RSET
RAJAGIRI SCHOOL OF
ENGINEERING & TECHNOLOGY
(AUTONOMOUS)

Project Report On

Cricket Data Analysis and Prediction

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science and Engineering

By

Aarya Purushothaman (U2103003)

Adithyakrishnan A (U2103015)

Aleena Roy (U2103025)

Ali Thalhathe (U2103027)

Under the guidance of

Ms. Jisha Mary Jose

Department of Computer Science
Rajagiri School of Engineering & Technology (Autonomous)
(Parent University: APJ Abdul Kalam Technological University)
Rajagiri Valley, Kakkanad, Kochi, 682039
April 2025

CERTIFICATE

This is to certify that the project report entitled ***Cricket Data analysis and Prediction*** *is a bonafide record of the work done by* ***Aarya Purushothaman(U2103003)***, ***Adithyakrishnan A (U2103015)***, ***Aleena Roy (U2103025)*** *and* ***Ali Thalhathe (U2103027)*** *submitted to Rajagiri School of Engineering & Technology (RSET) (Autonomous) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (B. Tech.) in Computer Science during the academic year 2024-2025.*

Project Guide
Ms. Jisha Mary Jose
Assistant Professor
Dept. of CSE
RSET

Project Co-ordinator
Mr. Harikrishnan M
Assistant Professor
Dept. of CSE
RSET

Head of the Department
Dr. Preetha K G
Professor
Dept. of CSE
RSET

ACKNOWLEDGMENT

We wish to express our sincere gratitude towards **Rev. Dr. Jaison Paul Mulerikkal CMI**, Principal, RSET, and **Dr. Preetha K G**, Professor, Head of the Department of Computer Science for providing us with the opportunity to undertake our project, Cricket Data analysis and Prediction.

We are highly indebted to our project coordinator, **Mr. Harikrishnan M**, Assistant Professor, Dept. of CSE, for his valuable support.

It is indeed our pleasure and a moment of satisfaction for us to express our sincere gratitude to our project guide **Ms. Jisha Mary Jose** , Assistant Professor, Dept. of CSE, for her patience and all the priceless advice and wisdom she has shared with us.

Last but not the least, we would like to express our sincere gratitude towards all other teachers and friends for their continuous support and constructive ideas.

Aarya Purushothaman

Adithyakrishnan A

Aleena Roy

Ali Thalhathe

Abstract

This project aims to develop an application for comprehensive cricket league data analysis, including the Indian Premier League (IPL). The application integrates live match scores, historical performance data of players and teams, and provides real-time insights and analytics. The primary objective is to enhance professional analysis and fan engagement by offering a robust platform for accessing detailed match statistics, visual analytics, and predictive insights across IPL.The proposed solution involves creating an interactive dashboard which will serve as the main interface for users. This dashboard will display visual representation of player statistics which provides insights into batsmen's and bowler's statistics including run distribution , wicket patterns etc. It will feature in-depth player performance analytics such as barcharts, piecharts, scatterplots enabling users to visually analyze player performance trends during matches in IPL.

Contents

Chapter 1

Introduction

1.1 Background

Today cricket is played everywhere in the world and has become a data-loaded sport from its history with time. The game traditionally has been reduced to simple statistics as batting and bowling averages. It is now crawling at high speed into a swirl of analytical worlds where deeper insights can be gleaned into quite complex performances and strategies of players in matches. Every match now produces an unparalleled amount of detailed information, including ball-by-ball logging, player fitness, and contextual conditions such as pitch conditions. Modern machine learning and data-visualization tools would theoretically help teams develop performance optimization and strategy planning, not to mention data-backed decisions. ESPNcricinfo and several other populer data is used in cricket and has increased fan involvement via interactive dashboards and fantasy cricket insights. These achievements have raised the height of the gulf between raw data and actionable insight, and fundamentally transformed the way cricket is played and experienced.

1.2 Problem Definition

Cricket fans, teams and analyst groups face challenges in accessing a common platform that will provide up-to-date and comprehensive statistics and prediction insights of matches, players, and teams. The restriction hinders their ability to take data-driven decisions and improve match strategies, resulting in a lack of competitiveness and understanding of the game.

1.3 Scope and Motivation

Development of a web application which is intended to provide advanced analytics and machine learning techniques to analyze cricket data to yield meaningful information about the match scenarios and performance of the players. The application can analyze both the historical and real-time data to create predictive models and interesting analyses that can help the users perform pattern identification, understanding and analyzing of trends related to key performance drivers in cricket.

It is designed with the aim of usage by cricket teams, analysts, and fans to take informed and well-grounded decisions, more robust strategies, and increased interaction with the sport. Intuitive dashboards and data visualizations would allow users to have their complex analyses in an entirely accessible format. This platform aims to change the way one understands and experiences cricket, by narrowing down the gap between raw data to actual insights thereby increasing the strategic depth and enjoyment of the game.

1.4 Objectives

- To develop an interactive dashboard visualizing player statistics, offering insights into batsmens' and bowlers' performance, including run distribution and wicket patterns. It will feature bar charts, pie charts, and scatter plots for analyzing IPL performance trends.
- To incorporate historical players' and team information to predict outcomes of matches.
- To optimize fantasy cricket team selection by calculating player performance scores using historical IPL ball-by-ball data, applying predefined fantasy point rules for batting, bowling, and fielding, and incorporating recent performance adjustments to create a balanced and high-scoring team.
- To create a Chatbot that interacts with users where they may ask queries, obtain real-time updates, forecasts, and access historical data.
- To track trajectory of the ball and angle detection in cricket using YOLO model.
- To detect LBW in cricket using Computer Vision techniques.

1.5 Challenges

- Availability and Quality of Information: Cricket data is seldom structured or clean. It is either inconsistent, incomplete or unavailable. Especially real-time data can lead to inaccuracies in analysis and prediction.

- Accuracy and Complexity Models for Cricket: Cricket is highly complex as different pitches, players' form, weather and strategy impacts the accuracy of the models. Sudden player injury or sudden weather change is unpredictable and difficult to factor accurately in models.

1.6 Assumptions

- Time Frame: The analysis is restricted to a specific season(s) or date range.

- Data Privacy and Security: The platform presumes with regard to the entire law compliance for all data privacy regulations and secure handling of user data. It presumes that any personal information from the users collected shall and the data being processed would be secure in the storage and accordance with relevant data protection laws.

1.7 Societal / Industrial Relevance

The industrial importance of the cricket analysis and prediction project are:

- The project empowers teams to analyze the player performance or match conditions along with strategic decision-making through better-informed team selection and tactics.

- Interactive dashboards immerse fans with real time statistics and insights into the game.

- Fantasy Cricket Team Optimization aiding fans in making the ideal fantasy teams thereby boosting participation in fantasy sports which is a growing worldwide market.

- This project has the potential for state-of-the-art AI/ML application in processing big data. Logistic Regression, Random Forest, and a few optimization algorithms significantly improve the prediction accuracy.
- The enhanced team and player predictions will bring more traffic and revenue in fantasy sports and sports gaming.
- Broadcasters will use these insights and visualizations to engage their audiences.
- The methodologies for cricket like machine learning, optimization, and visualization can be scaled for use in other sports and lead to larger market applications.
- This project shall close the gap for all cricket enthusiasts, analysts, and teams;a unified environment offering well-formed analytics, predictive insight, and near real-time data integration.

Thus, in a nutshell, this project will add value to the sport itself;the game of cricket, in technical and commercial individual capacity while also paving way for similar enhancements in other sports.

1.8 Organization of the Report

The report is structured into five chapters:

- Chapter 1, 'Introduction' covers the background, definition of the problem, scope and motivation, objectives, challenges, assumptions, and social or industrial relevance.
- Chapter 2, 'Literature Survey' reviews the different domains of the project.
- Chapter 3, 'System Design' covers the system architecture, Tools and Technologies used, Team contributions, and the Project timeline.
- Chapter 4,'Results and Discussions' includes the results of the project.
- Chapter 5,'Conclusion & Future Scope' summarizes the features of the project and also the future capabilities of the project.

Chapter 2

Literature Survey

2.1 Introduction

Cricket analytics has significantly advanced with the incorporation of machine learning, optimization techniques, and data visualization tools[1]. Existing studies emphasize the role of predictive analytics in forecasting match outcomes, selecting fantasy teams, and evaluating player performance in real time. These insights have paved the way for comprehensive data-driven systems that cater to analysts, teams, and fans.[2]

Fantasy sports analytics has mushroomed as a crucial niche in sports data science, blending historical performance metrics[3], optimization techniques, and machine learning strategies to refine various team selection strategies. H. Zhao et al.[4] proposed a multi-objective optimization framework for the selection of football team members, emphasizing how mathematical models can be applied to optimize the composition of teams.[5] Learning from this, the project has implemented a rule-based team optimizing system in which statistics–such as runs, wickets, boundaries, and fielding points–are used to quickly generate the most favorable playing XI[6]. This methodical approach helps the fantasy player to make data-driven decisions based on constant scoring rules and performance trends.[7]

As another capability to enhance match forecasting, we implemented a win predictor with machine-learning models[3] including Logistic Regression, Random Forest, Decision Trees, and K-Nearest Neighbors (KNN). [8]Random Forest was ultimately used to predict outcomes of matches after training and testing on parameters such as toss results, venue, and recent player form[9]. This process of selecting the model is analogous to the one demonstrated by Karatas et al. in highlighting comparative model evaluation to improve performance across imbalanced datasets. Predictive analytics further assist users in constructing the most likely outcomes of matches to optimize their team selection accordingly.

In real-time interaction and access, we created an interactive chatbot using LangChain. In our architecture, cricket data are converted from a CSV file into vector embeddings using the pre-trained Hugging Face model and stored in a FAISS vector store[10]. User queries are then processed through a RetrievalQA chain that combines the LLaMA language model to generate relevant and accurate answers backed by the dataset. This structure is reminiscent of the one used in JAICOB[11], showcasing the potential of data-driven chatbots to provide personalized insights. Complementing it is the interactive dashboard built in Python, with Pandas and Plotly, for visualizing match data and player performance through bar charts, pie charts, and scatter plots, giving a clear presentation of trends from which users can draw actionable insights. [12]

From a computer-vision standpoint, our project relies on YOLOv8 for cricket ball tracking, allowing for object detection and trajectory estimation[13]. A custom dataset created using extracted bowling action video frames with labelled ball images allows for the accurate analysis of bounce angles and motion paths, respectively[14]. Also, we have implemented an automated test for LBW detection based on HSV color filtering for tracking the ball, RGB and edge detection for recognizing the batsman, and on contours for identifying the pitch area[1]. The logic of LBW will analyze "pad-before-wicket" situations, assisting in decision-making without expensive multi-camera systems. This setup draws from the work of Nayak et al.citenayak2023computer, which discusses a cost-effective computer-vision-based LBW detection framework.[15]

2.2 Summary and Gaps Identified

In the recent years, advances in sports data science have brought phenomenal improvement in the accuracy and efficiency of fantasy team optimization, match prediction, and performance analysis. A variety of machine learning techniques, such as ensemble techniques, decision tree, and optimization algorithms, were successfully used in outcomes prediction concerning assisting users in fantasy sports decision making. Deep learning computer vision, especially some models like YOLOv8, enhanced object tracking in sports videos, with applications such as ball tracking and trajectory predictions. And also, RetrievalQA and large language models as chatbots made cricket data more intelligible and interactive. These advancements point out how AI plays an important role in the intelli-

gentization and user-friendly ability of sports analytics.

There are, of course, gaps in research in these areas. Most of the predictive models had developmental efforts directed toward football alone. There are yet no attempts to adapt to the dynamic and data-enriched environment of cricket. There is limited work regarding combining rule-based fantasy point systems with real-time match data to optimize fantasy team selection. Also, video analysis systems in cricket mostly depend on costly infrastructure and cannot survive light and motion fluctuations well. Very few studies ventured to build a low-cost, real-time LBW detection system based on basic computer vision principles without using predictive modeling techniques. With respect to cricket, few development works still exist on chatbots tailored to this sport, particularly those that make use of vector databases to avail of context-aware response-prompting capabilities. This project occupies the above-mentioned niches by collating rule-based optimization, machine-learning models, YOLOv8 tracking, and LangChain-based chatbot into a single platform-an affordable cricket analytics port.

2.2.1 Gaps Identified

- **Limited Adaptation of Cricket into Optimization Models:** In a very limited way, one can claim that most optimization setups where fantasy sports have been developed emphasize football, leaving cricket out concerning its unique dynamics, scoring patterns, and player roles.

- **Integrated Rule-Based and Real-Time Team Selection Systems Do Not Exist:** For team predictions, existing solutions consider either historical data or machine learning separately instead of some systems integrating real-time updates with rule-based fantasy point calculations to optimize team selection on the fly.

- **Spike in Costs and Complexities with Video-Based Approaches:** Proprietary ball-tracking systems and LBW detection systems, such as Hawk Eye, require their specific multi-camera setups, rendering them costly and inaccessible for mass use or for academic prototyping.

- **Neglected Development of Cricket-Specific Chatbots:** Even when some of the rules for smart chatbots are available, very few implementations develop intel-

ligent, cricket-specific chatbots that can retrieve accurate, context-aware answers from structured cricket datasets using vector embedding and LLMs.

2.2.2 Summary

This chapter spans an entire analysis of the integration of ML and AI within cricket analytics, highlighting major advancements while also pointing out existing gaps. The researchers have proposed various machine-learning-model-based predictions on match outcomes and player performances through the analysis of old data with available statistical features. For example, algorithms such as random forests have already been used to demonstrate the prediction of results of the Indian Premier League (IPL) matches, thereby indicating how data-driven process approaches can be instrumental in strategic decision-making. Furthermore, deep learning techniques such as YOLOv5 and YOLOv8 have been implemented for better object detection and tracking in cricket videos, hence allowing a keen analysis of ball trajectory and player movement.

However, owing to the existing research gaps, much further work remains. Optimization models remain mainly football-centric and have thus far failed to capture the peculiarities of cricket's dynamics, scoring patterns[16], and player roles. The other limitation of the current solutions is that they rely on either historic data or machine learning in order to model team selection without real-time updates on rule-based fantasy point calculations. High costs associated with proprietary ball-tracking systems, such as Hawk-Eye, which are hard to justify for general use or prototyping for academic purposes. Furthermore, there is apparently no intelligent cricket-specific chatbot that can accurately retrieve context-wise answers from structured cricket datasets using vector embeddings and large language models. Filling these gaps will substantially advance cricket analytics, resulting in better predictions, cheaper solutions, and interactive tools specific to cricket needs.

Title	Methodologies Used
Multi-Objective Optimization for Football Team Member Selection	Genetic Algorithm (GA) for multi-objective optimization in team composition.
Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset	Machine Learning algorithms including K-Nearest Neighbor (KNN), Random Forest (RF), Gradient Boosting (GB), Adaboost, Decision Tree (DT), and Linear Discriminant Analysis (LDA); Synthetic Minority Oversampling Technique (SMOTE) for addressing data imbalance.
JAICOB: A Data Science Chatbot	Modular cognitive agent architecture; Machine Learning models; Natural Language Understanding (NLU) algorithms.
Advancing in Cricket Analytics: Novel Approaches for Pitch and Ball Detection Employing OpenCV and YOLOv8	OpenCV for image processing; YOLOv8 for object detection and tracking.
Computer Vision and Image Segmentation: LBW Automation Technique	HSV color filtering and contour analysis for ball detection; RGB color-based detection and edge detection for batsman identification; Contour-based techniques for pitch detection; Logic implementation for LBW scenario prediction.

Figure 2.1: Comparison of the papers

Chapter 3

System Design

There is a huge amount of data about sports today: player statistics and match results, historical records, and performance metrics. Just as science or data analytics has developed to mine this data for insights, possibilities for predicted outcomes, or strategic optimization. The Cricket Data Analysis System proposed here will process cricket data, analyze and visualize it, and inform actors like analysts, coaches, players, and fantasy team enthusiasts. It will utilize the modular infrastructure and the many options it opens for analysis of cricket data, including ingestion and preprocessing, analysis, and visualization. Among the core functionalities with which this system will be outfitted include trend identification, match outcome prediction made by machine learning models, and visual reporting for easy understanding and interpretation. It will feature an architecture that is proximate to scalability, thus allowing the management of historical and real-time data for use in multi-use cases. Among other considerations for the user interface aspects, the project describes the design of the system in data processing pipelines and the integration of machine learning. Its ultimate goal should be development into a complete user-friendly platform that would convert raw cricket information into tangible insights.

3.1 System Architecture

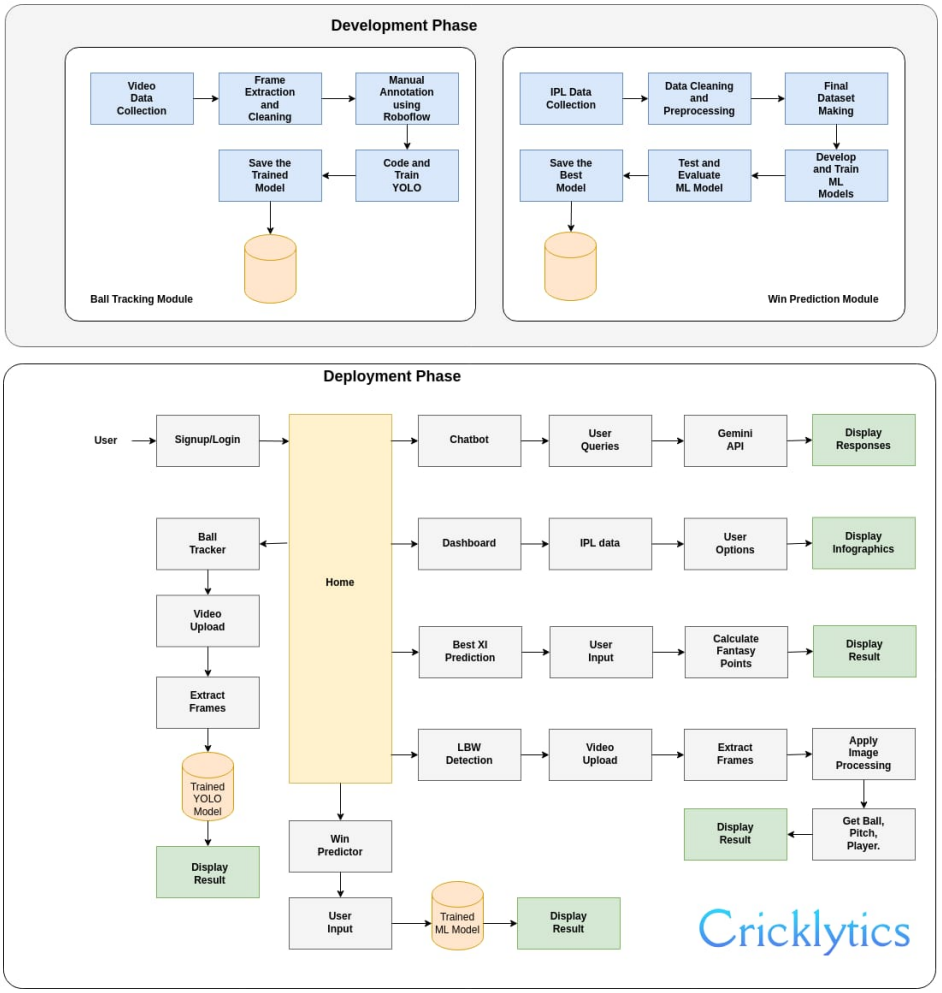


Figure 3.1: Architecture diagram

3.2 Component Design

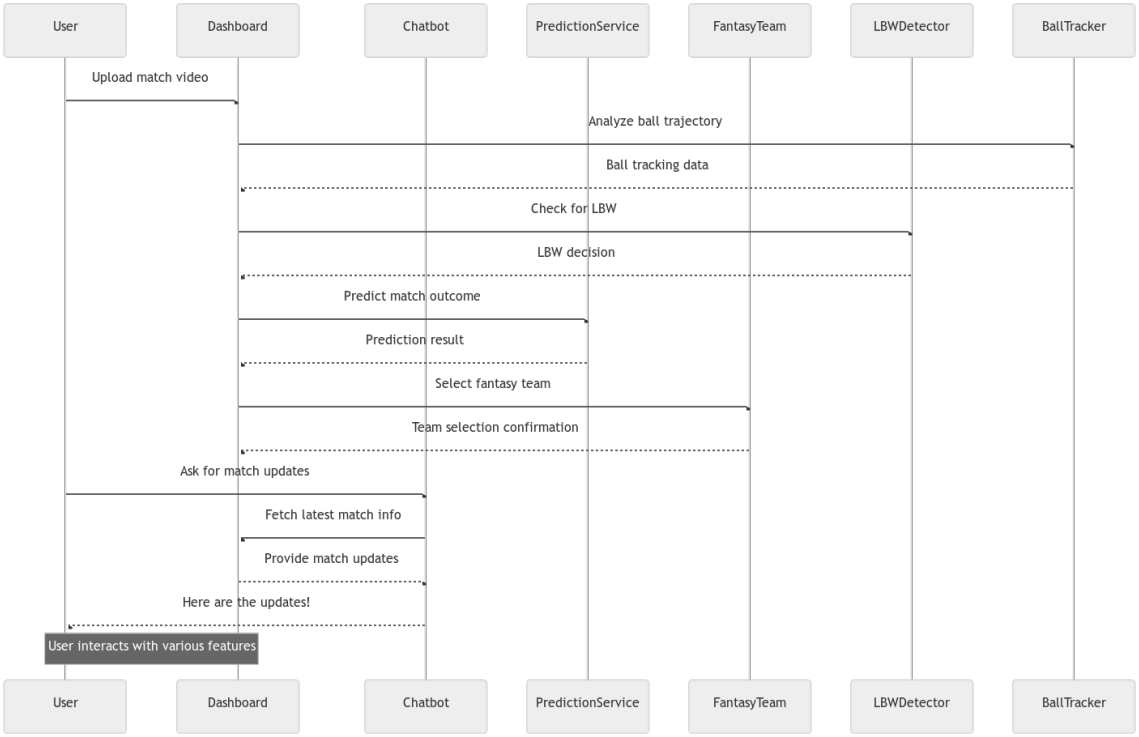


Figure 3.2: Sequence Diagram

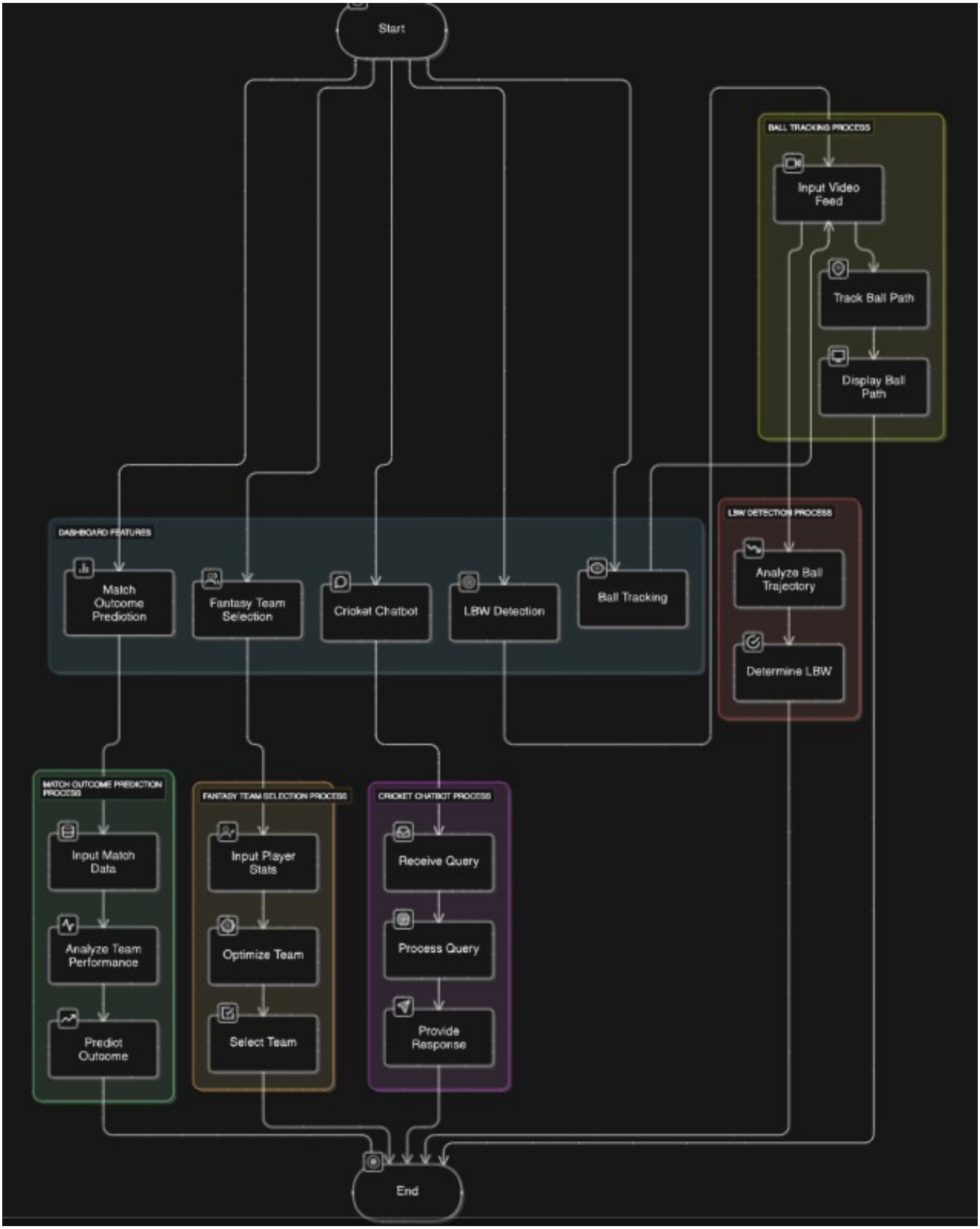


Figure 3.3: Workflow Diagram

It illustrates the interaction of the user with the cricket AI chatbot, team selection module, and match prediction system. Here follows the description of each of them with its purpose:

1. User: The User interacts with the Web application for a variety of tasks, such as asking various statistics related to player and different teams, asking for match predictions, and looking for insights. Therefore, here are the user's very simple actions:

    - Starting a conversation with the bot. .
    - Analysing players and teams based on different metrics.
    - Viewing player descriptions, predictions, or tother insights.

2. Team Selection Module: It gives feedback to the user by means of acknowledging actions or showing outcomes. It:

    - Accepts input from the Cricket AI Chatbot regarding the players that are chosen by the user..
    - Confirms when a team has been successfully created.
    - Display the selected team players to the user through the chatbot.

3. Match Prediction System: With the help of the chosen team and match data, the Match Prediction System produces insights and forecasts. It:

    - Get a request from the chatbot: prediction results for upcoming matches..
    - Processes data to generate predictions.
    - Delivers predictions and insights back to chatbot for the use of a user..

4. Cricket ball tracking: It predict the trajectory of a cricket ball in videos using advanced object detection and tracking techniques powered by YOLOv8.It:

    - The system extracts frames from cricket videos, detects the cricket ball, and predicts its future trajectory, providing detailed insights into its motion.
    - It also calculates the angle of the ball's motion, identifying key events like bounces..

5. Cricket LBW detection: A Python-based system for detecting Leg Before Wicket (LBW) scenarios in cricket videos. It utilizes computer vision techniques with OpenCV and the cvzone library.It:

    - It identifies key elements in a cricket video feed: the ball, the batsman, and the pitch.
    - It analyzes the motion and position of the ball relative to the batsman and the pitch, the system attempts to predict potential LBW situations.

6. Chatbot Interaction: The user initiates conversation with the chatbot. The chatbot harmlessly prompts the user for player selection. The user selects players, to which the Team Selection Module will send the respective processing. The user then receives confirmation that the team was created successfully. The bot asks for predictions on matches according to the team selected. The Match Prediction System then generates predictions for the match and provides it to the chatbot. The predictions and insights are displayed to the user. The user can ask the system for more questions or clarifications with the predictions. The user can also be in contact with the chatbot while doing other things, like selecting a team.

3.3 Algorithm Design

1. Logistic Regression

Logistic regression entails a linear modality whereby some binary classifications can be enabled. In fact, it can tell whether a certain team is likely to win a match or not based on the input features. More appropriately, the categorical outcome must be either that of winning or losing, or yes or no.

In cricket, for example, one could develop a function that predicts winning or losing chances in a given match based on the number of runs scored, the number of wickets taken, batting average, bowling average, and so on.[17] So, a good example is that, suppose, by using logistic regression, we may be able to specify a certain probability showing that the team will win: on aggregate, measuring top batsman scores, whether performing well or poorly at the moment, bad or good weather conditions etc.

Cricket Application:

- Logistic regression can be used to predict whether a team will win or lose based on factors such as current score, wickets in hand, overs remaining, and historical match data.
- It can predict whether a batsman will score more than a certain number of runs (e.g., 50 runs) or whether a bowler will take a wicket in a specific over based on past performance and match conditions.

Advantages:

- Simple, easy to understand.
- Efficient for tasks of binary classification.

Disadvantages:

- Inapplicable on linear feature-to-feature relations only.
- Treats the set of features independent and could poorly perform for stronger relationships among these features.

2. Random Forest

Random Forest is an ensemble learning approach. It works basically with the collective knowledge of a couple of decision trees. In the case of model trees with actual decision trees, many of those are created, and then each decision tree's outcome is averaged to produce a final output, either for a classification or regression problem.

Random forest is an aggregation learning ensemble. It typically works with the collective knowledge of multiple decision trees. In this case, it's a model with decision trees; many of these are created and combined to finally produce an outcome linearly for classification or regression.

Working:

- Random Forest algorithms create a number of decision trees with thousands of possible decisions selected randomly from the data and features. Each tree then makes its individual predictions, and the final output is determined by averaging the votes of the trees for the classification problem and averaging their predictions in the case of regression.

Being a Random Forest algorithm, random forest generates numerous decision trees ranging into thousands by randomly selecting subsets of data and features. Each tree provides its unique prediction, and the final output is mainly from the average prediction value if it is a regression problem or the majority voting count for a classification problem.

- Each decision tree is trained on a random subset of the data.
- At every split in the decision tree, it randomly selects a subset of the features to decide the split.
- The ensemble output is produced by aggregation of the prediction of all individual trees.

Cricket Application:

- Predict if a specific player from the historical datasets of cricket can score more runs than a certain defined value or take more wickets than a particular defined range.
- To ascertain if an outcome in a match is more often favorable or unfavorable for any team, several input features such as player statistics, team form history, venue conditions, etc. can be employed.

Advantages:

- Can handle both classification and regression problems.
- Handles a mix of numerical and categorical data well.
- Reduces overfitting compared to individual decision trees.
- Effective with complex, non-linear relationships between features.

Disadvantages:

- It can be computationally intensive when there are many trees.
- Not as interpretable as single decision trees.

3. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a simple yet powerful supervised learning algorithm that can be used for both classification and regression tasks. Unlike ensemble methods like Gradient Boosting Machines (GBM)[], KNN is a non-parametric and instance-based learning algorithm, meaning it does not make strong assumptions about the data distribution.

How it works:

- KNN makes predictions by finding the 'k' closest data points in the training set to a given input and assigns the output based on these neighbors.
- For classification tasks, it assigns the majority class among the 'k' nearest neighbors.

  For regression tasks, it averages the values of the 'k' nearest neighbors.
- The choice of the distance metric (Euclidean, Manhattan, Minkowski, etc.) plays a crucial role in determining similarity.

Cricket Application:

- KNN can be used to classify players into categories such as batsman, bowler, or all-rounder based on features like strike rate, economy rate, and past performance.
- It can help predict the playing role of a new player by comparing them with similar past players.
- Can be applied for similarity-based player recommendation systems, where players with similar stats can be suggested as replacements in fantasy leagues.

Advantages:

- Simple and Intuitive: Easy to understand and implement, requiring minimal assumptions.
- Brain manages and copes with complex non-linear relationships or interactions of a factor that does not exist independently with others..

- Sometimes, they could just produce better predictions than forest when it comes to making a good prediction..

Disadvantages:

- it's quite easy to get overfitted either when there are too many trees involved, or when the model hasn't been tuned properly..
- Non-parametric: Works well with different types of data without assuming a specific distribution.
- Handles Non-Linear Data: Since it does not assume linear relationships, it can work well for complex problems.

4. Decision Tree

A Decision Tree is a non-parametric supervised learning algorithm used for both classification and regression tasks. It splits the data into subsets based on the value of input features, forming a tree-like structure where each internal node represents a decision based on a feature, each branch represents an outcome of the decision, and each leaf node represents a final prediction.

The core idea is to recursively divide the dataset into subsets that are as pure as possible[1] (i.e., have similar target values), using criteria like Gini Impurity, Information Gain, or Entropy for classification, and Mean Squared Error for regression. How it works:

- The algorithm starts at the root node and splits the data on the feature that results in the highest information gain or lowest impurity.
- It continues splitting recursively until a stopping criterion is met (like maximum depth or minimum samples per leaf).
- Once built, the tree is used for prediction by traversing from the root to a leaf node based on feature values.

Cricket Application:

- Predict match outcomes: Based on features like team scores, number of overs left, wickets remaining, venue stats, and historical match data.

- It can help predict the playing role of a new player by comparing them with similar past players.

Advantages:

- Easy to understand and visualize.
- It Can handle both numerical and categorical data and captures non-linear relationships and interactions between features.
- Requires little data preprocessing (no need to scale or normalize features).

Disadvantages:

- Prone to overfitting: Especially when the tree is deep.
- Unstable: Small changes in the data can result in a completely different tree.
- It doesn't perform well on very complex datasets unless pruned properly.

5.YOLOv8 Model

YOLO (You Only Look Once) is a state-of-the-art deep learning-based object detection model that enables real-time detection of objects, making it highly effective for tracking fast-moving objects like a cricket ball. YOLOv8, the latest version, enhances accuracy and efficiency with improved architecture and model optimizations.

How it works:

- Frame Extraction: The system extracts frames from cricket match videos.
- Object Detection: YOLOv8 detects the cricket ball in each frame with high precision.
- Tracking & Trajectory Prediction: Using object tracking algorithms , the movement of the ball across frames is tracked, allowing the system to analyze its trajectory.
- Motion Analysis: The detected trajectory can provide insights into factors like speed, bounce, swing, and deviation.

Cricket Application:

- Ball Tracking: Helps track the ball's motion in real-time, aiding in match analysis.
- Bounce Prediction: Predicts where the ball will bounce based on previous frames, useful for bowler and batsman analysis.
- Umpire Decision Assistance: Can assist in validating no-balls, LBW decisions, and boundary checks.

Advantages:

- Real-Time Processing: YOLOv8 is optimized for speed, making it suitable for live match analysis.
- High Accuracy: Detects small, fast-moving objects like a cricket ball effectively.
- Versatile: Works with various video qualities and lighting conditions.

Disadvantages:

- Computationally Intensive: Requires high-end GPUs for real-time performance.
- Occlusion Challenges: The ball can be temporarily hidden behind players or the pitch, affecting detection.[18]
- Dataset Dependency: Performance relies on a well-labeled dataset with diverse cricket scenarios.

1. Probabilistic Models for Outcome Predictions in Matches:

Understand the algorithms that may be utilized for probabilistic modeling of different outcomes of matches with respect to winning, losing, and drawing on inputs such as the weather, player availability, forms in recent games, and so on.

Utilize Logistic Regression as a binary classification algorithm based on historical data to predict whether a team is most likely to win or lose in a match..

Using random forest or decision tree , model for complex and non-linear interdependencies among variables team composition, venue, and weather conditions, along with player performance.

2. Optimal Team Selection:

In our approach, player selection is performed using a rule-based system, where players

from two competing teams are compared directly against each other using performance metrics and predefined rules. Each player is evaluated based on factors such as recent form, average fantasy points, batting and bowling stats, and match context (venue, pitch conditions, opponent strength, etc.).

Instead of relying on complex machine learning models like Random Forest or GBM,[19] we map each player from one team to a counterpart in the opposing team (e.g., openers vs. openers, main bowlers vs. main bowlers), and then select the player with better statistical performance and impact under similar conditions.

Conclusion

- Logistic Regression is a favoured statistical tool for straight binary classification problems.
- Random forests find it easy to carve through the multi-dimensional data complexity, making it a valuable asset for classification problems having various features.
- KNN provides a very good predictive accuracy while capturing highly complex interactions inside the data, which makes it an excellent candidate for models where high precision is a necessity but the computational cost is a limiting factor.

3.4 Tools and Technologies

3.4.1 Software Requirements

Google Colab Notebook: It is a collaborative space for data preprocessing, exploratory analysis, and running machine learning models. It permits analysts to write and document Python or R scripts to test hypotheses and generate real-time insights.

Essential Libraries:

- Python Libraries (pandas, Matplotlib,Dash,Ultralytics): The module pandas is for manipulating and cleaning data, which makes match statistics and player performance handling much easier.
- Matplotlib: It's a module into which any form of charting can be very effectively rendered for insight into the form of a player, a team, and a match performance.

- R Libraries (ggplot2, plotly): ggplot2: Enabling the generation of specific and beautiful graphs for cricket data. plotly: Allowing interactive visualizations to get to granular details such as ball-by-ball analysis.

HTML/CSS, Javascript: It is the front-end part of the system providing an interactive user interface to get into the analysis tools and dashboards. HTML forms the skeleton which CSS then dresses up in styles and JavaScript then makes it responsive and dynamic in experience.

3.5 Dataset Identified

The source of the data for the project is IPL cricket dataset obtained from cricinfo website and kaggle . It is very reliable in providing detailed cricket data in the way of huge ball-by-ball details of matches among the available formats, Test, One Day International, and T20. Their datasets contain metadata for each match-such as date, location, teams, toss decision, result; player information-roles, performance, and statistics; and ball-by-ball data attached-finer granularity for outcomes of delivery, runs, extras, and wickets.

3.6 Module division

1. Dashboard design and development:
    - A Balanced View of Graphics: The use of bar graphs, line graphs, pie charts, scatter plots, and heat maps to present the most critical statistics and trends.
    - Advanced Calculation : It enables the preparation of sophisticated measures such as win/loss ratios and player rankings.
2. Match outcome prediction: Develop and train machine learning models such as Logistic regression, Random forest, and others like KNN and decision tree to predict the winning probability of a live match ,that can be updated incrementally by new live data when it is available, maintaining the updated models relevant and accurate as well.
3. Fantasy Cricket Team Building:

- Rule Based Approach: It selects an optimal cricket team based on player performance metrics. The system processes updated match data to evaluate players using predefined rules that assign fantasy points for batting, bowling, and fielding contributions. The selection algorithm considers constraints such as team composition (batsmen, bowlers, all-rounders, wicketkeeper), credit limits, and player balance from both teams.

4. Cricket Ball Tracking:: It predict the trajectory of a cricket ball in videos using advanced object detection and tracking techniques powered by YOLOv8.The system extracts frames from cricket videos, detects the cricket ball, and predicts its future trajectory, providing detailed insights into its motion.It also calculates the angle of the ball's motion, identifying key events like bounces.

5. Cricket LBW detection: A Python-based system for detecting Leg Before Wicket (LBW) scenarios in cricket videos. It utilizes computer vision techniques with OpenCV and the cvzone library.It identifies key elements in a cricket video feed: the ball, the batsman, and the pitch. • It analyzes the motion and position of the ball relative to the batsman and the pitch, the system attempts to predict potential LBW situations.

6. Interactive chatbot development: This module facilitates intelligent, data-driven query handling using a Retrieval-Augmented Generation approach. Cricket-related data is first ingested from a structured CSV file, then transformed into dense vector embeddings using a pre-trained Hugging Face model. These embeddings are indexed and stored efficiently in a FAISS vector store to support high-speed similarity-based retrieval. When a user submits a query, a LLaMA-based language model integrated with a RetrievalQA chain retrieves the most relevant documents from the vector store and generates a cricket-specific response grounded in the contextual information. This ensures that the chatbot provides accurate, domain-specific insights by strictly referencing the provided dataset.

3.7 Key Deliverables- Expected outputs

The work delivers a comprehensive interactive dashboard designed to visualize critical cricket statistics and performance trends. It incorporates diverse visualization techniques including bar graphs, line and pie charts, scatter plots, and heat maps to represent data effectively. Advanced metrics such as win/loss ratios, player rankings, and team performance indicators are dynamically computed to assist users in exploring insightful patterns. The dashboard aims to provide both casual fans and analysts with a balanced, data-driven perspective on ongoing and historical cricket matches.

A robust machine learning module is developed to predict match outcomes using models such as Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), and Decision Trees. These models are trained on historical match data and incorporate features like team composition, venue statistics, and player form. Additionally, the system supports incremental updates with live data, allowing predictive models to adapt and remain accurate over time. This ensures real-time relevance in forecasting match results and understanding game dynamics.

In the fantasy cricket team selection component, a rule-based system is used to evaluate and compare players from two teams using performance metrics such as runs scored, wickets taken, strike rate, and economy. Fantasy points are allocated based on predefined rules for batting, bowling, and fielding contributions.This model is expected to give best team for a match considering player match-ups against each other.

The project also introduces video analysis modules for deeper match insights. A ball tracking system using YOLOv8 detects and predicts the trajectory of the cricket ball from video footage, identifying events like bounces and calculating angular motion. Separately, an LBW (Leg Before Wicket) detection system utilizes computer vision techniques (OpenCV and cvzone) to analyze ball movement relative to the batsman and pitch, flagging potential LBW incidents. Complementing these tools is an NLP-powered chatbot, built using spaCy and NLTK, capable of answering cricket-related queries by extracting data through SQL-based retrieval from structured cricket databases. Scheduled data ingestion with Apache Airflow ensures the chatbot remains up to date with the latest statistics.

3.8 Work Division

Aarya Purushothaman- Optimal team selection and Web application integration.

Adithyakrishnan A- Cricket ball trajectory and Dashboard development.

Aleena Roy - Match win Prediction and Dashboard devolopment.

Ali Thalhathe - Implementation of chatbot and LBW detection.

3.9 Project Timeline

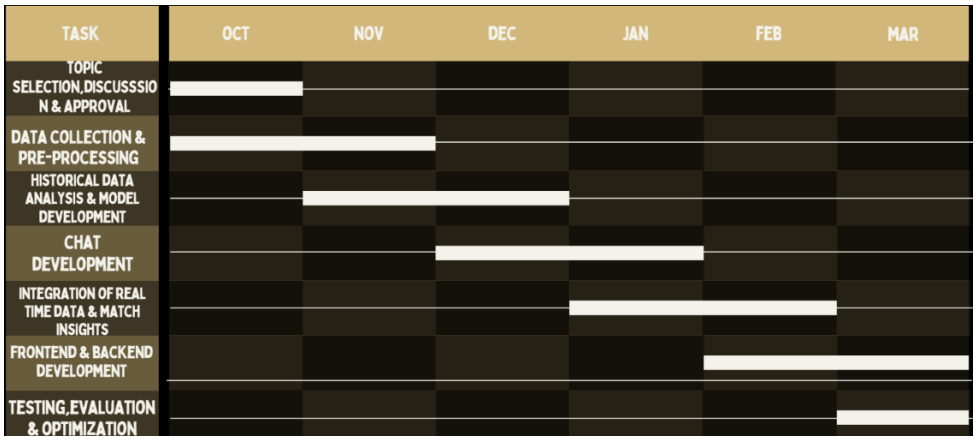


Figure 3.4: Project Schedule

Chapter 4

Results and Discussions

The primary objective of this project was to develop an advanced cricket analytics platform integrating real-time data processing, machine learning predictions, and interactive dashboards.

4.1 Optimized Cricket Team for a Particular Match

The optimized cricket team was generated using a rule-based system that scored players according to consistent performance metrics such as runs scored, wickets taken, boundaries hit, and fielding points. This has resulted in a statistically well and balanced Playing XI meant for fantasy league scoring systems.

The strength of this rule-based approach lies in its simplicity and speed. Users can fast-track their decision-making, relying on hard facts supplied by the scoring logic rather than dive into any manual data crunching. Decisions become fair since they are based purely on objective performance data for an evaluated list of players.

The drawback is that being static, this model does not cater dynamically for match-specific conditions such as pitch happenings, weather changes, or last-minute injury fitness tests. Player form changes might apply adversely when overall season stats are fortified, while these might not be taken into account during real-time game selection. However, given these limitations, correct would be the fantasy user wanting a dependable baseline team selection strategy.

Future versions could have predictive analytics and real-time data feeds integrated to provide more adaptability and accuracy.

4.2 Fantasy Points Distribution Across Player Roles

From the analysis, it is clear that all-rounders have a huge overall impact since they can contribute to both batting and bowling. In this respect, all-rounders are important in team selection. Batsmen performed well under good conditions but lacked consistency, while bowlers maintained a steady flow of points by regularly taking wickets.

Although wicketkeepers were able to add fewer points overall, they added much value by catching, stumping, and scoring quick runs when batting higher up in the order.

The above findings prove that a balanced team has the best luck maximizing fantasy points; therefore, the more the all-rounders in such a composition, the better it is supplemented by batsmen and bowlers in form.

4.2.1 Match Outcome Prediction

The match outcome prediction module relied upon machine learning models which are trained on factors such as toss results, venue, team strength, and recent player forms. Among the tested models, Random Forest stood out in accuracy and consistency; a boastful ability of handling data imbalance and thereby capturing the non-linear interaction among the features.

The predictive potential is useful for fantasy users as it factors in the match result prediction[20], which affects their team combinations. However, the model has its limitations, not taking into account the influence of real-time factors on sudden injuries, pitch changes, or weather.

Nevertheless, being reliable as a benchmark for match prediction, the model accentuates the everyday applications of machine learning in sports analytics. Improvements should try to include live-match data and finer detail on player metrics.

4.2.2 Chatbot Performance and User Interaction

The chatbot, integrated using FAISS and LLaMA with LangChain, answered very rightly and relevantly to the queries raised by end-user from the cricket dataset. It had enabled fast, context-with knowledge retrieval of elements such as player stats, match results, and teams summaries from CSV data converted into vector embeddings.

Minimal touches were needed in the User side as the bot was able to take multiple

natural languages queries. It also differed from usual FAQ-based bots in one other good factor[21], i.e. by providing better personalization and flexibility. However, it was entirely dependent on the dataset's quality; every outdated or incomplete data had the possibility of affecting the accuracy of answers derived.

Overall, chat-bots just opened the front of the user to cricket knowledge instantly and conversationally, making it a great tool for fantasy players as well as cricket lovers. More improvements expected were real-time APIs and adding voice input-output for better accessibility.

4.2.3 LBW Detection and Ball Tracking

Integration of advanced computer vision techniques and machine-learning algorithms was made to promote cricket analytics with the least Indian laws of cricket detection and ball tracking.

LBW Detection:

The LBW detection system combined HSV color filtering for ball tracking with RGB and edge detection for identifying the batsman and pitch area. This enabled detection of "pad-before-wicket" scenarios without relying on very expensive multi-camera setups. The method proved to be efficient and accurate enough for first-order LBW decisions. However, it was sometimes caused, due to interference, by a mixture of lighting conditions, occluding objects, and inconsistent camera angles, which led to false positives or missed detections. In spite of these limitations, the approach has strong potential as a low-cost decision support tool, especially in some amateur or training environments.

Ball Tracking:

The ball tracking module, carried out by YOLOv8, was highly accurate in detecting and tracking the cricket ball through a video frame. The model worked on a custom dataset to show the position of the ball and its trajectory for every frame, even when fast deliveries were done.[17] The system proved to compute the bounce angle and trajectories effectively, making it useful for performance studies and decision-making. In general, the setup was stable; nevertheless, the system somewhat lost accuracy in poor lighting conditions or when the ball blended into the background. Some of the consistency could be improved by the incorporation of motion prediction or multi-frame tracking.

4.3 Result

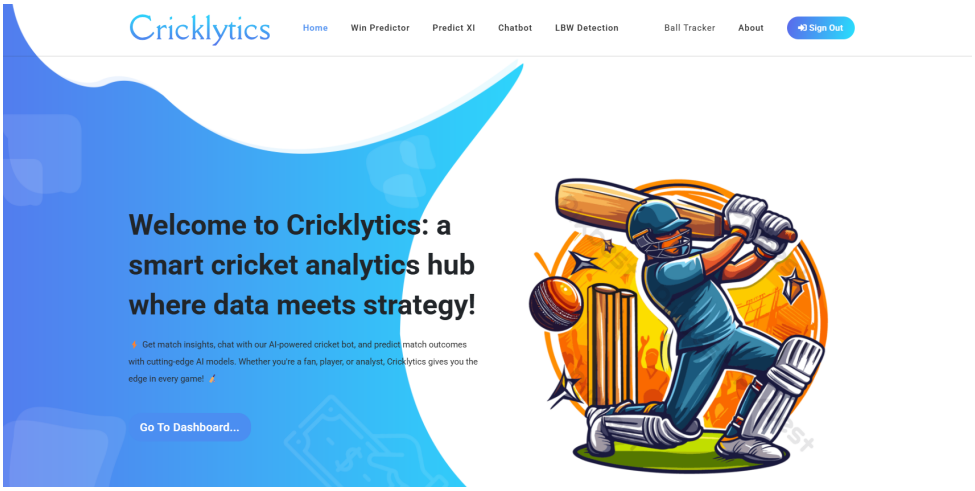


Figure 4.1: Home Page

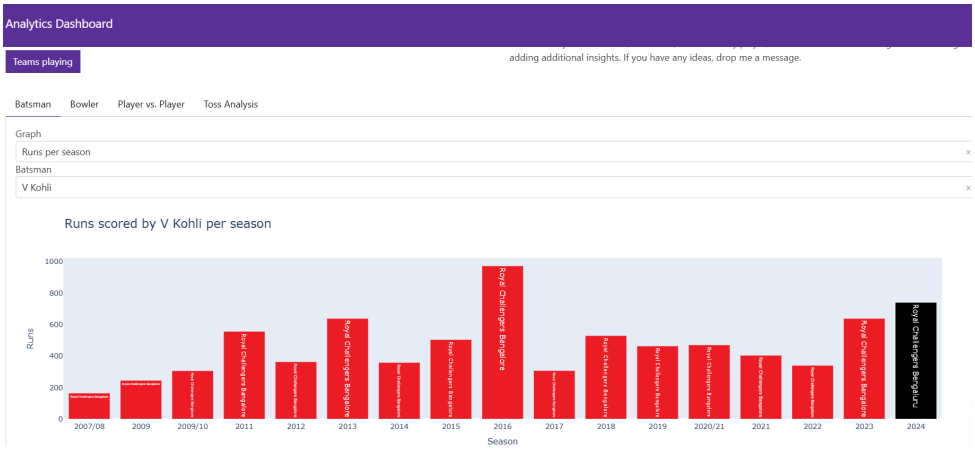


Figure 4.2: Dashboard

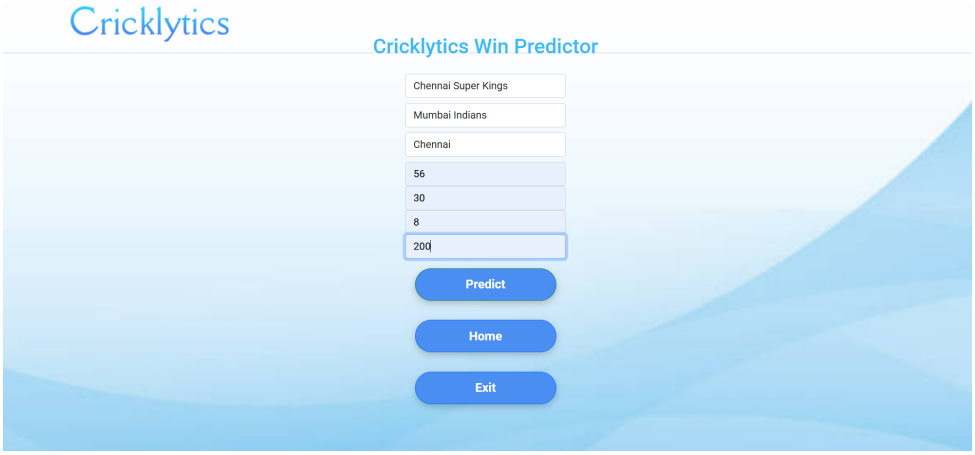


Figure 4.3: Win Prediction

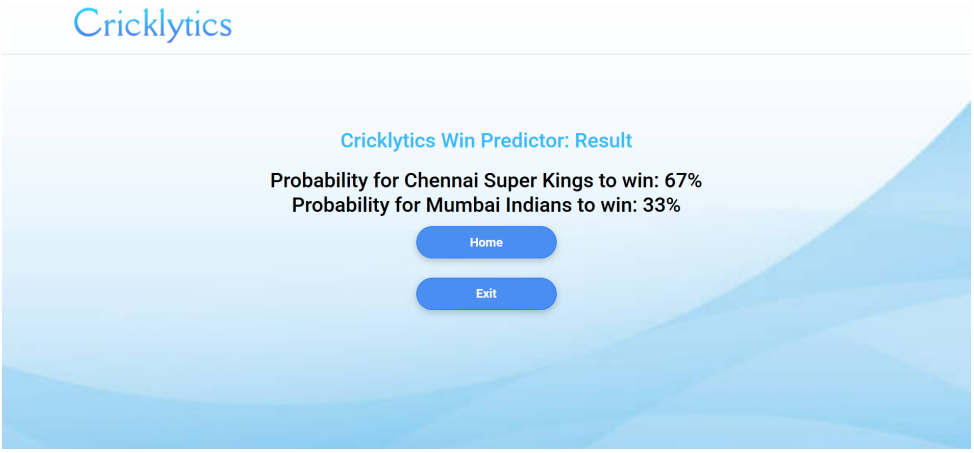


Figure 4.4: Win Prediction Result

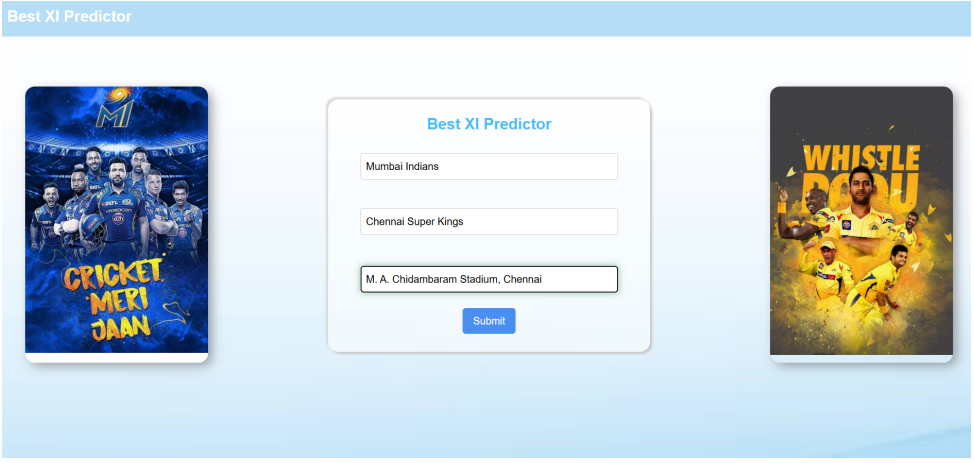


Figure 4.5: Team Selection

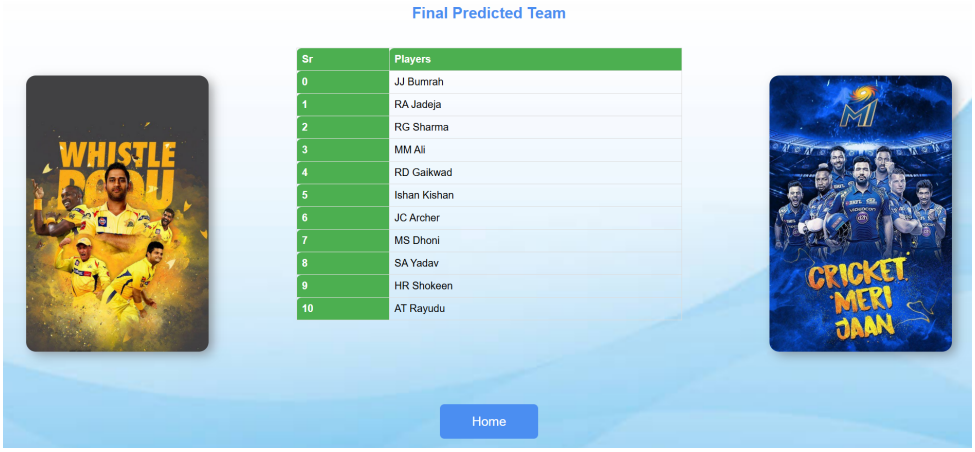


Figure 4.6: Team Selection Result

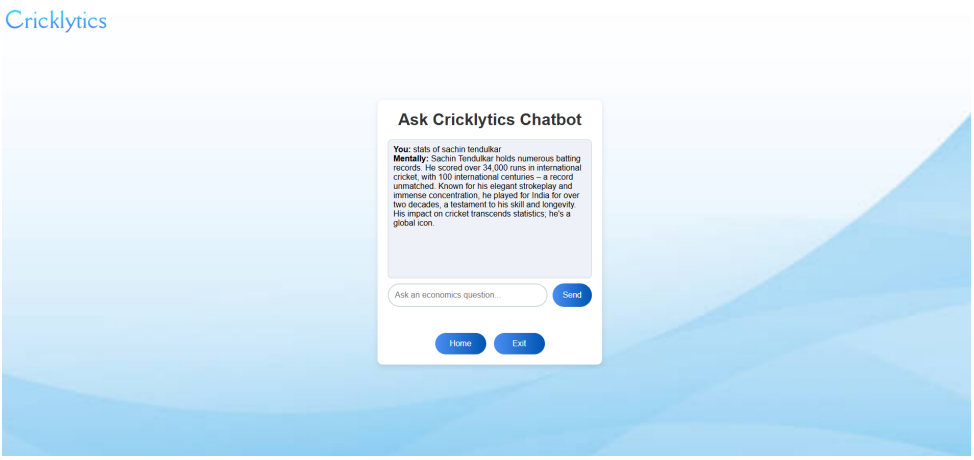


Figure 4.7: Chatbot

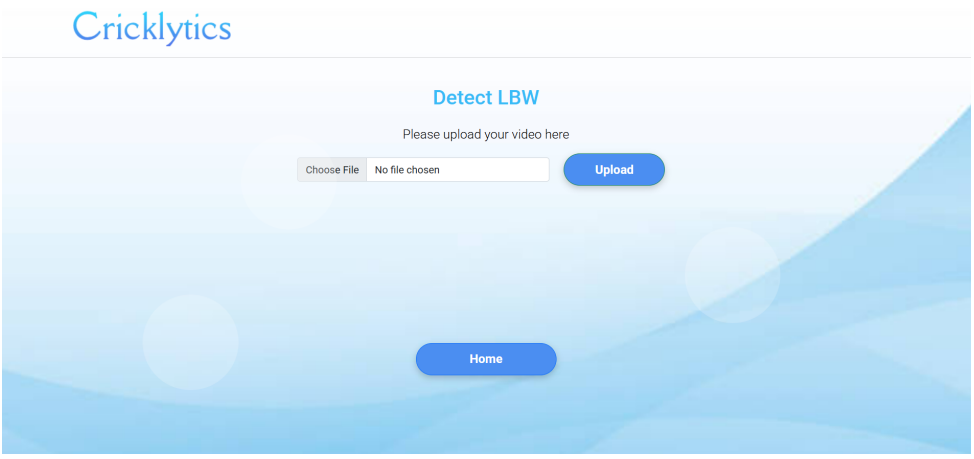


Figure 4.8: LBW Detection

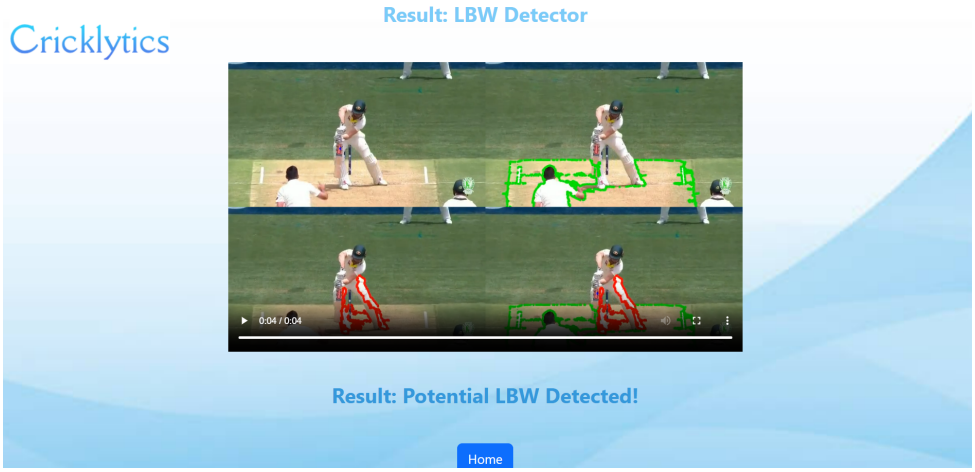


Figure 4.9: LBW Detection Result

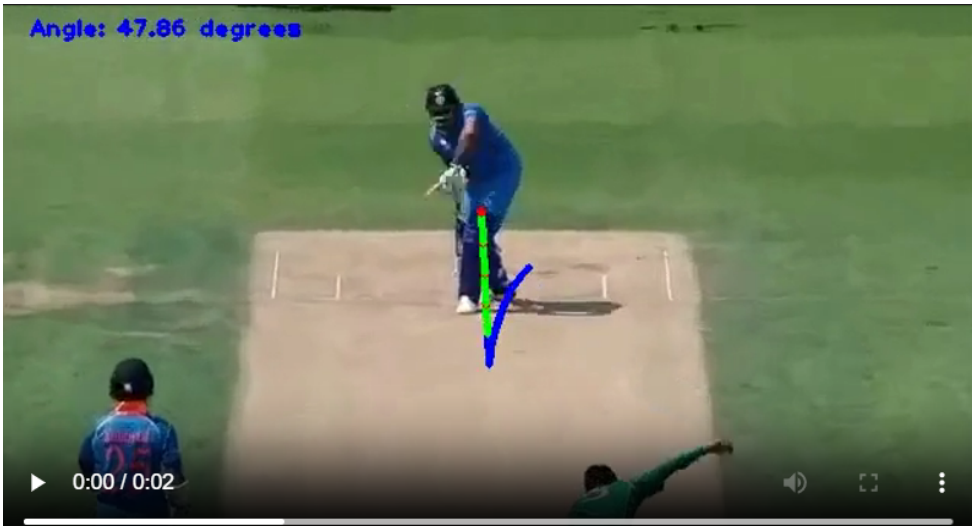


Figure 4.10: Ball Tracking

Chapter 5

Conclusions & Future Scope

The cricket analysis system integrates multiple AI-driven features to enhance match insights and decision-making. The cricket ball tracking and LBW detection modules leverage advanced computer vision techniques to analyze player actions and ball movement, improving accuracy in umpiring decisions. The integrated dashboard provides a centralized platform for visualizing key statistics, enabling users to explore historical and real-time match data. The chatbot, powered by both a pretrained model and Gemini Flash, ensures interactive and informative cricket-related discussions. The match prediction model successfully forecasts outcomes using machine learning, while the team optimization feature suggests the best team lineup based on player performance and match conditions. Together, these features create a comprehensive cricket analytics system, demonstrating the potential of AI in sports technology.

Future developments can focus on enhancing ball tracking and LBW detection by integrating real-time video processing and reinforcement learning for improved accuracy. The dashboard can be expanded to include live match updates and predictive analytics, making it more dynamic. The chatbot can be further trained with real-time data integration and multilingual support for a more personalized user experience. In match prediction, deep learning models can be incorporated for better accuracy by considering factors like pitch conditions, player fatigue, and weather. The team optimization feature can evolve into an AI-driven assistant for team selection, factoring in real-time performance metrics and opposition analysis. Additionally, a mobile-friendly application can make these features more accessible to users worldwide.

References

[1] R. Y. Rajesh and G. Sindhu, "Predict the game analysis of cricket match winning using k-nearest neighbor and compare prediction accuracy over support vector machine," in *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 2022, pp. 685–689.

[2] H. Zhao, H. Chen, S. Yu, and B. Chen, "Multi-objective optimization for football team member selection," *Ieee Access*, vol. 9, pp. 90 475–90 487, 2021.

[3] M. S. Hossain, U. K. Rumpa, M. S. Hossain, S. I. Shovon, M. T. Huque, and M. Hasan, "One day international cricket match score prediction using machine learning approaches," in *2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE)*, 2024, pp. 1–6.

[4] G. Karatas, O. Demir, and O. K. Sahingoz, "Increasing the performance of machine learning-based idss on an imbalanced and up-to-date dataset," *IEEE access*, vol. 8, pp. 32 150–32 162, 2020.

[5] A. Chauhan and V. Bhatia, "Cricket activity detection using computer vision," in *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. Waknaghat, India: IEEE, 2020, pp. 428–433.

[6] M. M. Hatharasinghe and G. Poravi, "Data mining and machine learning in cricket match outcome prediction: Missing links," in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*. IEEE, March 2019, pp. 1–4.

[7] M. Boeker and C. Midoglu, "Soccer athlete data visualization and analysis with an interactive dashboard," in *MultiMedia Modeling. MMM 2023*, ser. Lecture Notes in Computer Science, D.-T. Dang-Nguyen *et al.*, Eds., vol. 13833. Springer, Cham, 2023.

[8] Z. U. Rehman, M. M. Iqbal, H. Safwan, and J. Iqbal, "Predict the match outcome in cricket matches using machine learning," *MJAIAS*, vol. 3, no. 1, May 2022.

[9] A. K. J. Rani, "Ipl auction player price prediction using machine learning regression algorithms," 2019, iEEE Manuscript.

[10] R. Y. Rajesh and G. Sindhu, "Predict the game analysis of cricket match winning using k-nearest neighbor and compare prediction accuracy over support vector machine," in *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. Greater Noida, India: IEEE, 2022, pp. 685–689.

[11] D. Carlander-Reuterfelt, Á. Carrera, C. A. Iglesias, O. Araque, J. F. S. Rada, and S. Muñoz, "Jaicob: A data science chatbot," *IEEE Access*, vol. 8, pp. 180 672–180 680, 2020.

[12] P. Madnur, P. Shetty, G. Parashetti, S. Varur *et al.*, "Advancing in cricket analytics: Novel approaches for pitch and ball detection employing opencv and yolov8," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*. IEEE, 2024, pp. 1–8.

[13] N. Rodrigues *et al.*, "Cricket squad analysis using multiple random forest regression," in *1st IEEE International Conference on Advanced Information Technology (ICAIT)*. IEEE, 2019, pp. 104–108.

[14] N. Perattur, V. K, V. S. Kushwah, T. P. Sai, S. Rekha, and R. Subramanyam, "Algorithms for performance prediction in cricket – factors influencing the player," in *2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN)*. Indore, India: IEEE, 2024, pp. 1047–1051.

[15] S. Raajesh, N. Martin, J. Jiji, A. Nair, and H. H, "Cricket team selection and player analysis using data analytics," in *2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*. Kothamangalam, Kerala, India: IEEE, 2024, pp. 1–6.

[16] M. S. Hossain, U. K. Rumpa, M. S. Hossain, S. I. Shovon, M. T. Huque, and M. Hasan, "One day international cricket match score prediction using machine learning approaches," in *2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE)*. Rourkela, India: IEEE, 2024, pp. 1–6.

[17] A. Verulkar, S. Hegde, S. Pai, S. Patil, and L. Pramiee, "Enhanced classification of cricket batting shots using advanced machine learning and computer vision techniques," in *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2025, pp. 1–8.

[18] ——, "Enhanced classification of cricket batting shots using advanced machine learning and computer vision techniques," in *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. Bhopal, India: IEEE, 2025, pp. 1–8.

[19] P. R. Kamble, A. G. Keskar, and K. M. Bhurchandi, "Ball tracking in sports: a survey," *Artificial Intelligence Review*, vol. 52, pp. 1655–1705, 2019.

[20] P. Mondal, A. Ghosh, A. Sinha, and S. Goswami, "A study of interrelation between ratings and user reviews in light of classification," in *Advances in Intelligent Systems and Computing*. Springer, 2020, vol. 937, pp. 689–697.

[21] K. Pandya and M. Holia, "Automating customer service using langchain: Building custom open-source gpt chatbot for organizations," *arXiv preprint arXiv:2310.05421*, 2023.

Appendix A: Presentation

CRICKET DATA ANALYSIS AND PREDICTION

1

GUIDE

Ms JISHA MARY JOSE
ASSISTANT PROFESSOR
DEPT OF CSE

TEAM MEMBERS

- ALI THALHATHE U2103027
- ALEENA ROY U2103025
- ADITHYAKRISHNAN A U2103015
- AARYA PURUSHOTHAMAN U2103003

PURPOSE & NEED

2

- The purpose of this project is to develop a web application that uses advanced analytics and machine learning to analyze cricket data, providing actionable insights for match outcomes and player performance.

- This system will empower cricket teams, analysts, and fans to make informed decisions, develop strategies , and enhance overall engagement with the game.

OBJECTIVE

3

- Develop an interactive dashboard to visualize IPL player statistics using bar charts, pie charts, and scatter plots.

- Integrate historical player and team data to predict match outcomes.

- Optimize fantasy cricket team selection using predefined scoring rules for batting, bowling, and fielding.

- Implement a chatbot for user interaction, providing real-time updates, forecasts, and historical insights.

- Utilize YOLO for ball trajectory tracking and angle detection in cricket.

- Apply computer vision techniques for LBW detection.

LITERATURE SURVEY

Title	Methodologies Used
Multi-Objective Optimization for Football Team Member Selection	Genetic Algorithm (GA) for multi-objective optimization in team composition.
Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset	Machine Learning algorithms including K-Nearest Neighbor (KNN), Random Forest (RF), Gradient Boosting (GB), Adaboost, Decision Tree (DT) and Linear Discriminant Analysis (LDA); Synthetic Minority Oversampling Technique (SMOTE) for addressing data imbalance.
JAICOB: A Data Science Chatbot	Modular cognitive agent architecture; Machine Learning models; Natural Language Understanding (NLU) algorithms.
Advancing in Cricket Analytics: Novel Approaches for Pitch and Ball Detection Employing OpenCV and YOLOv8	OpenCV for image processing; YOLOv8 for object detection and tracking.
Computer Vision and Image Segmentation: LBW Automation Technique	HSV color filtering and contour analysis for ball detection; RGB color-based detection and edge detection for batsman identification; Contour-based techniques for pitch detection; Logic implementation for LBW scenario prediction.

PROPOSED METHODS

Data Collection and Integration

- Integrate data from sources like available datasets ,espncricinfo.
- Use ETL (Extract, Transform, Load) processes to gather, clean, and transform both real-time and historical data for analysis.

Dashboard Design and Development

- IPL cricket data is analyzed using Python, Pandas, and Plotly to create interactive visualizations of player performance.
- The system merges ball-by-ball and match-level data to generate insights through bar charts, pie charts, scatter plots, and tables, helping users understand trends and optimize team selection.

6

Machine Learning Model Development

- Train and deploy ML models such as Logistic Regression, Random Forest, KNN and Decision tree to predict match outcomes and player performance.
- Compared and evaluated all the models and chose the best one in terms of accuracy.

Fantasy Cricket Team Building

- The system calculates player fantasy points based on batting, bowling, and fielding performances, with adjustments for recent form and performance against specific opponents.
- It combines historical performance (50%) and calculates current form (50%) to rank players, ensuring an optimal and adaptive fantasy team selection.

7

Cricket ball Tracking

- YOLOv8 is used for object detection and tracking to analyze the cricket ball's trajectory in videos.
- The system calculates the ball’s motion angle and detects key events like bounces for performance analysis.
- A custom dataset was created by extracting frames from bowling videos and using various cricket ball images for training.

Cricket LBW Detection

- Detects the cricket ball using HSV color filtering and contour analysis.
- Identifies the batsman and pitch using RGB-based filtering, edge detection, and contour analysis.
- Implements an LBW classification function to analyze ball motion and detect potential pad-before-wicket scenarios.

Interactive Chatbot

- The cricket-related data is loaded from a CSV file, converted into vector embeddings using a pre-trained Hugging Face model, and stored in a FAISS vector store for efficient information retrieval.
- User queries are processed by a LLaMA language model integrated with a RetrievalQA chain, which retrieves the most relevant documents from the vector store and generates cricket-specific responses based on the provided context.
- This approach ensures that all answers are drawn from the provided data, enabling users to get accurate insights related to cricket in an automated and interactive manner.

ARCHITECTURE DIAGRAM

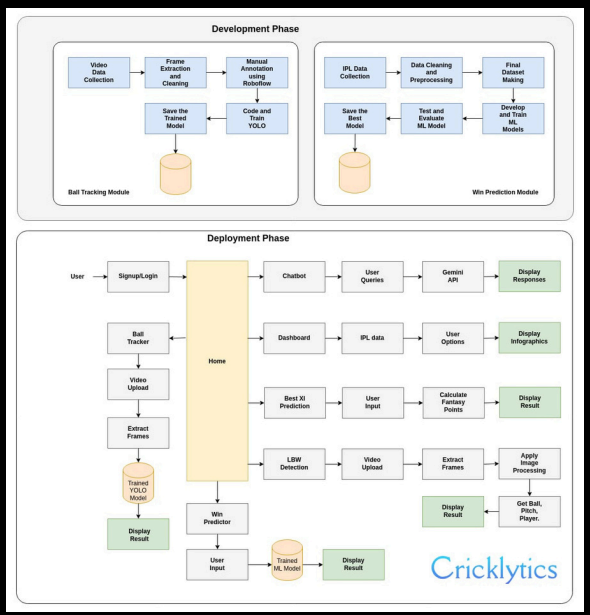


fig1-Architecture diagram

SEQUENCE DIAGRAM

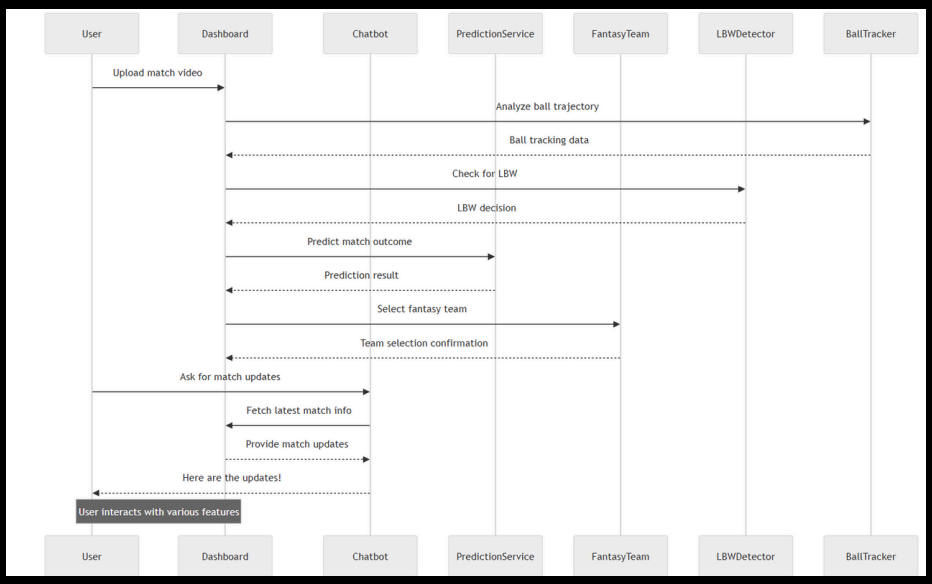


fig2-sequence diagram

MODULE WISE DIVISION

1 Dashboard Design and Development

- Data Collection & Preprocessing: Extract and clean ball-by-ball and match-level IPL data for analysis.
- Statistical Analysis & Insights: Analyze player performance (runs, wickets, economy rates) and trends like player vs player, toss impact, and venue-based performance.
- Visualization & Dashboard Development: Develop interactive visualizations (bar charts, pie charts, scatter plots, tables) using Python, Pandas, and Plotly.
- User Interaction & Decision Support: Create a user-friendly interface for data-driven decision-making for fans, analysts, and fantasy players.

2 Match outcome prediction

- Train and deploy ML models such as Logistic Regression, Random Forest, KNN and Decision tree to predict match outcomes and player performance.
- New features along with historical match data to analyze patterns and make predictions about which team is likely to win given the current situation.
- The trained models was assessed using various measures to check its accuracy and effectiveness in predicting match outcomes. The results were stored for comparison
- A Comparison of all four models were done on the basis of their precision, accuracy, recall and F1 score.

3 Fantasy Cricket Team Building

- Player Performance Calculation: Compute fantasy points based on batting, bowling, and fielding contributions (runs, wickets, economy rates, catches, etc.).
- Recent Form Adjustment: Integrate recent player form into the fantasy score using logarithmic scaling to adjust scores.
- Opponent Analysis: Factor in player performance against individual bowlers and teams, applying penalties for unfavorable matchups.
- Weighted Scoring: Combine historical performance (50%) and current form (50%) to generate an adaptive and accurate fantasy score.
- Player Ranking & Team Selection: Rank players by fantasy points and select the top performers to form the optimal fantasy team lineup.

4 Interactive chatbot development

1.CSV Data Loading and Vectorization:

- Load cricket-related data from a CSV using CSVLoader.
- Convert data into vector embeddings with HuggingFaceEmbeddings.
- Store embeddings in FAISS for fast similarity search and retrieval.

2.Question-Answering Setup and LLaMA Integration:

- Integrate LLaMA 2 language model with RetrievalQA chain for efficient query answering.
- Use FAISS to retrieve relevant documents based on the query.
- Custom PromptTemplate guides the model to provide cricket-specific answers.

3.Answer Generation and Contextual Response:

- Generate responses based on the retrieved context from FAISS and the LLaMA model.
- Ensure all answers are grounded in the CSV data, ensuring relevance and accuracy.
- Control response creativity using the temperature parameter.

5 Cricket Ball Tracking

- Data Collection & Dataset Creation: Manually extract frames from different bowling videos and use various cricket ball images to create a custom dataset for training the model.

- Object Detection & Tracking: Implement YOLOv8 for detecting and tracking the cricket ball in video frames, analyzing its trajectory and motion.

- Angle Calculation & Event Detection: Calculate the ball's motion angle and detect key events such as bounces for performance analysis.

- Real-Time Application: Integrate the system for real-time cricket analysis, including player performance analysis and ball trajectory tracking.

6 Cricket LBW Detection

- Ball Detection: Utilizes HSV color filtering and contour analysis to detect the cricket ball in video frames.
- Batsman Detection: Combines RGB color-based detection and edge detection techniques to identify the batsman, with tunable color ranges and Canny thresholds for optimization.
- Pitch Detection: Employs contour-based techniques to identify the pitch area in the video feed.
- LBW Detection Logic: Implements a function to analyze ball motion and predict potential LBW scenarios, focusing on "pad before wicket" events.

ASSUMPTIONS

1 Advanced Player and Team Analysis

It is assumed that cricket analytics dashboard provides detailed insights into player performance using IPL match and ball-by-ball data. It allows users to visualize stats like runs, wickets, strike rate, and dismissal types interactively for both batsmen and bowlers.

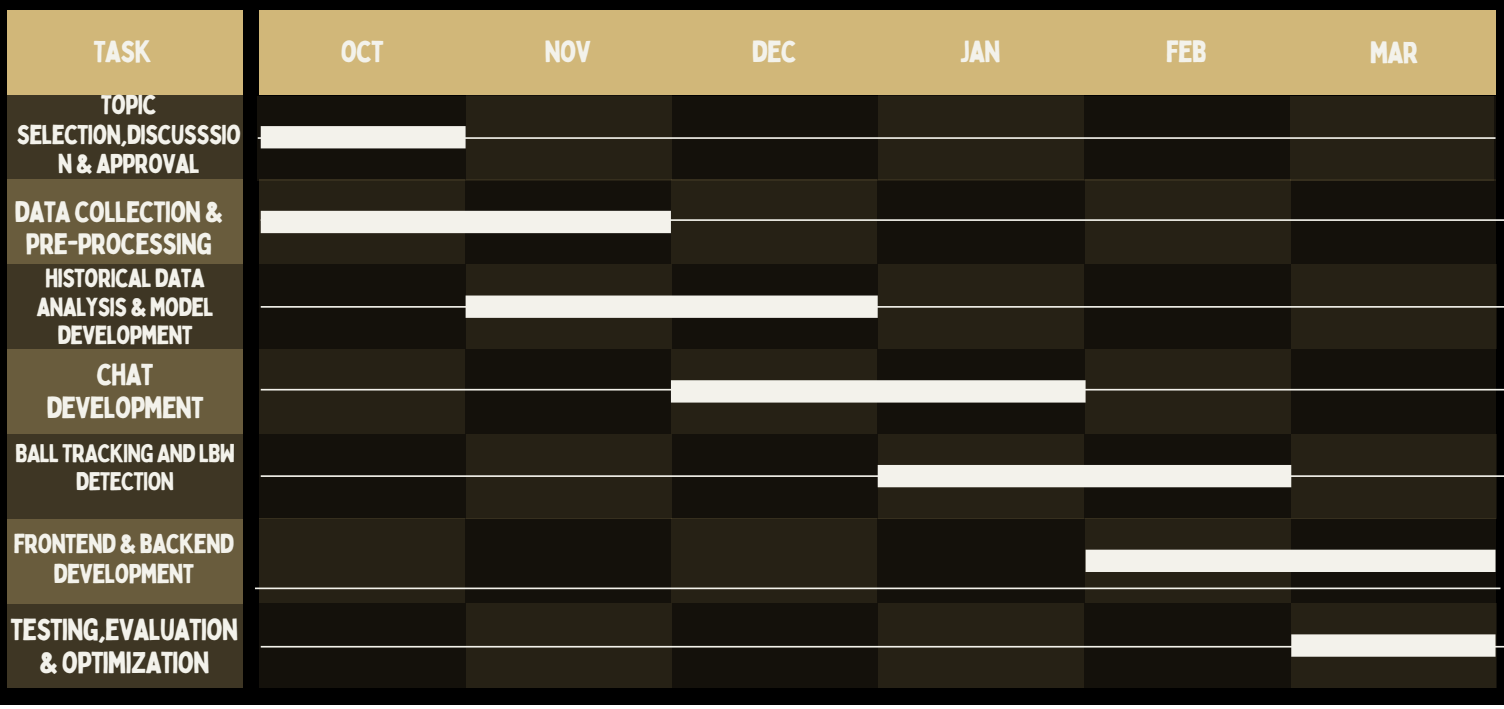
2 Data Privacy and Security

The platform assumes full compliance with data privacy regulations and secure handling of user data.It is assumed that all personal information from users (if collected) and the data being processed are stored and handled securely, complying with relevant data protection laws.

SOFTWARE/HARDWARE REQUIREMENTS

- Google Colab Notebook
- YOLOv8 Model
- Python Libraries (dash,opencv,pandas,Matplotlib)
- R Libraries(ggplot2,plotly)
- HTML/CSS/Javascript/Flask

GANTT CHART



RISK & CHALLENGES

1.Data Quality and Availability:

- Cricket data might not always be available in structured or clean formats, especially historical data.
- Inconsistent, incomplete, or unavailable data, especially real-time data, can lead to inaccuracies in analysis and predictions.

2.Model accuracy and complexity:

- Cricket is influenced by various factors like pitch conditions, player form, weather, and strategies, making it hard to develop highly accurate models.
- Factors like unexpected player injuries or weather changes are unpredictable and difficult to account for accurately in models.

EXPECTED OUTCOME

The expected outcome of this project is to develop a comprehensive cricket analytics platform that will significantly enhance the fan experience.

1 PREDICTION USING RANDOM CLASSIFIER

- Predicts match outcomes, player performances, and key events using historical data and Random Forest classifiers.
- Provides probabilistic predictions based on patterns found in past IPL data, offering insights into match trends and team dynamics.

2 INTERACTIVE CHATBOT

- Offers real-time responses to user queries related to cricket matches, player statistics, and predictions.
- Uses natural language processing to enhance user interaction, offering a conversational interface for accessing historical data and forecasts.

3 CRICKET BALL TRACKING

- Detects and tracks the trajectory of the cricket ball in videos using YOLOv8, providing insights into ball motion and events like bounces.
- Calculates the ball's angle and motion for analyzing performance, with potential real-time applications for player performance improvement.

4 CRICKET LBW DETECTION

- Uses computer vision techniques to detect potential LBW scenarios by analyzing the ball's position, batsman’s location, and pitch in video feeds.
- Predicts "pad before wicket" events using a function that classifies ball motion in relation to the batsman’s stance.

5 INTERACTIVE DASHBOARD

- Provides a visual interface to explore IPL player statistics, trends, and match data, using interactive charts and graphs for analysis.
- Allows users to filter and analyze historical data, player vs player stats, toss impacts, and performance at various venues.

6 TEAM OPTIMIZATION

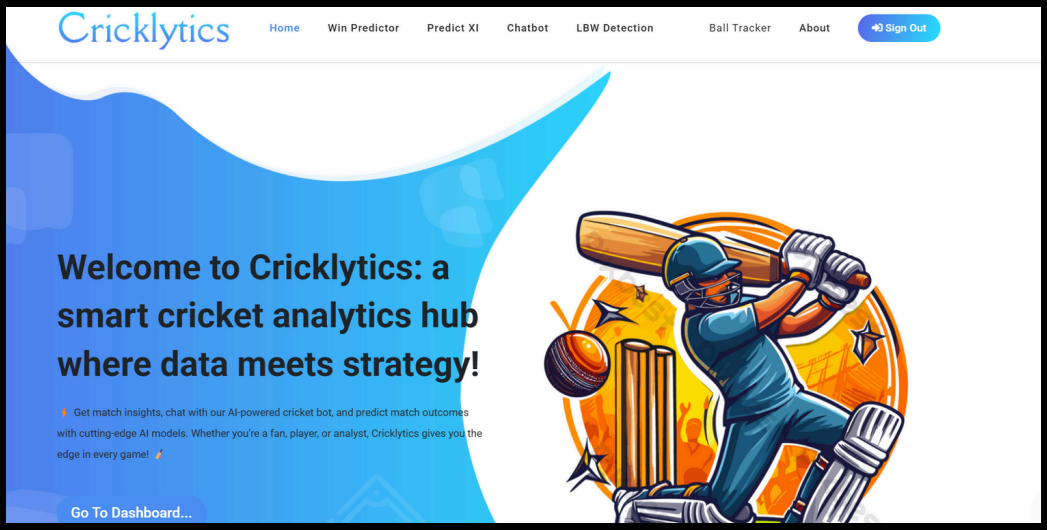
- Optimizes fantasy team selection by calculating player fantasy points based on past performance, including batting, bowling, and fielding statistics.
- Incorporates recent form and opponent performance to adjust player rankings and create the best possible team lineup based on adaptive scoring rules.

WORK BREAKDOWN & RESPONSIBILITIES

1. Aarya Purushothaman
   Optimal team selection and Web application integration.
2. Adithyakrishnan A
   Cricket ball trajectory and Dashboard development.
3. Aleena Roy
   Match win Prediction and Dashboard devolopment.
4. Ali Thalhathe
   Implementation of chatbot and LBW detection.

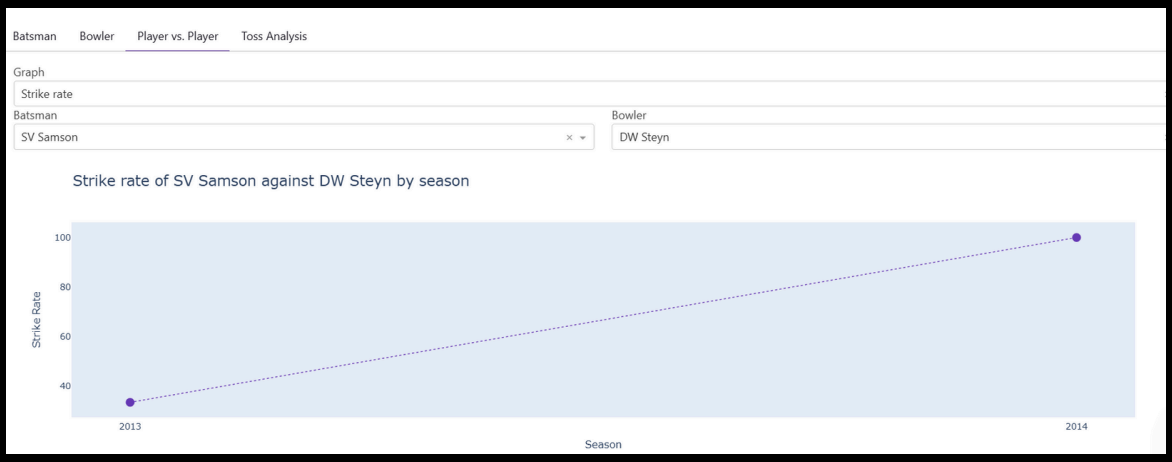
RESULTS

Web Interface

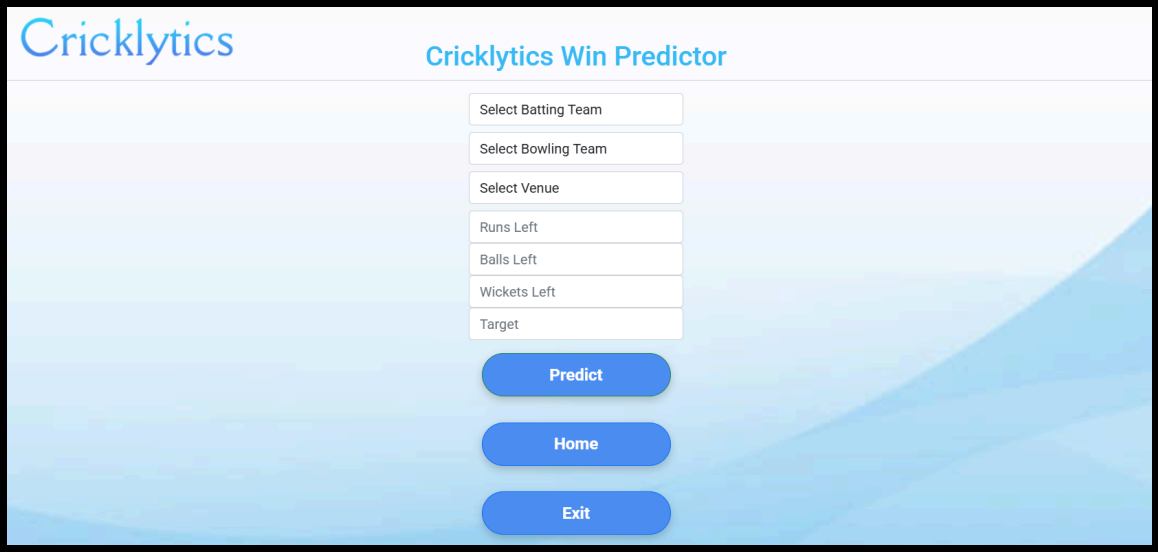


Analytics Dashboard

Player vs Player analysis



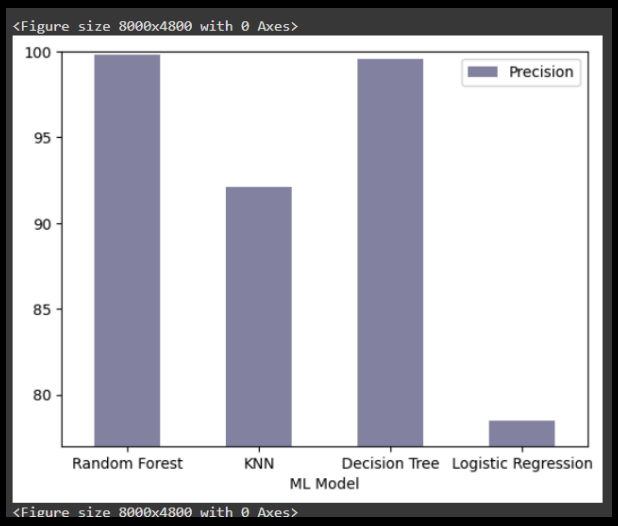
Win predictor



Comparison of different Models

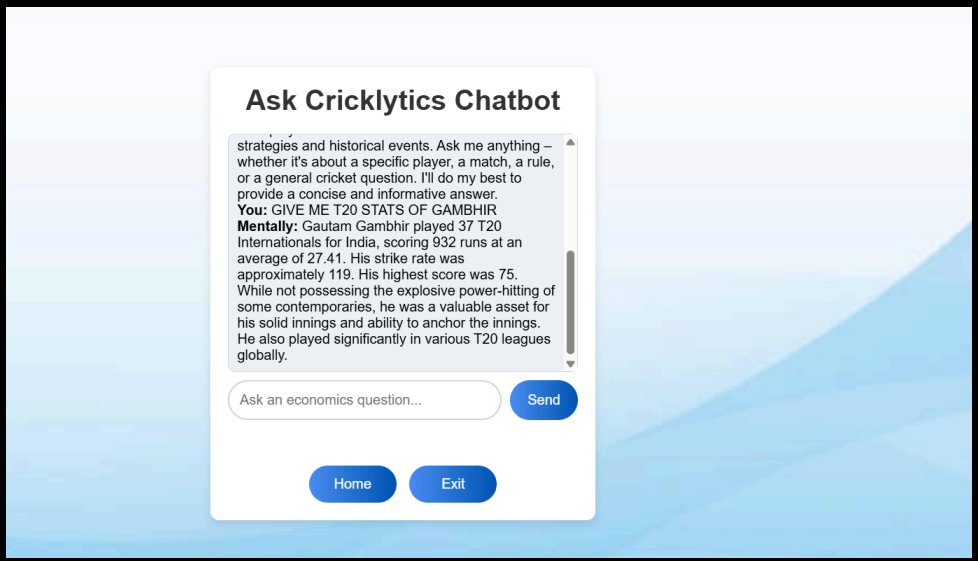
```
result
```

	ML Model	Accuracy	f1_score	Recall	Precision
0	Random Forest	99.892	99.899	100.000	99.798
1	KNN	92.896	93.413	94.737	92.126
2	Decision Tree	99.569	99.595	99.595	99.595
3	Logistic Regression	78.471	80.198	81.984	78.488

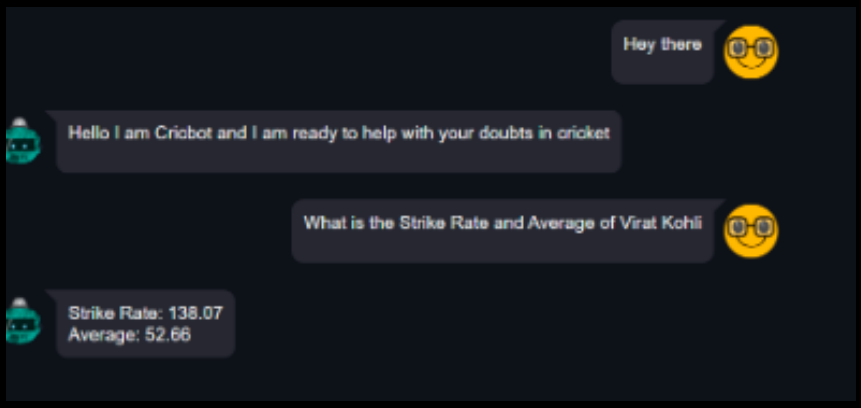


Precision Scores

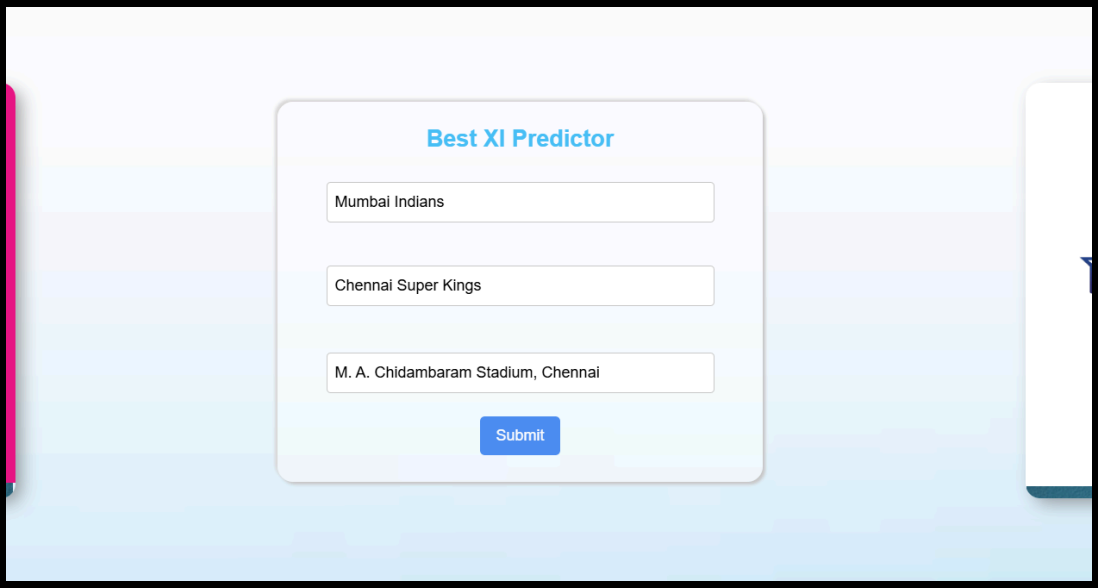
Chatbot answering queries



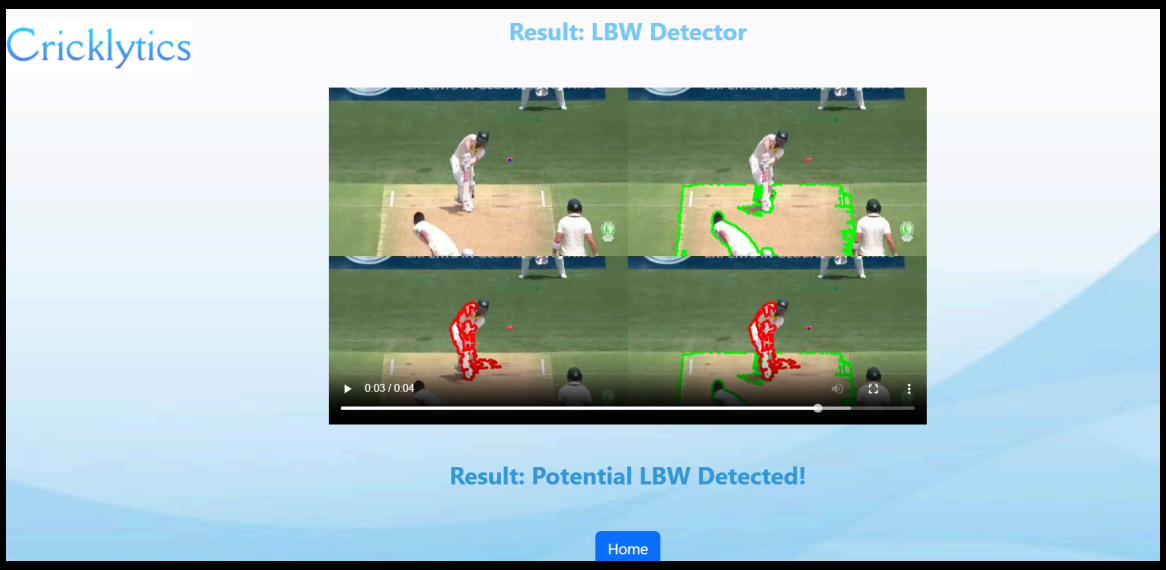
Chatbot using Llama2



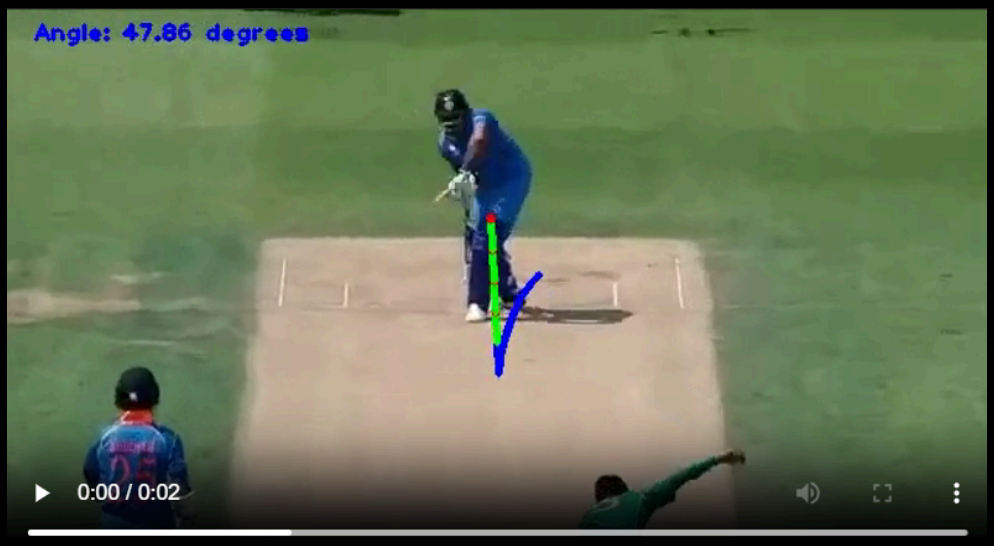
Optimal team selection



LBW Detection



Cricket Ball Tracking



CONCLUSION

In summary, our proposed system revolutionizes cricket data analysis by leveraging historical match data and advanced analytics to provide comprehensive insights.Our approach aims to deliver more accessible and interactive cricket analysis compared to existing systems, offering fans a deeper understanding of the game. The combination of real-time data integration, machine learning-driven predictions, and an intuitive chatbot interface sets a new standard for fan interaction and decision-making in the cricketing world.

REFERENCES

[1] H. Zhao, H. Chen, S. Yu and B. Chen, "Multi-Objective Optimization for Football Team Member Selection," IEEE Access, vol. 9, pp. 90475-90487, 2021, doi: 10.1109/ACCESS.2021.3091185.

[2] G. Karatas et al., “Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset,” IEEE Access, vol. 8, pp. 32150-32162, 2020.

[3] D. Carlander-Reuterfelt et al., "JAICOB: A Data Science Chatbot," IEEE Access, vol. 8, pp. 180672-180680, 2020, doi: 10.1109/ACCESS.2020.3024795.

[4] P. Madnur et al., "Advancing in Cricket Analytics: Novel Approaches for Pitch and Ball Detection Employing OpenCV and YOLOv8," 2024 IEEE 9th Int. Conf. for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-8, doi: 10.1109/I2CT61223.2024.10544224.

[5] J. Nayak et al., "Computer Vision and Image Segmentation: LBW Automation Technique," in Intelligent Systems, ICMIB 2023, Lecture Notes in Networks and Systems, vol 728, Springer, 2024.

THANKYOU

Appendix B: Vision, Mission, Programme Outcomes and Course Outcomes

Vision, Mission, Programme Outcomes and Course Outcomes

Institute Vision

To evolve into a premier technological institution, moulding eminent professionals with creative minds, innovative ideas and sound practical skill, and to shape a future where technology works for the enrichment of mankind.

Institute Mission

To impart state-of-the-art knowledge to individuals in various technological disciplines and to inculcate in them a high degree of social consciousness and human values, thereby enabling them to face the challenges of life with courage and conviction.

Department Vision

To become a centre of excellence in Computer Science and Engineering, moulding professionals catering to the research and professional needs of national and international organizations.

Department Mission

To inspire and nurture students, with up-to-date knowledge in Computer Science and Engineering, ethics, team spirit, leadership abilities, innovation and creativity to come out with solutions meeting societal needs.

Programme Outcomes (PO)

Engineering Graduates will be able to:

1. Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern Tool Usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and Team work: Function effectively as an individual, and as a member or leader in teams, and in multidisciplinary settings.

10. Communication: Communicate effectively with the engineering community and with society at large. Be able to comprehend and write effective reports documentation. Make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team. Manage projects in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Programme Specific Outcomes (PSO)

A graduate of the Computer Science and Engineering Program will demonstrate:

PSO1: Computer Science Specific Skills
The ability to identify, analyze and design solutions for complex engineering problems in multidisciplinary areas by understanding the core principles and concepts of computer science and thereby engage in national grand challenges.

PSO2: Programming and Software Development Skills
The ability to acquire programming efficiency by designing algorithms and applying standard practices in software project development to deliver quality software products meeting the demands of the industry.

PSO3: Professional Skills
The ability to apply the fundamentals of computer science in competitive research and to develop innovative products to meet the societal needs thereby evolving as an eminent researcher and entrepreneur.

Course Outcomes (CO)

Course Outcome 1: Model and solve real world problems by applying knowledge across domains (Cognitive knowledge level: Apply).

Course Outcome 2: Develop products, processes or technologies for sustainable and socially relevant applications (Cognitive knowledge level: Apply).

Course Outcome 3: Function effectively as an individual and as a leader in diverse teams and to comprehend and execute designated tasks (Cognitive knowledge level: Apply).

Course Outcome 4: Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level: Apply).

Course Outcome 5: Identify technology/research gaps and propose innovative/creative solutions (Cognitive knowledge level: Analyze).

Course Outcome 6: Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level: Apply).

Appendix C: CO-PO-PSO Mapping

COURSE OUTCOMES:

After completion of the course the student will be able to

SL.NO	DESCRIPTION	Blooms' Taxonomy Level
CO1	Model and solve real world problems by applying knowledge across domains (Cognitive knowledge level:Apply).	Level 3: Apply
CO2	Develop products, processes or technologies for sustainable and socially relevant applications. (Cognitive knowledge level:Apply).	Level 3: Apply
CO3	Function effectively as an individual and as a leader in diverse teams and to comprehend and execute designated tasks. (Cognitive knowledge level:Apply).	Level 3: Apply
CO4	Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level: Apply).	Level 3: Apply
CO5	Identify technology/research gaps and propose innovative/creative solutions (Cognitive knowledge level:Analyze).	Level 4: Analyze
CO6	Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level:Apply).	Level 3: Apply

CO-PO AND CO-PSO MAPPING

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO1	2	2	2	1	2	2	2	1	1	1	1	2	3		
CO2	2	2	2		1	3	3	1	1		1	1		2	
CO3									3	2	2	1			3
CO4					2			3	2	2	3	2			3
CO5	2	3	3	1	2							1	3		
CO6					2			2	2	3	1	1			3

3/2/1: high/medium/low

JUSTIFICATIONS FOR CO-PO MAPPING

MAPPING	LOW/MEDIUM/HIGH	JUSTIFICATION
101003/ CS722U.1-PO 1	M	Knowledge in the area of technology for project development using various tools results in better modeling.
101003/ CS722U.1-PO 2	M	Knowledge acquired in the selected area of project development can be used to identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions.
101003/ CS722U.1-PO 3	M	Can use the acquired knowledge in designing solutions to complex problems.
101003/ CS722U.1-PO 4	M	Can use the acquired knowledge in designing solutions to complex problems.
101003/ CS722U.1-PO 5	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
101003/ CS722U.1-PO 6	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
101003/ CS722U.1-PO 7	M	Project development based on societal and environmental context solution identification is the need for sustainable development.
101003/ CS722U.1-PO 8	L	Project development should be based on professional ethics and responsibilities.

101003/ CS722U.1-PO 9	L	Project development using a systematic approach based on well defined principles will result in teamwork.
101003/ CS722U.1-PO 10	M	Project brings technological changes in society.
101003/ CS722U.1-PO 11	H	Acquiring knowledge for project development gathers skills in design, analysis, development and implementation of algorithms.
101003/ CS722U.1-PO 12	H	Knowledge for project development contributes engineering skills in computing & information gatherings.
101003/ CS722U.2-PO 1	H	Knowledge acquired for project development will also include systematic planning, developing, testing and implementation in computer science solutions in various domains.
101003/ CS722U.2-PO 2	H	Project design and development using a systematic approach brings knowledge in mathematics and engineering fundamentals.
101003/ CS722U.2-PO 3	H	Identifying, formulating and analyzing the project results in a systematic approach.
101003/ CS722U.2-PO 5	H	Systematic approach is the tip for solving complex problems in various domains.
101003/ CS722U.2-PO 6	H	Systematic approach in the technical and design aspects provide valid conclusions.

101003/ CS722U.2-PO 7	H	Systematic approach in the technical and design aspects demonstrate the knowledge of sustainable development.
101003/ CS722U.2-PO 8	M	Identification and justification of technical aspects of project development demonstrates the need for sustainable development.
101003/ CS722U.2-PO 9	H	Apply professional ethics and responsibilities in engineering practice of development.
101003/ CS722U.2-PO 11	H	Systematic approach also includes effective reporting and documentation which gives clear instructions.
101003/ CS722U.2-PO 12	M	Project development using a systematic approach based on well defined principles will result in better teamwork.
101003/ CS722U.3-PO 9	H	Project development as a team brings the ability to engage in independent and lifelong learning.
101003/ CS722U.3-PO 10	H	Identification, formulation and justification in technical aspects will be based on acquiring skills in design and development of algorithms.
101003/ CS722U.3-PO 11	H	Identification, formulation and justification in technical aspects provides the betterment of life in various domains.
101003/ CS722U.3-PO 12	H	Students are able to interpret, improve and redefine technical aspects with mathematics, science and engineering fundamentals for the solutions of complex problems.

101003/ CS722U.4-PO 5	H	Students are able to interpret, improve and redefine technical aspects with identification formulation and analysis of complex problems.
101003/ CS722U.4-PO 8	H	Students are able to interpret, improve and redefine technical aspects to meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
101003/ CS722U.4-PO 9	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
101003/ CS722U.4-PO 10	H	Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools for better products.
101003/ CS722U.4-PO 11	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
101003/ CS722U.4-PO 12	H	Students are able to interpret, improve and redefine technical aspects for demonstrating the knowledge of, and need for sustainable development.
101003/ CS722U.5-PO 1	H	Students are able to interpret, improve and redefine technical aspects, apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

101003/ CS722U.5-PO 2	M	Students are able to interpret, improve and redefine technical aspects, communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
101003/ CS722U.5-PO 3	H	Students are able to interpret, improve and redefine technical aspects to demonstrate knowledge and understanding of the engineering and management principle in multidisciplinary environments.
101003/ CS722U.5-PO 4	H	Students are able to interpret, improve and redefine technical aspects, recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.
101003/ CS722U.5-PO 5	M	Students are able to interpret, improve and redefine technical aspects in acquiring skills to design, analyze and develop algorithms and implement those using high-level programming languages.
101003/ CS722U.5-PO 12	M	Students are able to interpret, improve and redefine technical aspects and contribute their engineering skills in computing and information engineering domains like network design and administration, database design and knowledge engineering.
101003/ CS722U.6-PO 5	M	Students are able to interpret, improve and redefine technical aspects and develop strong skills in systematic planning, developing, testing, implementing and providing IT solutions for different domains which helps in the betterment of life.

101003/ CS722U.6-PO 8	H	Students will be able to associate with a team as an effective team player for the development of technical projects by applying the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
101003/ CS722U.6-PO 9	H	Students will be able to associate with a team as an effective team player for Identify, formulate, review research literature, and analyze complex engineering problems
101003/ CS722U.6-PO 10	M	Students will be able to associate with a team as an effective team player for designing solutions to complex engineering problems and design system components.
101003/ CS722U.6-PO 11	M	Students will be able to associate with a team as an effective team player use research-based knowledge and research methods including design of experiments, analysis and interpretation of data.
101003/ CS722U.6-PO 12	H	Students will be able to associate with a team as an effective team player, applying ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
101003/ CS722U.1-PS O1	H	Students are able to develop Computer Science Specific Skills by modeling and solving problems.
101003/ CS722U.2-PS O2	M	Developing products, processes or technologies for sustainable and socially relevant applications can promote Programming and Software Development Skills.

101003/ CS722U.3-PSO3	H	Working in a team can result in the effective development of Professional Skills.
101003/ CS722U.4-PSO3	H	Planning and scheduling can result in the effective development of Professional Skills.
101003/ CS722U.5-PSO1	H	Students are able to develop Computer Science Specific Skills by creating innovative solutions to problems.
101003/ CS722U.6-PSO3	H	Organizing and communicating technical and scientific findings can help in the effective development of Professional Skills.