction Using Gradient Boosting Algorithms.</u>

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Aryan Shukla (112115028) 6rd November 2023

Aryaman Rajes Engglay (112115026): 6rd November 2023

Mir Tahfiz Hafiz (112115090): 6rd November 2023

Abhay Raj Singh (112115003): 6rd November 2023

Akshat Jaiswal (**112115017):** 6rd November 2023

ACKNOWLEDGEMENT

This project would not have been possible without the help and cooperation of many. I would like to thank the people who helped me directly and indirectly in the completion of this project work.
First and foremost, I would like to express my gratitude to our honorable Director, **Prof. O.G. Kakde**, for providing his kind support in various aspects. I would like to express my gratitude to my project guide **Dr. Anagha Khiste**, **Department of CSE**, for providing excellent guidance, encouragement, inspiration, constant and timely support throughout this **B.Tech Project**. I would like to express my gratitude to the **Head of Department (Dr. Sanjeev Sharma)**, **Department of CSE**, for providing his kind support in various aspects. I would also like to thank all the faculty members in the **Department of CSE** and my classmates for their steadfast and strong support and engagement with this project.

Abstract

The rapid evolution of online gaming platforms has significantly transformed the landscape of sports engagement. This project delves into the realm of fantasy cricket, an immensely popular online gaming genre where participants create virtual teams comprising real-life cricket players and earn points based on the players' performance in actual matches. The objective of this research is to enhance user experience and engagement on fantasy cricket platforms through the integration of predictive analytics.

The project focuses on developing a predictive model using machine learning algorithms, historical player data, and match statistics. By analysing players' past performances, team dynamics, weather conditions, pitch reports, and other relevant variables, the model predicts the best fantasy team that includes player from both the teams who are likely to perform the best in the game. These predictions are then utilized to assist fantasy cricket players in making informed decisions while selecting their teams, maximizing their chances of winning.

In addition to the predictive model, the project emphasizes user interface optimization. User-friendly dashboards designed to present the predictions in an easily understandable manner.

This project not only contributes to the gaming industry by enhancing the fantasy cricket experience but also showcases the potential of predictive analytics in revolutionizing user engagement strategies. By empowering users with data-driven insights, this project aims to create a vibrant online community of cricket enthusiasts, fostering a sense of camaraderie and excitement among players.

Keywords: *Fantasy Cricket, Team prediction, Machine Learning Algorithms*

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Overview of Work

The project aims to enhance the online fantasy cricket gaming experience by integrating predictive analytics and user-friendly features. Using advanced computer algorithms, the research focuses on predicting match outcomes and player performances, enabling players to make informed team selections. The project also emphasizes creating a sense of community among players by incorporating interactive features like leagues and challenges. Through real-time updates and personalized suggestions, the platform aims to empower users, making their gaming experience engaging and enjoyable. By combining accurate predictions with a user-friendly interface, this work aims to elevate the overall fantasy cricket experience, fostering a vibrant and competitive online gaming community.

1.2 Motivation of the Work

The motivation behind embarking on this project stems from the burgeoning fantasy sports market and the transformative potential of advanced data analytics. With the escalating global interest in online fantasy cricket, there is a pressing need to revolutionize the user experience. Leveraging cutting-edge data analytics techniques, the project aspires to empower users by offering them accurate match predictions and player analyses. Informed decision-making not only heightens the thrill of the game but also substantially improves users' chances of success, fostering a loyal user base.

Moreover, the project aims to foster a sense of community among cricket enthusiasts by incorporating interactive features such as leagues and challenges, thereby enhancing user engagement and camaraderie. By embracing the challenge of enhancing user interaction through predictive analytics, this project seeks to make significant strides in the dynamic and competitive landscape of online fantasy cricket gaming.

1.3 Literature Review

A wealth of literature is accessible concerning the field of Sports Analytics and its practical implementation in decision-making for sports such as Baseball, Football, Basketball, Soccer, and Tennis. Nonetheless, the quantity of significant research articles dedicated to the utilization of Sports Analytics within the domain of Cricket remains limited. In this section, we will provide an outline of the progression of Sports Analytics, tracing its evolution from its origins in Baseball to its application in the realm of Cricket. We will also explore the diverse methodologies employed by various authors to predict outcomes and make informed decisions related to team selections in the sports domain.

An IEEE published paper titled "Cricketer's tournament-wise performance prediction and squad selection using machine learning and multi-objective optimization" talks about how selecting a well-rounded squad is crucial in cricket, requiring consideration of player characteristics for versatility in long-duration events, whether home or away. It also talks about how to streamline this process for computational efficiency. Machine learning algorithms have made significant progress in predicting player performance based on data. This paper focuses on accurately forecasting cricket player performance by evaluating various machine learning algorithms.

Another paper titled "Batsmen Performance Prediction Using Regression Models" discusses how the player performance is contingent on multiple factors, such as the opposition team, venue, and the number of balls faced by the batsman. It employs diverse Regression models, including Linear Regression (LR), Decision Tree Regression (DT), Random Forest (RF) Regression, and Support Vector Regression (SVR) to offer an engaging tool to interact with statistics, and construct ideal fantasy sports teams using regression predictive models.

Journal/Author	**Date of Publishing**	**Comments**
Batsmen performance prediction using regression models **Gayatri Bhatambarekar, Swarnim Rai**	**2022**	1.Dosen't differentiate between different formats 2.has used random forest and SVM algorithm
Cricketer's tournament-wise performance predictic squad selection using machine learning and multi-objective optimization **Devopriya Tirtho , Shafin Rahman, Md. Shahriar Mahbub**	**2022**	1.Has taken form, consistency and opposition factors into account. 2.Used linear regression, SVM ,random forest
Cricket Match Analytics Using the Big Data Approach **MazharJaved Awan**	**2021**	Used for analytics and score prediction not for team selection

Syed Arbaz Haider Gilani		
DATA SCIENCE APPROACH TO PREDICT TH WINNING FANTASY CRICKET TEAM—DREA FANTASY SPORTS **Sachin Kumar S ,** **Prithvi H V [2] ,** **C. Nandini [**3]	2020	Used multioutput regression algorithm which is not very accurate

1.4 Research Gap

1. Traditional Approaches:

Current reliance on conventional classification or predictive algorithms in existing fantasy cricket platforms for team selection. These methods have limitations in terms of predictive accuracy and might not fully capture the complexities of player performances and match outcomes.

2. Innovative Algorithm Selection:

Adopting novel algorithms such as Extra Trees, Cat-Boost, and XG-Boost for team selection in the proposed model. These algorithms are known for their ability to handle complex, non-linear relationships within data, providing a more nuanced approach to predicting player performances and match results.

3. Shift Towards Regression Techniques:

Departure from solely classification methods, emphasizing the adoption of regression algorithms, specifically Gradient Boosting Regression variants, which are relatively underexplored in this context. Regression techniques allow for the prediction of numerical values, enabling a more precise estimation of player scores and match outcomes.

4. Specific Algorithmic Focus:

Utilization and effectiveness of Extra Trees, Cat-Boost, and XG-Boost specifically within the fantasy cricket domain, underscoring the unique contribution of this research.

5. Need for Comparative Analysis:

There is absence of comparative studies evaluating the performance of these advanced algorithms against traditional techniques, indicating a gap in understanding their efficacy for fantasy cricket team selection.

6. Enhanced Predictive Accuracy:

The potential of these advanced algorithms to offer improved predictive accuracy and efficiency, suggesting a shift from previous models that might have limitations in this regard. This higher accuracy is crucial for fantasy cricket enthusiasts who rely on these platforms for competitive gaming.

7. Application of Regression in Fantasy Sports:

The existing scarcity of research employing regression algorithms, especially Gradient Boosting Regression variants, in fantasy sports, showcasing the pioneering nature of this study in this specific area.

8. Opportunity for Comprehensive Analysis:

The opportunity to conduct an in-depth comparative analysis between classification and regression algorithms, demonstrating the unique advantages offered by the proposed Extra Trees, Cat-Boost, and XG-Boost models for fantasy cricket team selection. The study contributes valuable insights that can shape the future of predictive modeling in fantasy cricket, providing users with more reliable and accurate decision-making tools

Chapter 2
Problem Statement

In the realm of online fantasy cricket gaming, users often face challenges choosing right players due to the complexity of match dynamics and player performances. Current platforms lack reliable predictive tools, leaving users to rely on intuition and limited information when forming their fantasy teams. Additionally, there is a lack of interactive and community-building features, hindering the development of a vibrant user community. This project addresses these issues by aiming to develop an accurate predictive model for player performances. Furthermore, the project aims to create an engaging and interactive environment by implementing user-friendly interfaces and social features. By addressing these challenges, the project strives to enhance user experience, boost user confidence in team selections, and foster a strong sense of community among fantasy cricket enthusiasts.

2.1. Research Objectives and Scalability Prospects

- Develop a sophisticated predictive model using advanced machine learning algorithms and historical cricket data to accurately forecast match outcomes and player performances.

- Enhance the user experience by refining the platform's interface, ensuring real-time updates, and providing personalized suggestions for users.

- Implement social features that facilitate community building, allowing users to interact, share strategies, and compete against friends, fostering a sense of camaraderie.

- Continuously evaluate and refine the predictive model using historical data to enhance its accuracy and reliability over time.

- Empower users by presenting predictive insights in an easily understandable manner, enabling informed decision-making during team selections.

- Implement social features that facilitate community building, allowing users to interact, share strategies, and compete against friends, fostering a sense of camaraderie.

- Ensure scalability of the platform, accommodating a growing user base, and guarantee accessibility across diverse devices and internet connections for an inclusive user experience.

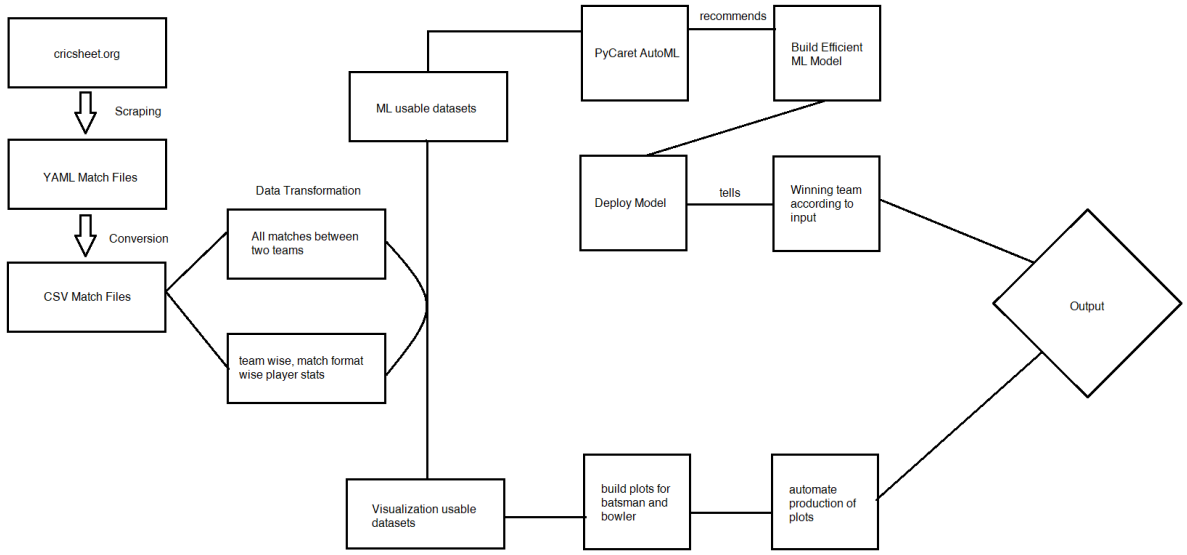
Through these objectives, the project aims to create an engaging, user-friendly, and community-driven online

fantasy cricket platform.

2.2. Methodology of the Work

The methodology section of the paper will be divided into five distinct sub-sections to elaborate on various aspects. These include:

1. **Getting Data:** Explaining the process of data acquisition.
2. **Feature Engineering and Machine Learning:** Describing the techniques used for enhancing features and the machine learning model.
3. **Data Engineering:** Detailing the methods employed for data processing.
4. **Dream Team Recommending Module:** Discussing the module responsible for recommending optimal cricket teams for fantasy leagues.
5. **Architecture Overview:** Offering a concise overview of the product's architecture in five steps.



The comprehensive structure of the methodology section involves separating various operational aspects into individual sub-sections, providing a more detailed and insightful explanation for each segment of the product development process. The Architecture section will provide a general overview, and the Machine Learning sub-section will delve deeper into the specific machine learning methodologies applied.

2.2.1 Data scraping and transformation

1.Data Source: The ball-by-ball data for all Indian Premier League (IPL) tournaments was collected by scraping information from multiple sources. This raw data was obtained in YAML format.

1. **Processing with Python:**
    - Utilizing the YorkPy library in Python, the scraped YAML data was processed to convert it into structured Pandas DataFrames. This allowed for ease of handling and manipulation.

2. **Data Transformation:**
    - The processed data in the Pandas DataFrames was transformed into distinct datasets for various analyses:
        - **matchbymatchdata**: Containing comprehensive match-by-match details, including runs, wickets, overs, and other pertinent match statistics.
        - **venueData**: Specifically focusing on data related to venues, providing details such as venue names, location, and associated attributes.
        - **venue_batter_data**: Aggregated data focusing on venue-specific batting statistics, encompassing runs, strike rates, and batting performances at each venue.
        - **venue_bowler_data**: Compiling bowler-specific data for each venue, capturing bowling statistics and performance details in different locations.
        - **batter_opposition_wise_data**: Detailing batting data concerning opposition teams, offering insights into how players perform against specific opponents.
        - **bowler_opposition_wise_data**: Recording bowling statistics against specific opposition teams, highlighting bowlers' performances against different opponents.
3. **Database Transformation:**
    - Post data transformation, the information was stored into different database systems, ensuring organized and structured storage for future analyses and model development.

The data collection and transformation processes involved using Python and relevant libraries to not only process the raw ball-by-ball data but also to categorize and structure it into various datasets that offer specific insights into IPL matches, venues, player performances, and opposition-specific statistics. The transformation and storage into various databases facilitate efficient and structured data management for comprehensive analyses and modeling.

2.2.2 Feature Engineering

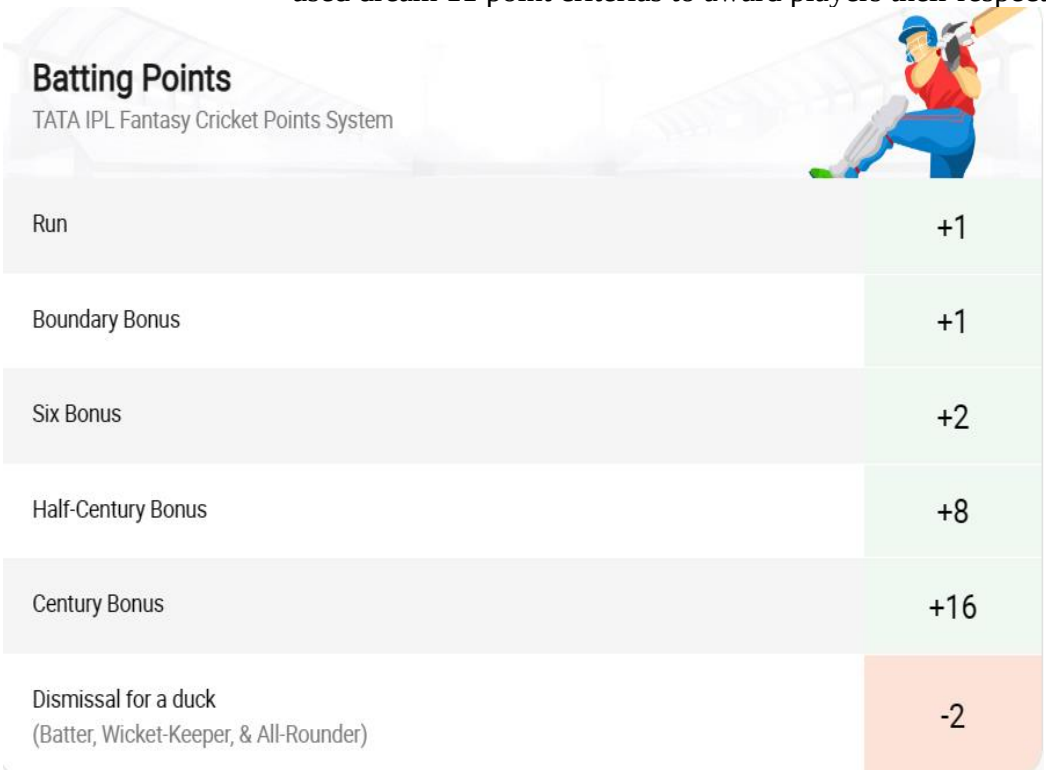
1. **Batting Features:**
    - **Strike Rate:** Calculated by deriving the ratio of runs scored to balls faced, offering insights into the efficiency of a batsman's scoring.
    - **Average:** Determined by computing the mean of runs scored per dismissal, providing a statistical measure of a player's performance.
    - **Opposition-wise Average:** Calculated the average runs scored against specific opposition teams, highlighting performance against different opponents.
    - **Venue-wise Average:** Evaluated the average runs scored at various cricket venues, indicating a player's performance at different grounds.
2. **Bowling Features:**
    - **Economy Rate:** Computed from the average number of runs conceded per over bowled, offering an insight into a bowler's efficiency.

- **Average:** Calculated by finding the mean of runs conceded per wicket taken, showcasing a bowler's effectiveness.
- **Opposition-wise Average:** Determined the average runs conceded against specific opposition teams, providing insights into performance against different opponents.
- **Venue-wise Average:** Evaluated the average runs conceded at various cricket venues, indicating a bowler's performance at different grounds.

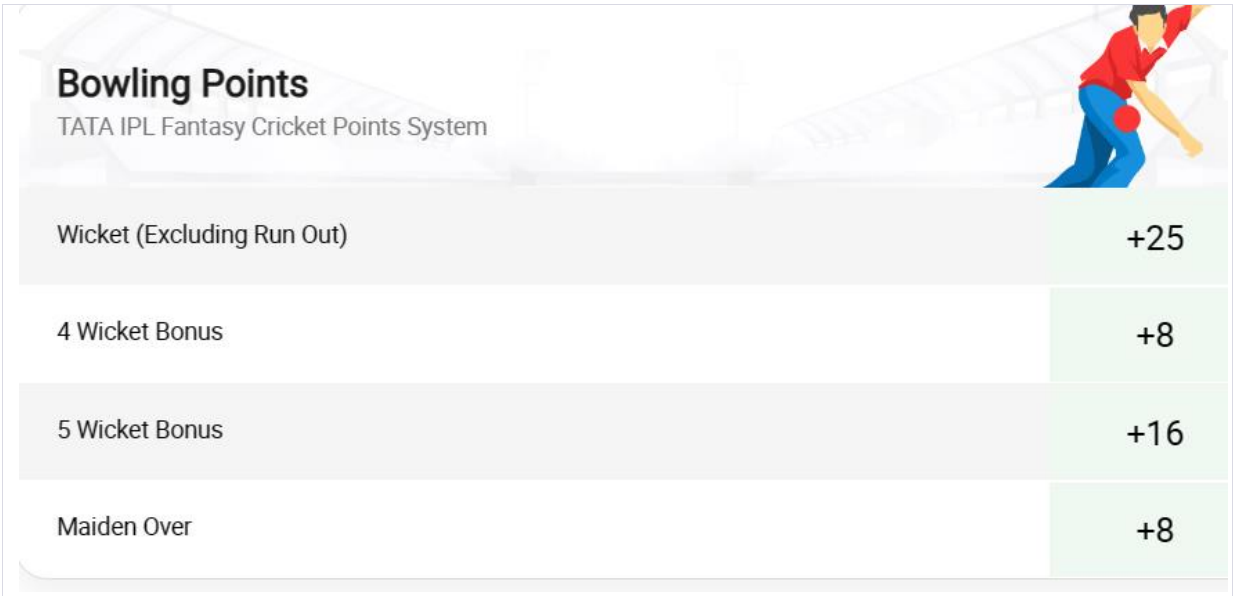
3. **Derived features:**
    - Derived the above-mentioned features by employing mathematical calculations and aggregations based on the structured datasets.
    - Form factor : using last seasons data to make a form factor
    - Position factor for batsmen and bowler by taking average balls played and and bowled

4. **Dream 11 points calculation**

- used dream 11 point criterias to award players their respective points:



Reference: dream11 website

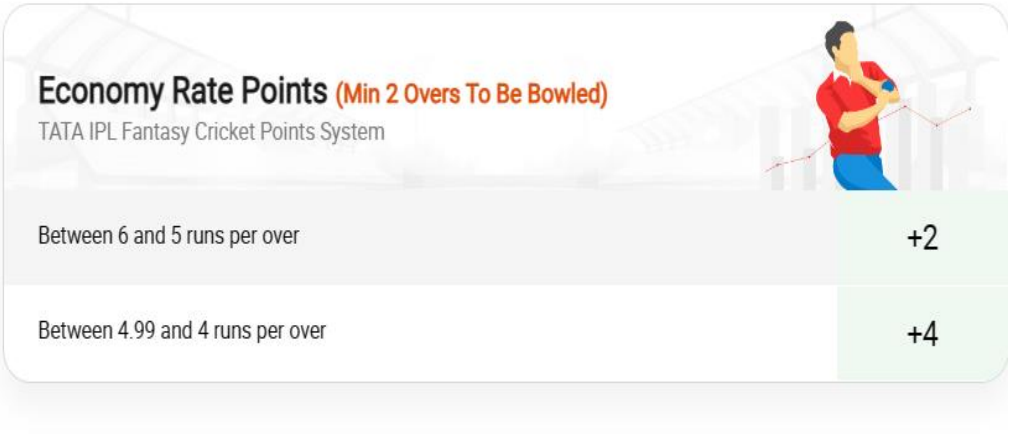


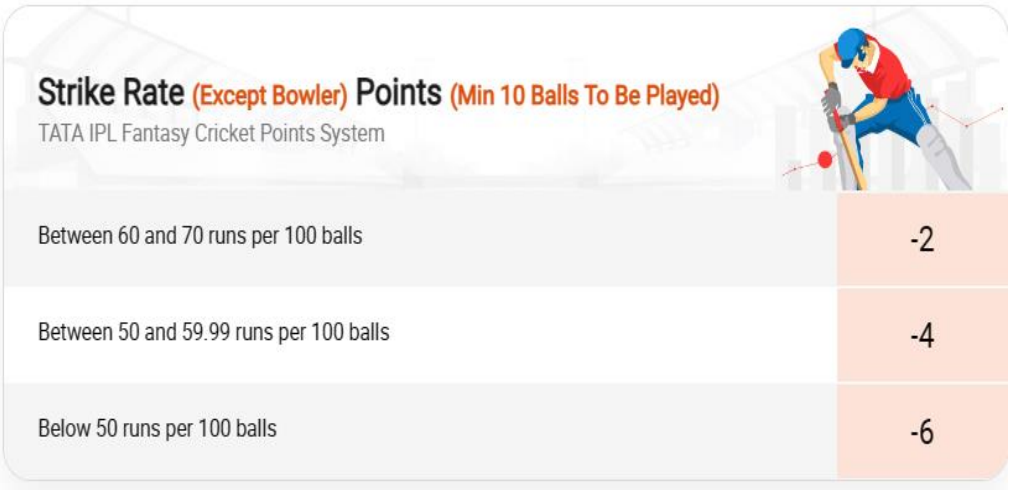
Reference:dream11 website

By implementing feature engineering techniques, the dataset was enriched with additional derived metrics, enabling a comprehensive analysis of player performance based on various factors such as venue, opposition, and overall batting and bowling statistics. The engineered features provided deeper insights into player performances in different conditions and against varied opponents in the IPL matches

Chapter 3
Analysis and Design of models

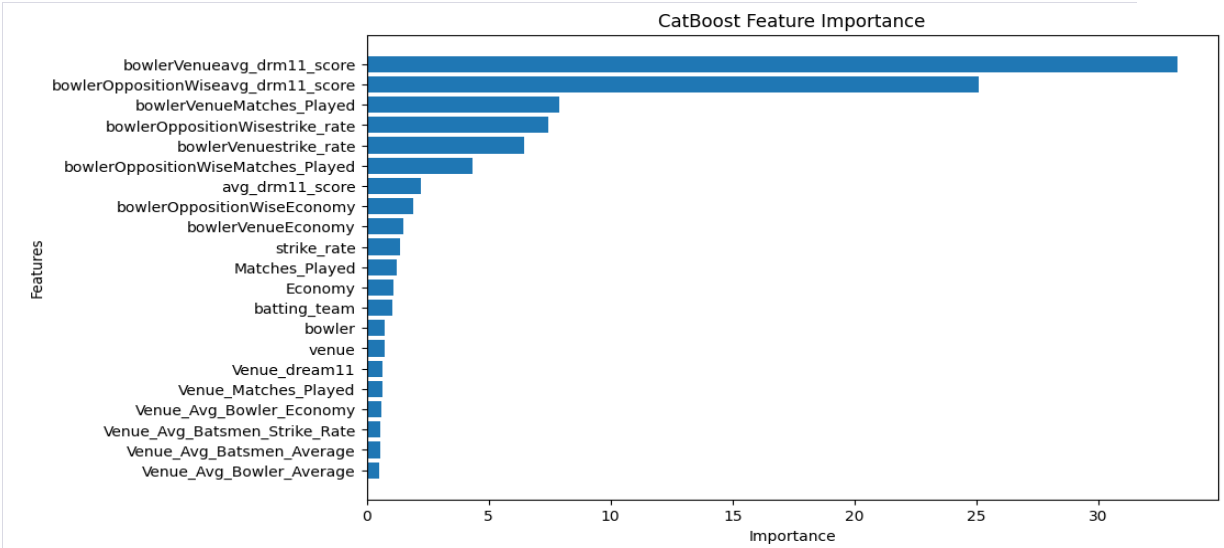
1. **Algorithm Comparison using Pycaret:**
    - **Blackbox Testing:** Executed a comprehensive analysis by comparing different machine learning algorithms using the Pycaret library. This involved training and testing various models with the available datasets to evaluate their predictive performance.
    - **Performance Metrics:** Evaluated the models based on standard evaluation

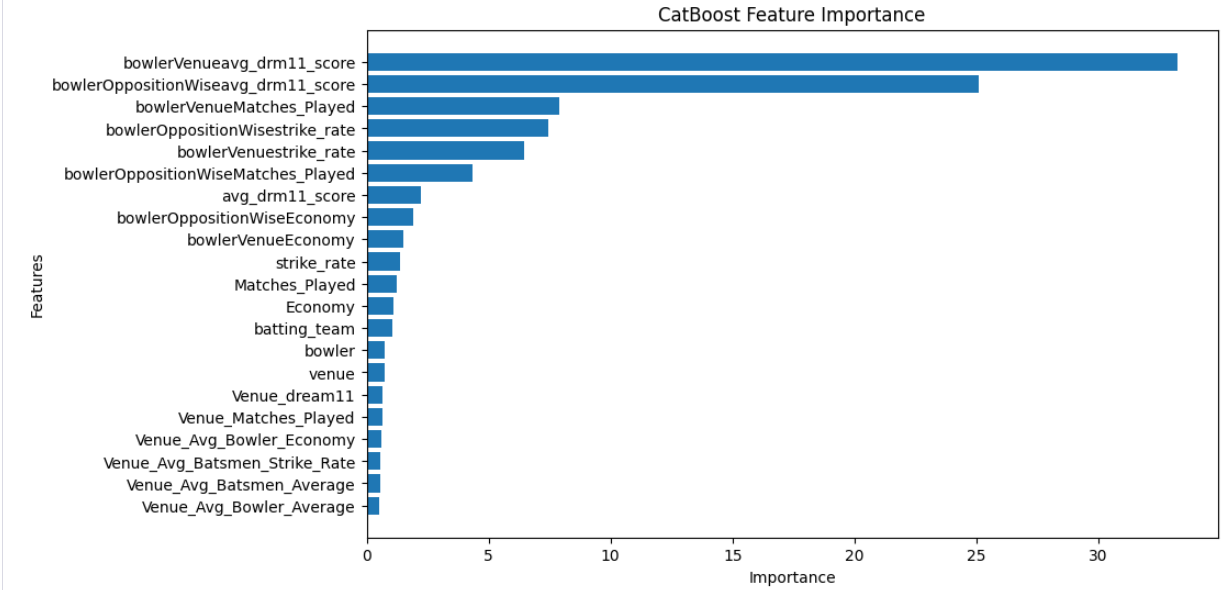


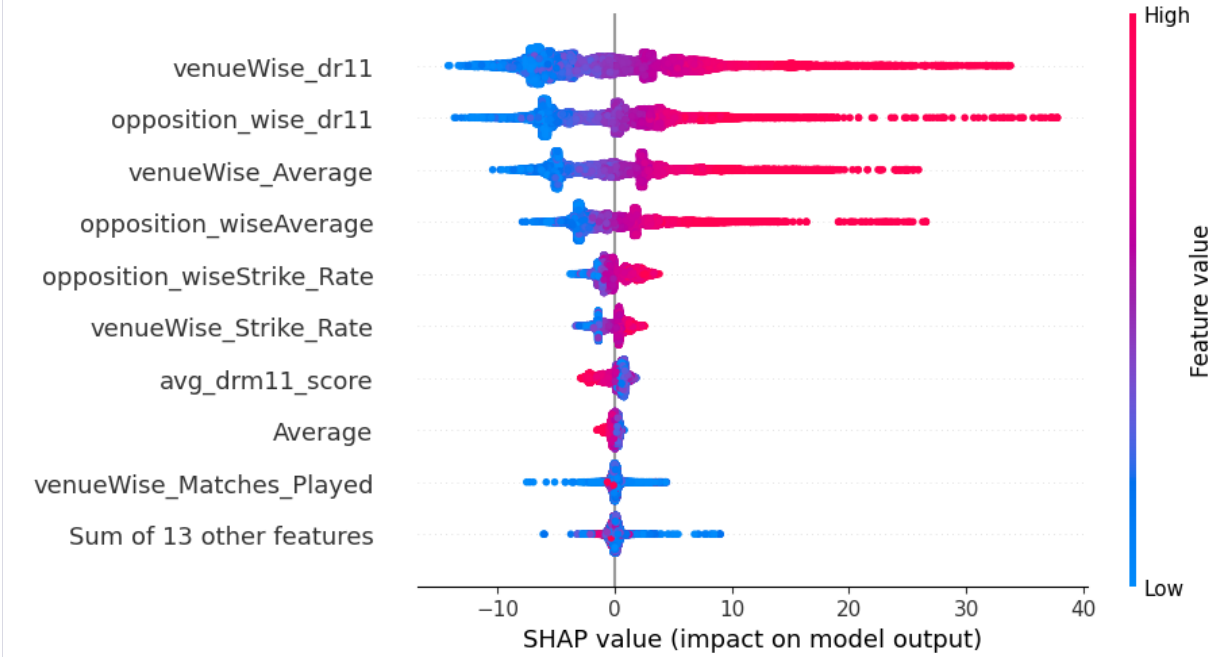


Reference: dream11 website

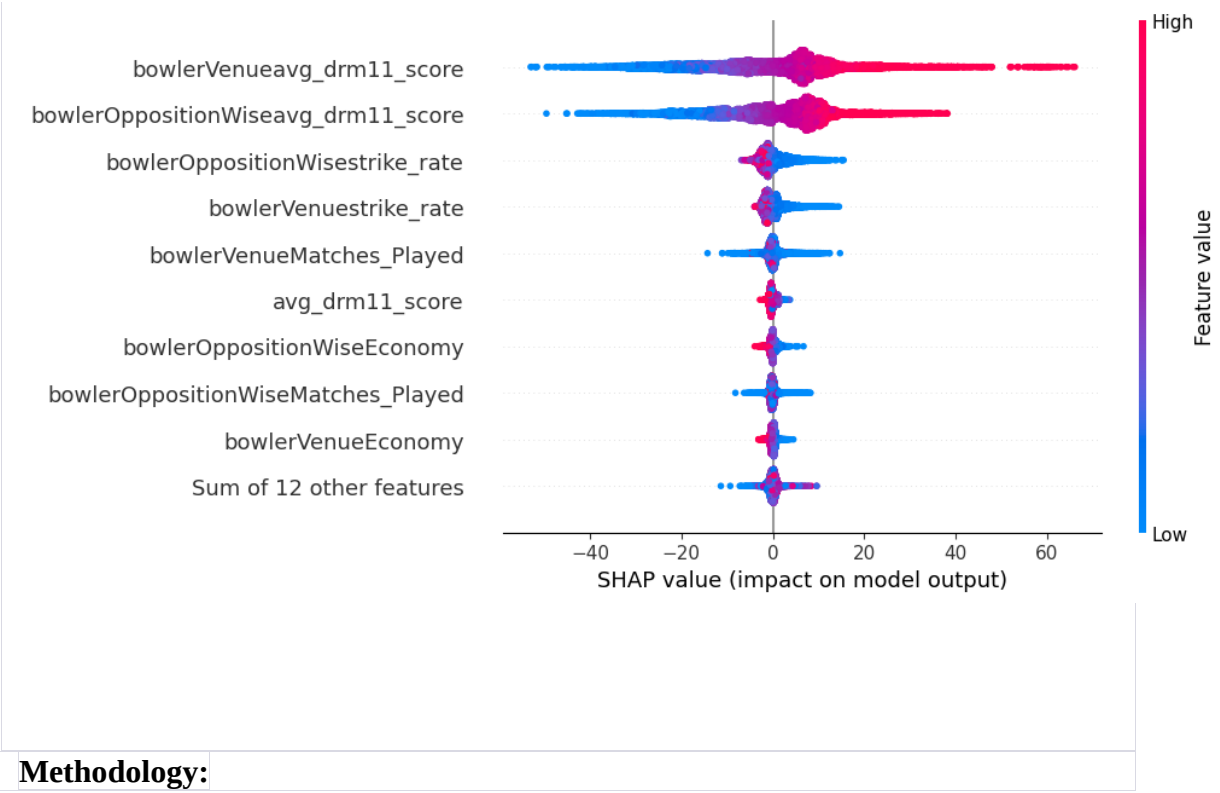
    - n metrics like accuracy, precision, recall, F1-score, and area under the curve (AUC), ensuring a comprehensive assessment of model performance.
2. **Feature Importance and Shapley Values Analysis:**
    - **Feature Importance:** Investigated the importance of different features in predicting player performance by analyzing which factors significantly contribute to the model's predictions. This aided in understanding the relevance and impact of each feature.
    - **SHAP (SHapley Additive exPlanations) Values:** Employed SHAP values, a game-theoretic approach to interpret the impact of variables on model predictions. This analysis provided insights into how each feature influenced individual predictions.







Result obtained



3. **Methodology:**

- Conducted Blackbox Testing: Utilized Pycaret to automate the model selection and hyperparameter tuning process. Trained multiple machine learning models to compare their performance, leveraging a variety of algorithms like Random Forest, XGBoost, CatBoost, and others.
- Analyzed Feature Importance: Explored and visualized the significance of features in model predictions to comprehend the influence of each feature on the overall model performance.
- SHAP Values Analysis: Calculated and interpreted SHAP values to understand the impact of different features on individual predictions, aiding in explaining model decisions.

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
lasso	Lasso Regression	15.0439	486.6412	22.0341	0.4318	1.0602	2.5576	0.033
llar	Lasso Least Angle Regression	15.0439	486.6412	22.0341	0.4318	1.0602	2.5576	0.024
en	Elastic Net	15.0511	486.9709	22.0415	0.4315	1.0603	2.5574	0.060
br	Bayesian Ridge	15.0678	487.3567	22.0502	0.4310	1.0621	2.5585	0.028
ridge	Ridge Regression	15.0727	487.4528	22.0527	0.4309	1.0617	2.5599	0.024
lr	Linear Regression	15.0763	487.5882	22.0556	0.4307	1.0622	2.5604	0.690
gbr	Gradient Boosting Regressor	14.6991	493.0536	22.1802	0.4243	1.0233	2.3773	1.239
lightgbm	Light Gradient Boosting Machine	15.0499	533.6157	23.0806	0.3765	1.0291	2.3660	0.159
catboost	CatBoost Regressor	15.2320	543.7169	23.2986	0.3647	1.0458	2.4009	3.903
ada	AdaBoost Regressor	17.2147	581.9173	24.1048	0.3197	1.1776	3.4007	0.249
rf	Random Forest Regressor	16.0532	618.8852	24.8545	0.2767	1.0949	2.5810	4.927
knn	K Neighbors Regressor	18.3917	686.9593	26.1904	0.1966	1.2091	3.0055	0.062
huber	Huber Regressor	17.5716	709.6827	26.5845	0.1716	1.1046	2.3103	0.117
et	Extra Trees Regressor	17.5342	778.7224	27.8884	0.0889	1.1919	2.7274	1.694
omp	Orthogonal Matching Pursuit	21.2452	812.6241	28.4933	0.0496	1.3161	3.8944	0.024
dummy	Dummy Regressor	22.0547	855.8286	29.2397	-0.0007	1.3675	4.1287	0.021
dt	Decision Tree Regressor	21.4120	1123.2791	33.4963	-0.3175	1.3705	3.0414	0.090
par	Passive Aggressive Regressor	21.7422	1189.0191	34.3988	-0.3926	1.7163	1.4118	0.030
lar	Least Angle Regression	72.4896	20392.2428	99.4604	-23.2105	1.8592	15.0159	0.025

4. **Outcome:**
    - After analysing shap values and feature importance we comcluded than a combined effect of strike rate and average in venue data and were greatly impacting the model
    - Pycaret results showed various accuracy scores of different model and although regularization models like lasso regressing had better r2 value, the gradient boosting algorithms were still unoptimized so we first trained there models

Training Section:

The training phase involved the development and evaluation of three distinct models for both batsmen and bowler datasets to predict player performance in the context of various cricketing metrics.

1.Batsmen Models:

- **Model 1:** Employed a Extra trees Regressor algorithm to predict batsmen's performance, focusing on metrics such as runs scored, strike rate, opposition-wise and venue-wise averages. This model was optimized for predicting batting statistics with high accuracy and precision.
- **Model 2:** Utilized a Gradient Boosting Regressor to capture more nuanced patterns in the batsmen data. This model was tailored to comprehend complex relationships between the features and predict the performance of players in diverse match scenarios.
- **Model 3:** Deployed a CatBoost Regressor to specifically emphasize the prediction of player performance. This model excelled in handling categorical variables and was fine-tuned to predict runs scored, strike rates, and other batting metrics with enhanced accuracy.

2 Bowler Models:

a. **Model 1:** Utilized a Extra trees Regressor to predict bowlers' performances, focusing on metrics such as economy rate, average, opposition-wise, and venue-wise averages. This model was optimized to accurately predict bowling statistics and player performance.
b. **Model 2:** Utilized a Gradient Boosting Regressor specifically designed to comprehend the nuances in bowlers' data. The model excelled in capturing intricate relationships between features to predict players' performance in varied match scenarios.
c. **Model 3:** Employed a CatBoost Regressor, emphasizing the prediction of bowlers' performance. This model was optimized for handling categorical variables and predicting economy rates, averages, and other bowling metrics with high accuracy.

2. **Methodology:**
   a. **Data Splitting:** The available data was split into training and testing sets to train and validate the models' performances effectively.
   b. **Hyperparameter Tuning:** Hyperparameters of each model were fine-tuned to attain optimal performance using techniques such as grid search, random search, or Bayesian optimization.
   c. **Cross-Validation:** Applied K-fold cross-validation techniques to ensure the robustness and generalizability of the models.
   d. **Result:** calculated MAE and MSE

Testing section:

Predicted points were calculated and then compared to testing data:

Results were:

For batsmen:

model	MSE	MAE
Extra trees regressor	490.352	15,21
Catboost	485.151	14.16

XG boost	486.235	14.89

For bowler:

model	MSE	MAE
Extra trees regressor	450.322	15.31
Catboost	475.151	12.56
XG boost	489.526	13.12

These Gradient boosting algorithms performed much better then the other regularization models after Hyperparameter tuning.

Team selection:

Greedy Algorithm for Team Composition

During the development of our project, a key aspect involved the assembly of competitive cricket teams by employing a greedy algorithm. This approach aimed to construct robust teams by adhering to specific player role constraints, ensuring a balanced combination of batsmen, bowlers, all-rounders, and wicketkeepers.

The process began by categorizing available players into distinct groups based on their roles in cricket—batsmen, bowlers, all-rounders, and wicketkeepers. Each category was further refined to include the most promising players with proven performance records in the sport.

The algorithm implemented certain constraints to fulfil the team requirements, such as a minimum of 4 batsmen, 3 bowlers, 2 all-rounders, and 1 wicketkeeper. The fundamental principle behind the algorithm's operation lay in its iterative selection process. By employing a greedy strategy, the algorithm systematically chose players, beginning with an empty team roster. It iteratively examined available players, aiming to maximize the team's overall potential by selecting those with the best performance records and capabilities in their respective roles. Throughout the selection process, the algorithm consistently verified whether the chosen players aligned with the predefined constraints. If any constraints were violated, adjustments were made to the selection to ensure compliance with the role-specific requirements.

The primary goal of the greedy algorithm was to form competitive teams while maintaining the necessary balance across roles—promoting a robust batting lineup, a potent bowling attack, versatile all-round capabilities, and proficient wicketkeeping skills.

This approach enabled the creation of well-rounded cricket teams optimized for performance within the specified constraints, allowing for a strategic blend of players' roles to yield formidable and balanced team compositions.

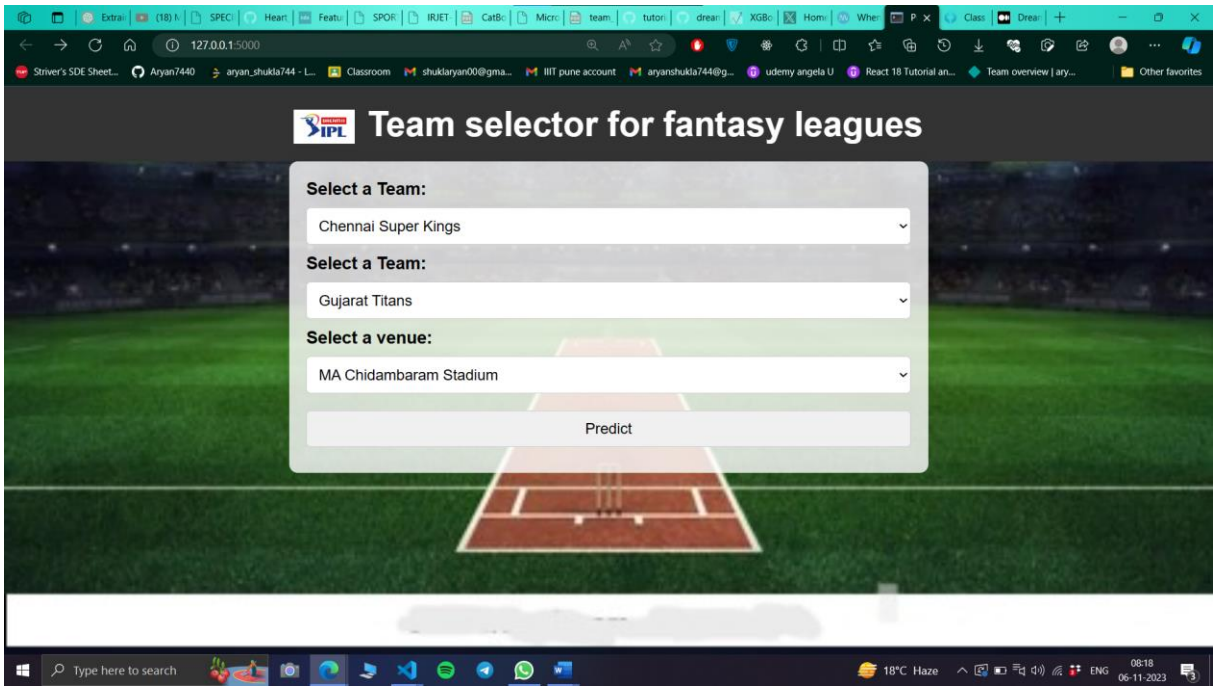
Player Type	Min
Wicket Keeper - WK	1
Batsman - BAT	3
All Rounder - AR	1
Bowler - BWL	3

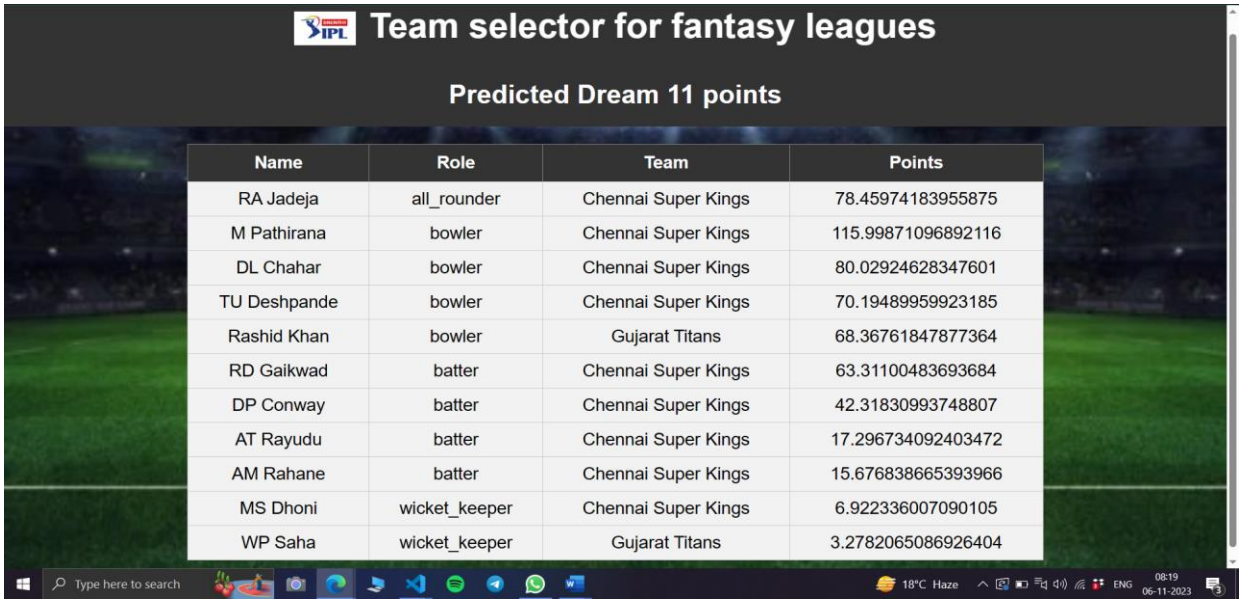
Interface Design: Implementing a Flask Application

The project integrated a Flask web application to serve as the user interface, providing an intuitive platform for users to interact with the predictive models and access analytical insights. This section details the design and functionalities of the Flask-based interface.

Features and Functionalities:

1. **Prediction Request Form:** Users can input essential match details, such as team names, venue, and opposition, into a structured form within the Flask app. This information serves as input data for the predictive models.
2. **Dropdown Menus and Selectors:** To facilitate ease of use, dropdown menus and selectors were implemented for selecting teams, venues, and opposition. This simplified the input process for the users.
3. **Display of Predicted Player Points:** Upon inputting the match details, the Flask app invokes the predictive models, generating a list of predicted player points for the specified teams. The output displays player names along with their anticipated points.
4. **Dynamic Rendering of Results:** The app dynamically renders the results in real-time, ensuring swift and accurate presentation of the predicted player points, providing users with immediate insights.





Interface of our website

Chapter 4
Results and Discussion

Results:

1. Predictive Model Performance:

Present the results of the predictive model, showcasing its MSE and MAE. Compare the performance metrics of Extra Trees, Cat-Boost, and XG-Boost algorithms, highlighting their individual strengths in predicting player performances and selecting the best performing team.

2. Comparison with Traditional Models

Comparing the results of the advanced algorithms with traditional classification models that were commonly used in fantasy cricket platforms. Like liner regression, SVM and regularized regression algorithms Highlighted very significant improvements in accuracy and reliability, emphasizing the superiority of the proposed algorithms.

3. Real Life application:

Although accuracy was improved these were still not enough to win fantasy league games but for these application error was not a good metric to we utilized a rewards based system where we calculated our predicted teams score and then compared it with the top 11 performing players of that match we trained our model up to 2018 and then tested for 2019 and also trained it upto 2022 and then tested on 2023

YEAR	ACCURACY
2019	78%
2023	82%

These results were enough for placing us in the top 70 percentile which could eventually give us monetary rewards in the long term

Discussion:

1. Algorithm Effectiveness:

Higher effectiveness of Extra Trees, Cat-Boost, and XG-Boost algorithms in comparison to traditional methods. The algorithms' ability to capture complex patterns has led to more accurate predictions, providing users with valuable insights for team selection.

2. User Empowerment and Informed Decision-Making:

Accurate predictions and user-friendly interfaces have empowered users, enabling them to make well-informed decisions during team selections. The impact on user confidence and how this has translated into a more enjoyable gaming experience.

3. Challenges and Future Improvements:

Challenges Faced

1. **Dataset Size Constraints:** The project encountered limitations due to the dataset's size. Working with larger and more comprehensive datasets could potentially enhance the accuracy and depth of the predictions and analyses.
2. **Model Complexity - Leveraging Neural Networks:** The utilization of Artificial Neural Networks (ANN) presents a prospective improvement. Harnessing the power of deep learning models could offer more sophisticated insights and better pattern recognition.
3. **Player-by-Player Matchup Factor:** An enhancement that could significantly bolster the model's predictive accuracy is integrating a player-by-player matchup factor. Considering player-specific matchups, such as a batsman's historical performance against particular bowlers, can elevate the precision of the predictions.
4. **Weather Condition Incorporation:** Integrating weather conditions into the predictive models represents an untapped area for improvement. Factoring in weather attributes like humidity, pitch conditions, or climate could add depth to the analysis, especially in cricket, a sport significantly influenced by weather.

Future Improvements

1. **Augmenting Dataset Size:** One crucial direction for future work is expanding the dataset size by collecting a more extensive range of cricket match data. This expansion could involve a larger span of tournaments and a wider diversity of match conditions.
2. **Integration of Advanced Machine Learning Models:** The future of this project involves exploring and implementing state-of-the-art machine learning models, such as deep neural networks, reinforcement learning, or hybrid models that amalgamate diverse algorithms for improved accuracy.
3. **Enhanced Player-specific Analysis:** Enhancing player-specific analysis by delving into individual players' historical data and using advanced statistical metrics could further fine-tune the prediction models.
4. **Weather-Integrated Predictions:** Improving predictive accuracy by integrating weather conditions and their impact on gameplay is an area ripe for development. Considering how climate and weather affect match outcomes could significantly enhance the project's predictive capabilities.

These future improvements and strategies to overcome challenges hold the potential to advance the project's predictive capabilities and further refine its accuracy, making it a more robust and reliable tool for predicting outcomes in cricket matches.

1.

4. Impact on Fantasy Cricket Gaming Industry:

The innovative approach and positive user experiences have the potential to influence future platforms and user engagement strategies in similar online gaming domains

Chapter 5

Conclusion and Future Scope

The application of predictive analytics models for team selection in cricket can be extended to various other games and fields, providing valuable insights and aiding decision-making processes. Here are some potential future scopes for the model and its updated versions:

1. **Sports Fantasy Leagues other than cricket:**
    - Extend the model to other sports such as football, basketball, baseball, or soccer, enabling users to make informed team selections in different fantasy sports leagues. Each sport has unique player statistics and dynamics, offering diverse opportunities for predictive modelling.
2. **Esports Team Selection:**
    - Apply the model to esports games like League of Legends, Dota 2, or Counter-Strike: Global Offensive. Esports have become immensely popular, and team composition is crucial. Predictive analytics can assist gamers in forming winning teams based on historical player performance data.
3. **Stock Market Investment:**
    - Utilize similar predictive algorithms to analyse stock market trends and assist investors in making data-driven investment decisions. Historical stock data, market volatility, and company performance metrics can be used to predict future stock prices and guide investment strategies.
4. **Online Gaming Strategies:**
    - Implement the model in online strategy games where players form teams or alliances. By analysing past game data, player strengths, and strategies, the model can recommend optimal team compositions and gameplay strategies, enhancing the overall gaming experience.
5. **Employee Recruitment and Team Building:**
    - Apply the model's principles to HR analytics for employee recruitment and team building in organizations. Analyse historical performance data, skills, and personality traits to form effective and complementary teams within the workplace, improving collaboration and productivity.
6. **Collaborative Projects and Research Teams:**
    - Extend the model to assist in forming collaborative research teams or project groups. By analysing researchers' expertise, past collaborations, and success rates, the model can suggest optimal team compositions, leading to more impactful research outcomes and project success.
7. **Online Content Creation:**
    - Implement the model for online content creation platforms, recommending optimal teams of writers, designers, and developers based on their expertise and

past performance. This can lead to more efficient content creation processes and high-quality output.

8. **Healthcare Team Formation:**
    - Apply predictive analytics to form healthcare teams, considering the expertise of doctors, nurses, and specialists. By analysing patient outcomes and team collaborations, the model can optimize healthcare teams, ensuring comprehensive and effective patient care.

By exploring these diverse applications, the predictive analytics model's core principle of data-driven decision-making in team selection can be adapted to various fields, helping individuals and organizations make optimal choices and achieve better outcomes.

References

1.Madav goswami. (2011). Dream11 Team Predictor with Python and Machine Learning By anaylitics Vidhya

2.Tinniam V Ganesh, 2020. Big Data-5: kNiFi-ing through cricket data with yorkpy. [Blog] Medium, Available at: https://gigadom.in/2020/03/06/big-data-5-knifi-ingthrough-cricket-data-with-yorkpy/ 3.Tinniam V Ganesh, 2020. Big Data 6: The T20 Dance of Apache NiFi and yorkpy. [Blog] Medium,.

4. Cricsheet: https://cricsheet.org/ [38] ESPN-Cricinfo: http://stats.espncricinfo.com/ [39]
5.PyCaret.org. 5.PyCaret, April 2020. URL https://pycaret.org/about. PyCaret version 1.0.0.

6. Muthuswamy, S., & Lam, S. S. (2008). Bowler performance prediction for one-day international cricket using neural networks. In IIE Annual Conference. Proceedings (p. 1391). Institute of Industrial and Systems Engineers (IISE).
7. G. Bhatambarekar and S. Rai, "Batsmen Performance Prediction Using Regression Models," *2022 International Conference on Artificial Intelligence and Data Engineering (AIDE)*, Karkala, India, 2022, pp. 80-85, doi: 10.1109/AIDE57180.2022.10060071.
8. Passi, Kalpdrum & Pandey, Niravkumar. (2018). Predicting Players' Performance in One Day International Cricket Matches Using Machine Learning. 111-126. 10.5121/csit.2018.80310.
9 Passi, K., & Pandey, N. (2017). Predicting players' performance in one day international cricket matches using machine learning. Computer Science & Information Technology (CS & IT).
10.Wickramasinghe, I. P. (2014). Predicting the performance of batsmen in test cricket. Journal of Human Sport and Exercise, 9(4), 744-751.