

# CRICKET DATA ANALYSIS AND PREDICTION



# 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

- 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

- 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
<b>Multi-Objective Optimization for Football Team Member Selection</b>	Genetic Algorithm (GA) for multi-objective optimization in team composition.
<b>Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset</b>	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.
<b>JAICOB: A Data Science Chatbot</b>	Modular cognitive agent architecture; Machine Learning models; Natural Language Understanding (NLU) algorithms.
<b>Advancing in Cricket Analytics: Novel Approaches for Pitch and Ball Detection Employing OpenCV and YOLOv8</b>	OpenCV for image processing; YOLOv8 for object detection and tracking.
<b>Computer Vision and Image Segmentation: LBW Automation Technique</b>	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 ,espnccricinfo.
- 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.

## 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.

## 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

9

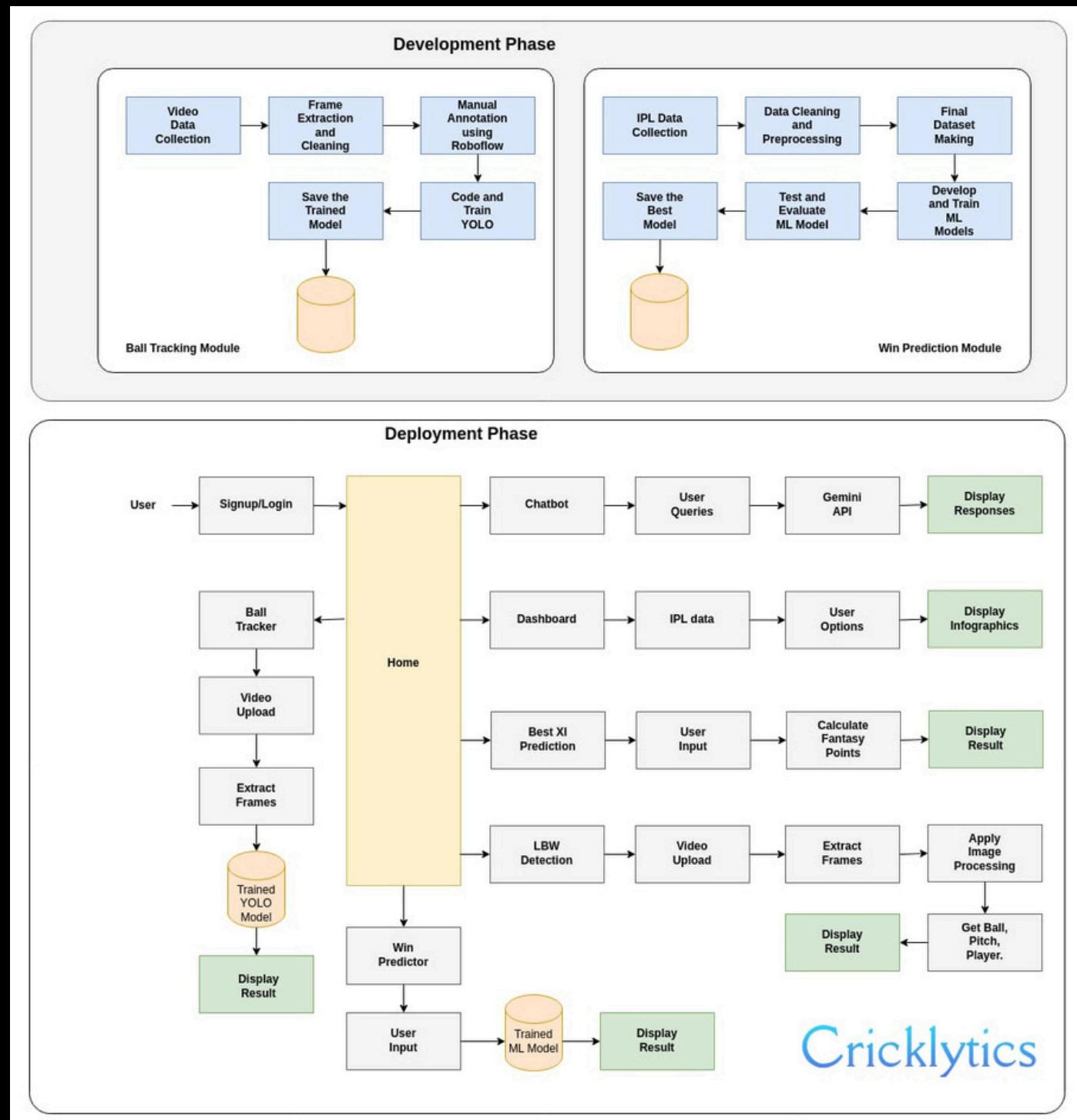


fig1-Architecture diagram

# SEQUENCE DIAGRAM

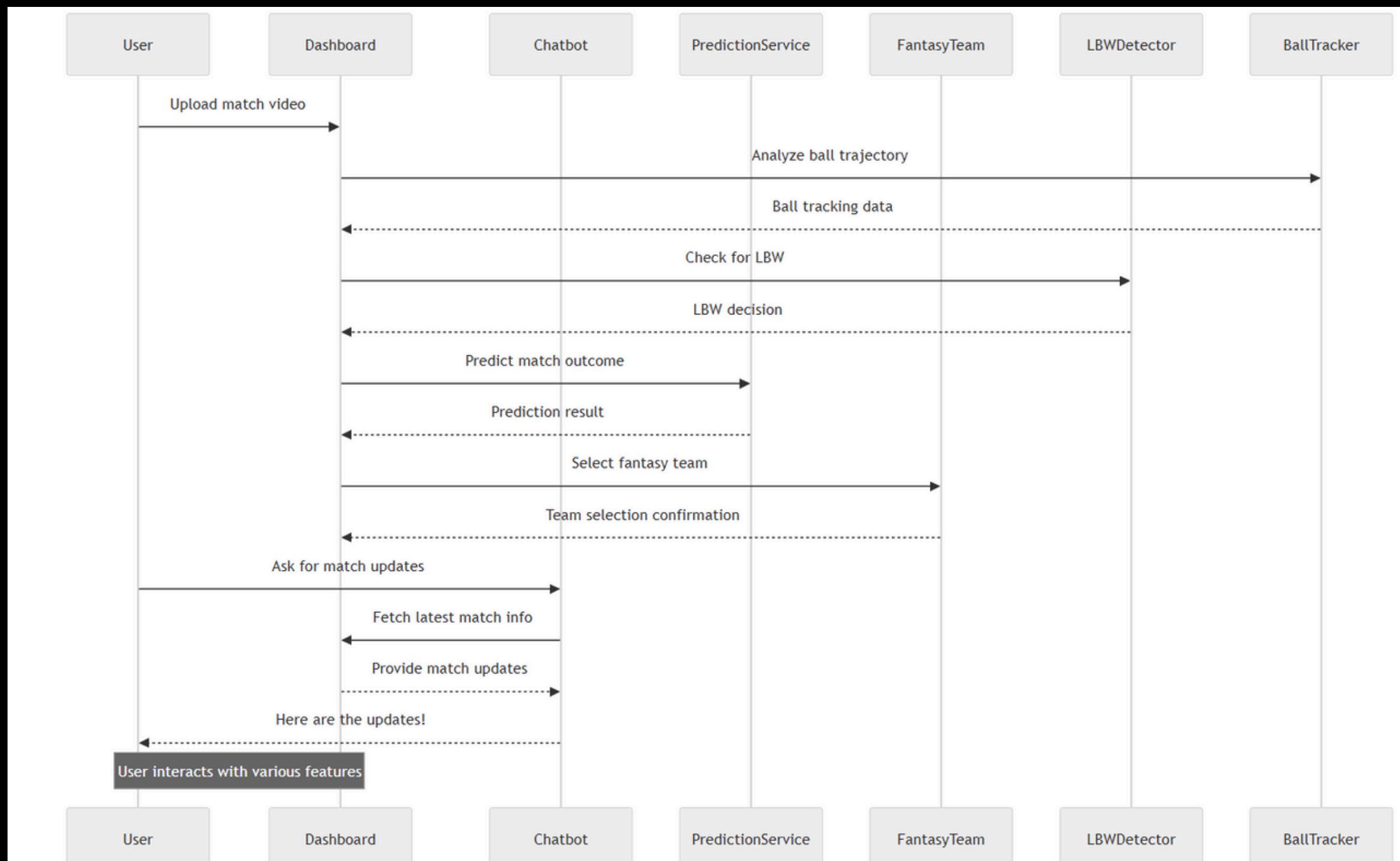


fig2-sequence diagram

# MODULE WISE DIVISION

11

## 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

14

### 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

15

- 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.

- 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.

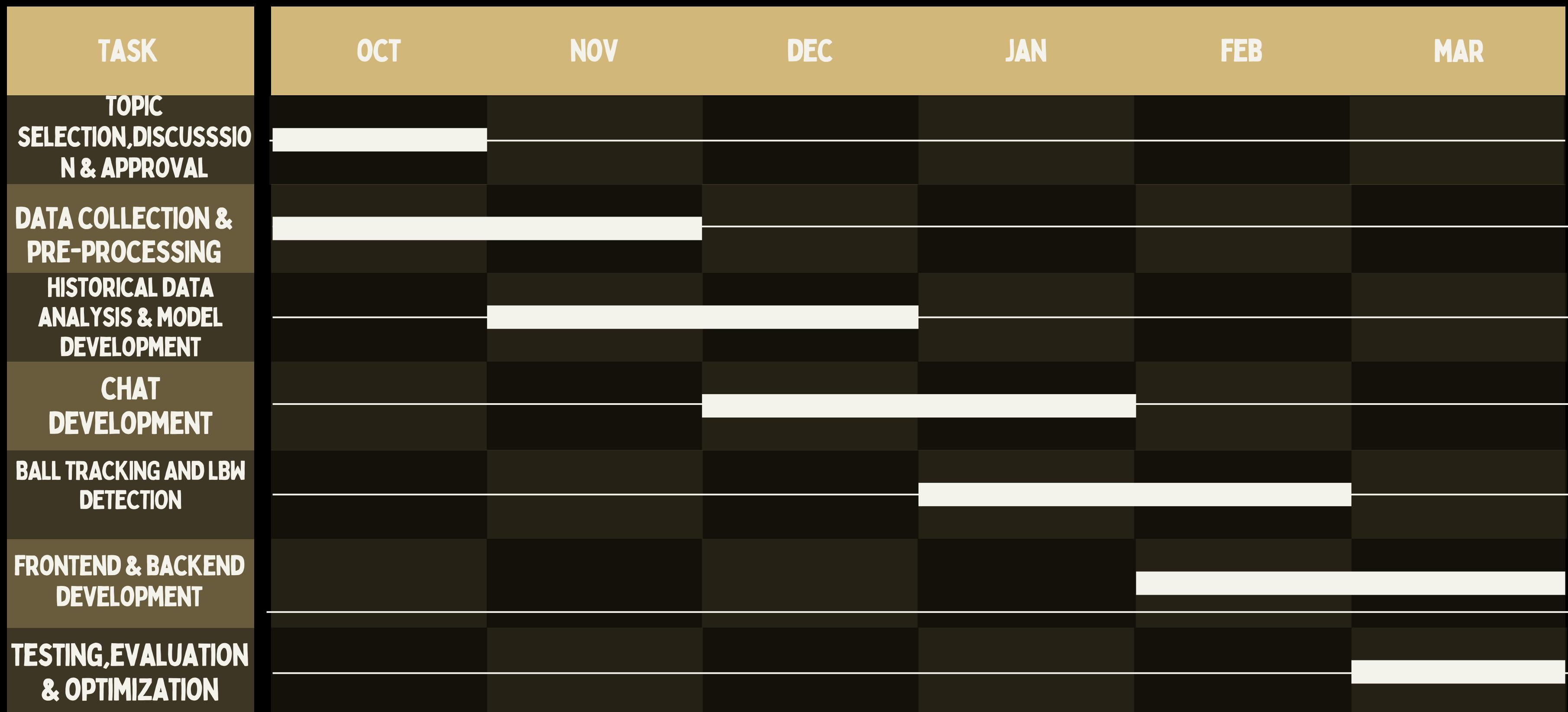
## 2 Data Privacy and Security

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.

## 22 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.

## 33 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.

## 55 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.

## 66 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 development.

4 Ali Thalhathe

Implementation of chatbot and LBW detection.

# RESULTS

# Web Interface

**Cricklytics**

Home Win Predictor Predict XI Chatbot LBW Detection Ball Tracker About [Sign Out](#)

Welcome to Cricklytics: a smart cricket analytics hub where data meets strategy!

Get match insights, chat with our AI-powered cricket bot, and predict match outcomes with cutting-edge AI models. Whether you're a fan, player, or analyst, Cricklytics gives you the edge in every game!

[Go To Dashboard...](#)



# Analytics Dashboard

Analytics Dashboard

Teams playing

adding additional insights. If you have any ideas, drop me a message.

Batsman   Bowler   Player vs. Player   Toss Analysis

Graph

Runs per season

Batsman

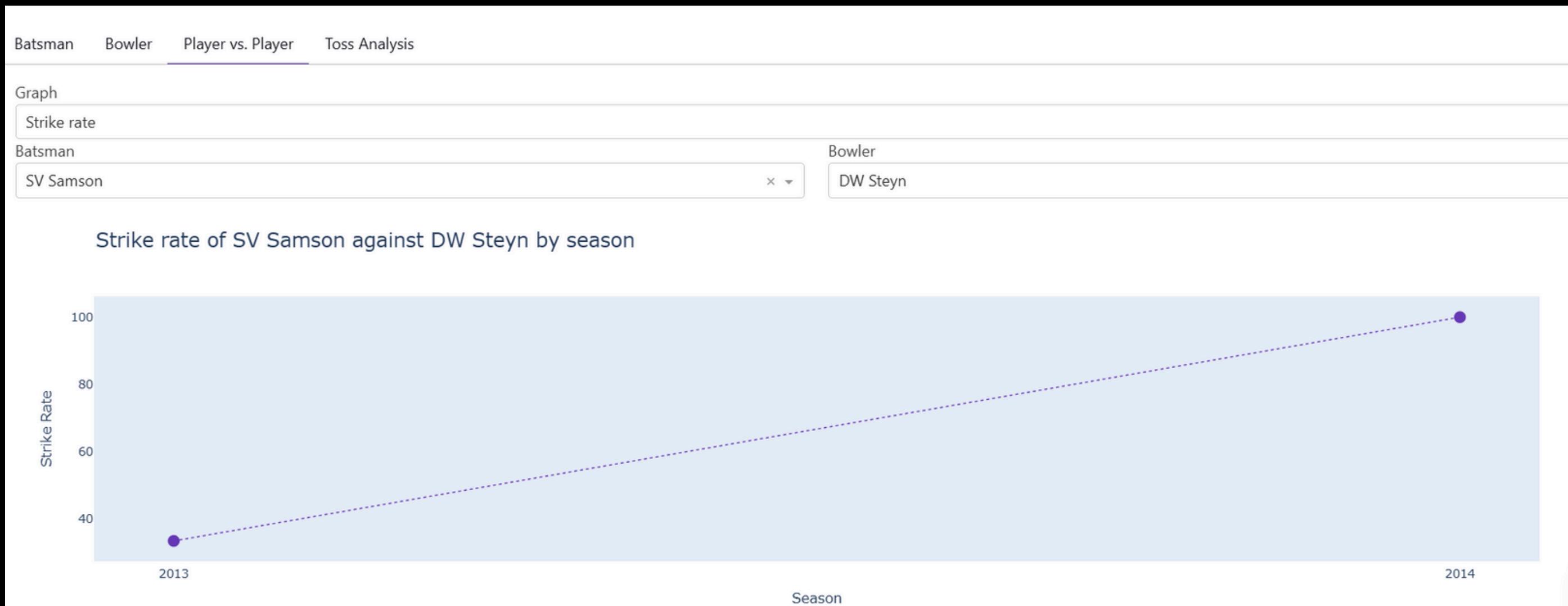
V Kohli

Runs scored by V Kohli per season

A bar chart titled "Runs scored by V Kohli per season" showing the total runs scored by Virat Kohli for each season. The Y-axis represents "Runs" from 0 to 1000, and the X-axis represents the "Season" from 2007/08 to 2024. The chart shows significant fluctuations in his performance over the years, with a notable peak in 2013 and a decline towards the end of the period shown.

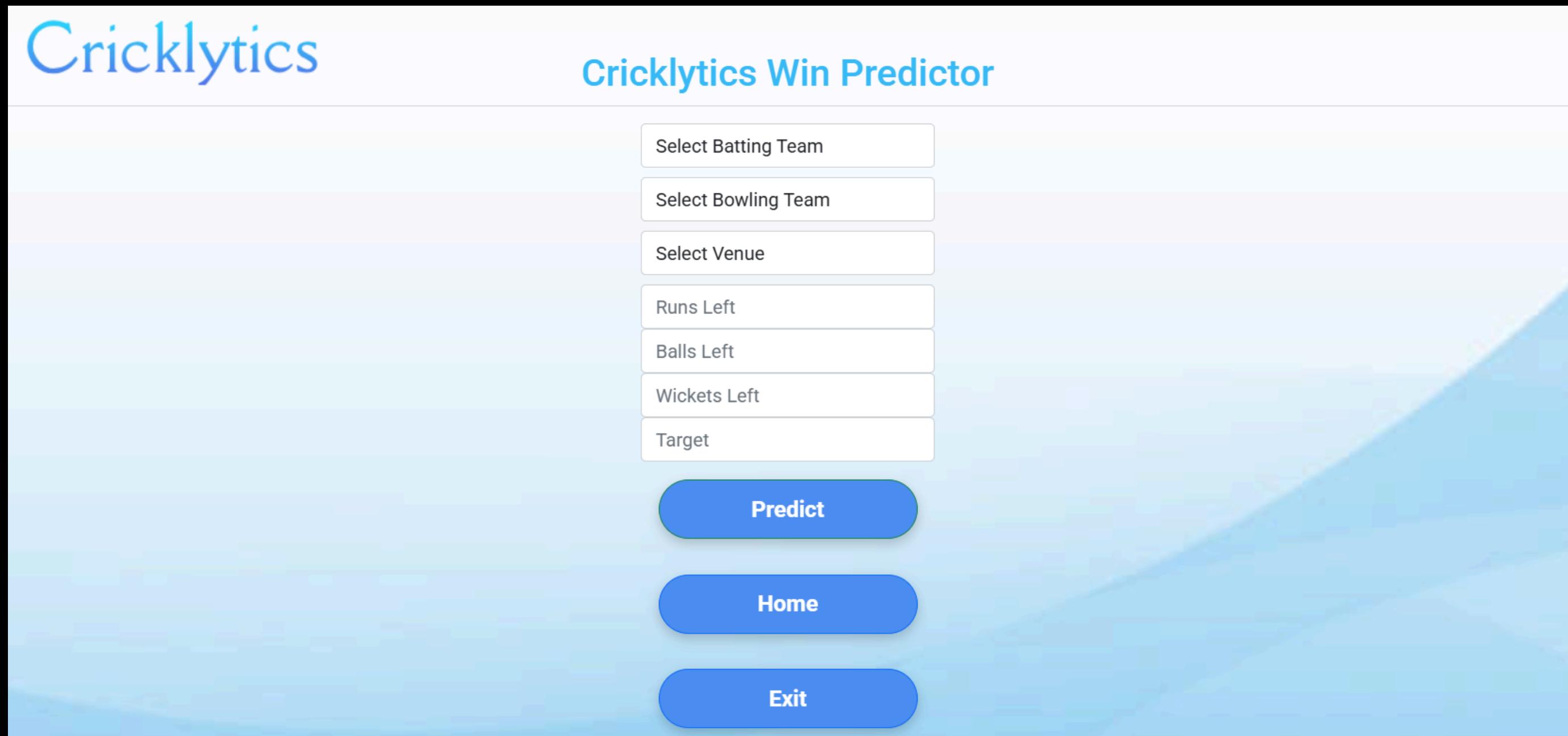
Season	Runs Scored
2007/08	~150
2009	~220
2009/10	~300
2011	~550
2012	~350
2013	~650
2014	~350
2015	~350
2016	~950
2017	~300
2018	~550
2019	~450
2020/21	~400
2021	~350
2022	~350
2023	~600
2024	~750

# Player vs Player analysis



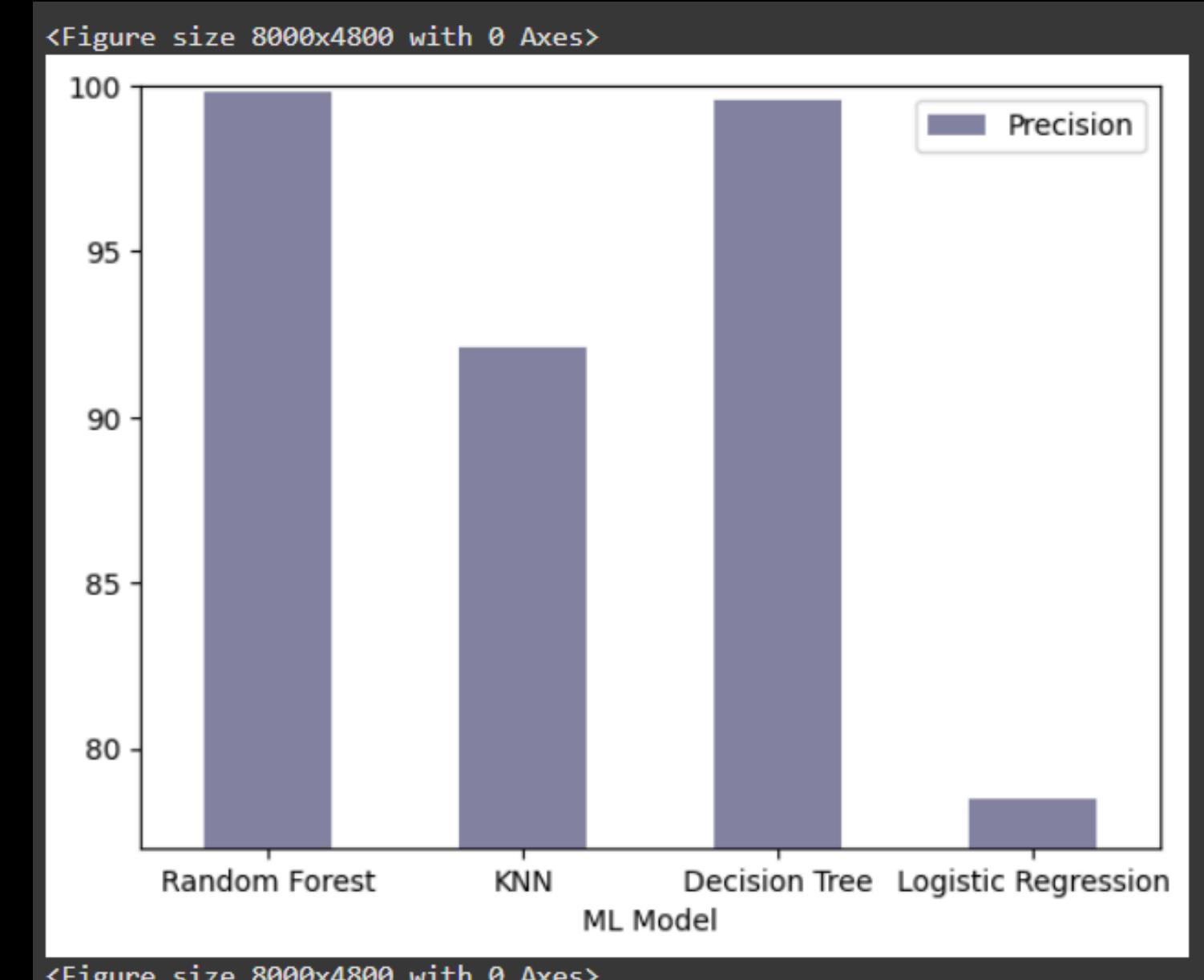


# Win predictor

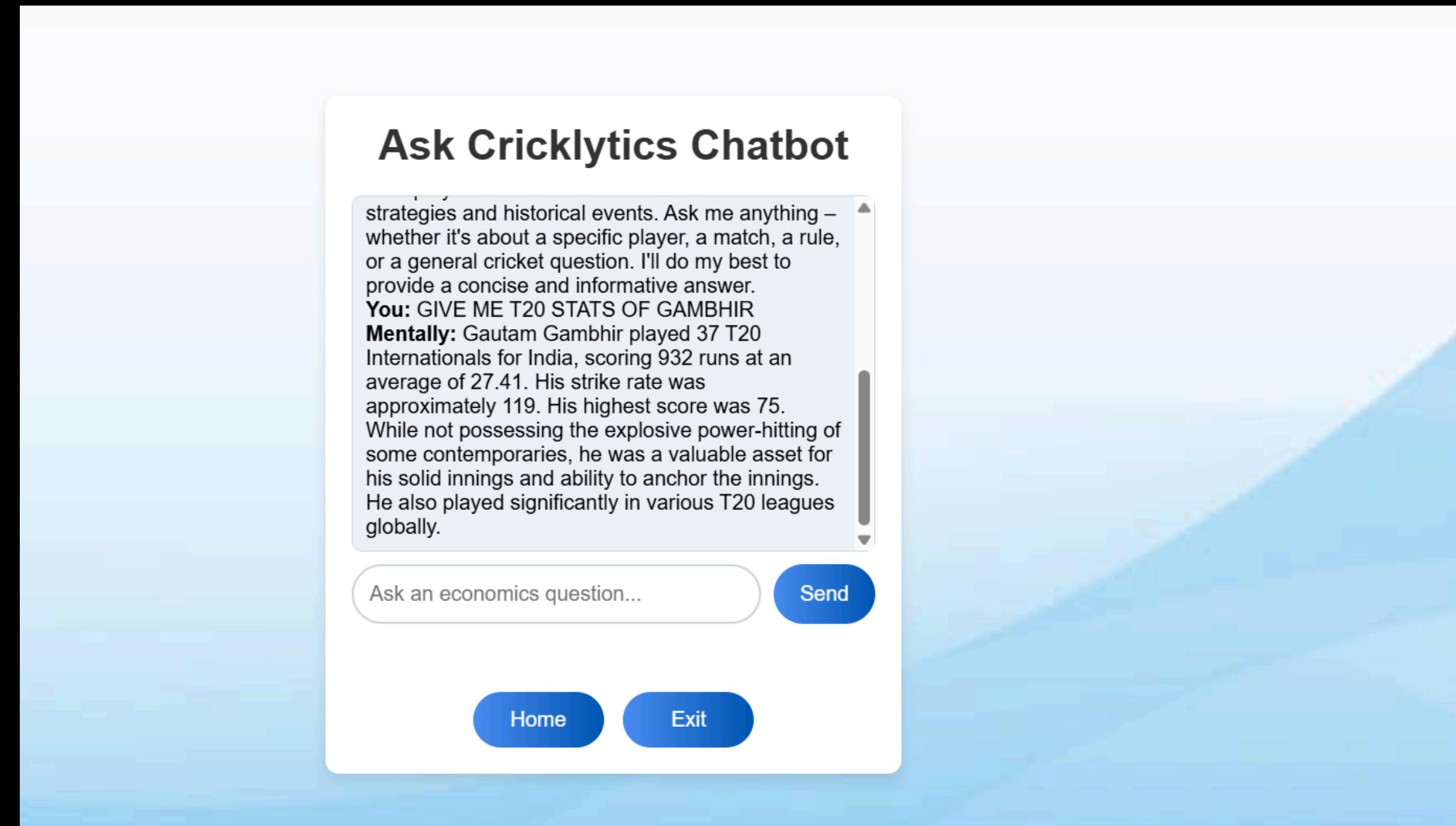


# Comparison of different Models

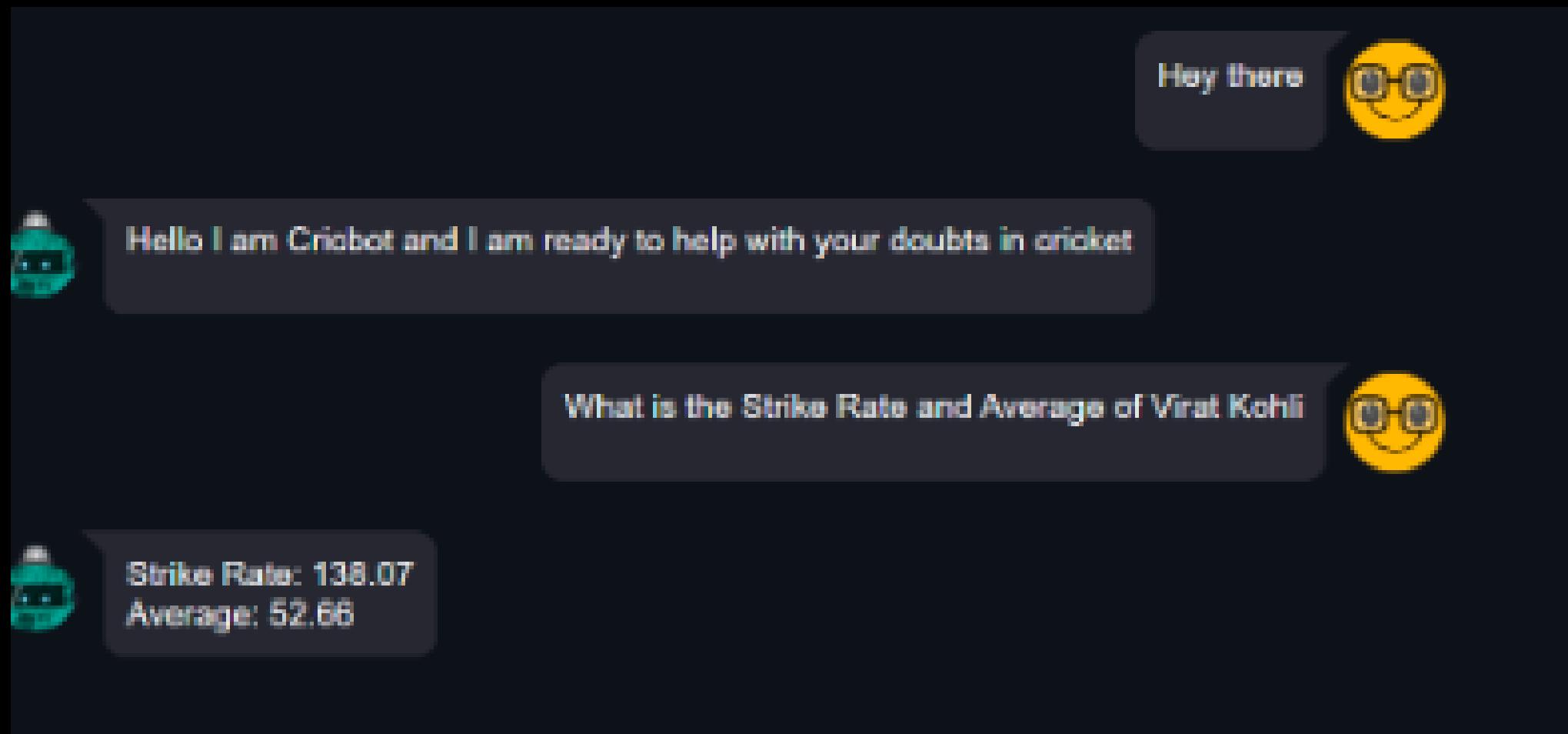
	ML Model	Accuracy	f1_score	Recall	Precision
0	Random Forest	99.892	99.899	100.000	99.798
1	KNN	92.896	93.413	94.737	92.126
2	Decision Tree	99.569	99.595	99.595	99.595
3	Logistic Regression	78.471	80.198	81.984	78.488



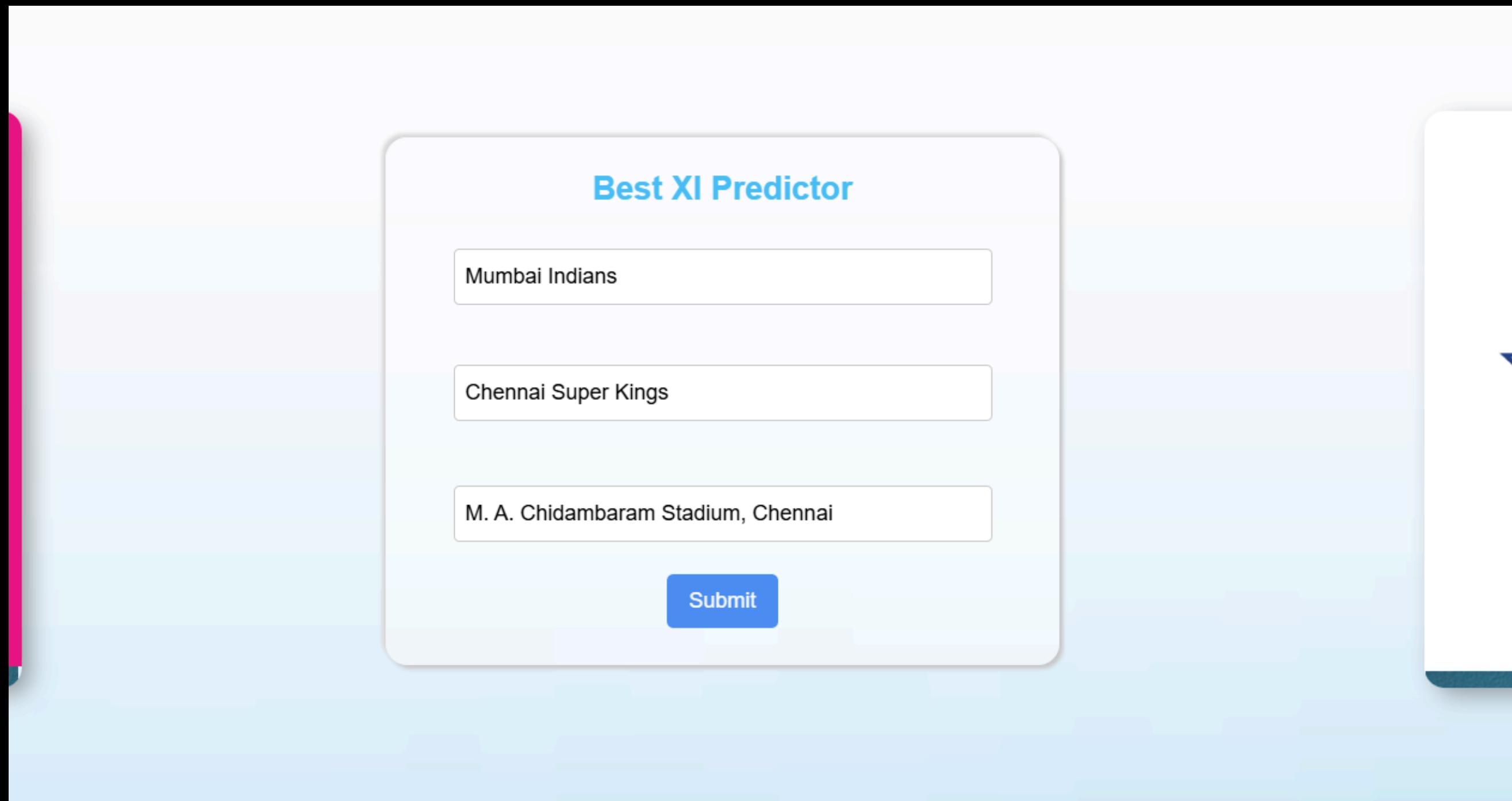
# Chatbot answering queries



# Chatbot using Llama2



# Optimal team selection



### Select Playing 11

Please select exactly 11 players for both teams.

Team 1	Team 2
<input checked="" type="checkbox"/> Ishan Kishan 1	<input checked="" type="checkbox"/> MS Dhoni 1
<input checked="" type="checkbox"/> RG Sharma 2	<input checked="" type="checkbox"/> Devon Conway 2
<input checked="" type="checkbox"/> SA Yadav 3	<input checked="" type="checkbox"/> RD Gaikwad 3
<input checked="" type="checkbox"/> JJ Bumrah 4	<input type="checkbox"/> AT Rayudu
<input checked="" type="checkbox"/> Akash Madhwai 5	<input checked="" type="checkbox"/> Shivam Dube 4
<input type="checkbox"/> Arjun Tendulkar	<input checked="" type="checkbox"/> MM Ali 5
<input checked="" type="checkbox"/> D Brevis 6	<input checked="" type="checkbox"/> RA Jadeja 9
<input type="checkbox"/> HR Shokeen	<input type="checkbox"/> Simarjeet Singh
<input checked="" type="checkbox"/> JP Behrendorff 11	<input checked="" type="checkbox"/> Subhranshu Senapati
<input checked="" type="checkbox"/> JC Archer 7	<input checked="" type="checkbox"/> Matheesha Pathirana
<input type="checkbox"/> K Kartikeya	<input checked="" type="checkbox"/> TU Deshpande 10
<input type="checkbox"/> Arshad Khan	<input type="checkbox"/> Bhagath Varma
<input checked="" type="checkbox"/> Tilak Varma 9	<input type="checkbox"/> Ajay Mandal
<input type="checkbox"/> Ramandeep Singh	<input checked="" type="checkbox"/> KA Jamieson 8
<input checked="" type="checkbox"/> TH David 8	<input type="checkbox"/> Nishant Sindhu
<input type="checkbox"/> T Stubbs	<input type="checkbox"/> Shaik Rasheed
<input checked="" type="checkbox"/> R Goyal 10	<input checked="" type="checkbox"/> BA Stokes 11
<input type="checkbox"/> N Wadhura	<input type="checkbox"/> AM Rahane

### Final Predicted Team

Sr	Players
0	JJ Bumrah
1	RG Sharma
2	RA Jadeja
3	MM Ali
4	BA Stokes
5	SA Yadav
6	TH David
7	Ishan Kishan
8	RD Gaikwad
9	Tilak Varma
10	MS Dhoni

# LBW Detection

Cricklytics

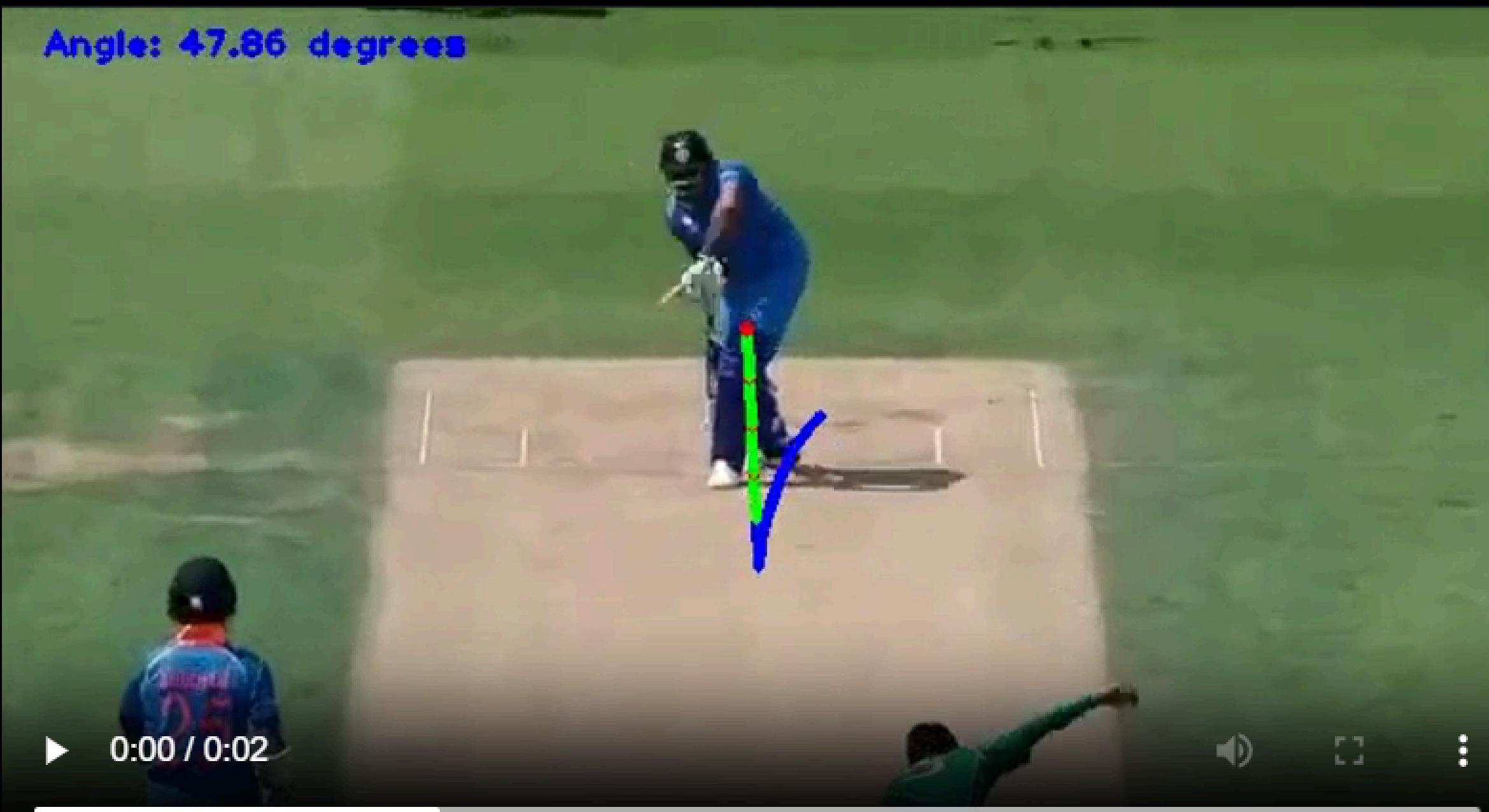
Result: LBW Detector

▶ 0:03 / 0:04

Result: Potential LBW Detected!

Home

# 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

