INDIAN PREMIER LEAGUE
WIN PROBABILITY PREDICTOR

A Project Report Submitted in

Partial fulfilment

of the Degree of

Bachelors of Computer Applications

Supervisor's Name – Dr. Pankaj Agarwal

Submitted by – 1. DEEP SHARMA

ROLL NO. - 2401201211

2. NIHARIKA
ROLL NO. - 2401201211

Semester – 2ND SEMESTER

.

KR MANGALAM UNIVERSITY

PROJECT CERTIFICATE

This is to certify that the project report entitled **Indian Premier League Win Probability Predictor** **s u b m i t t e d** to **<u>KR MANGALAM UNIVERSITY</u>**, HARYANA in partial fulfillment of the requirement for the award of the degree of **Bachelors of Computer Applications** (BCA), is the original work carried out by **Deep Sharma and Niharika, BCA 2nd semester, Session 2024-27,** under the guidance of **Dr. Pankaj Agarwal.** The matter embodied in this Project is a genuine work done by the student and has not been submitted whether to this University or to any other University / Institute for the fulfillment of the requirement of any course of study.

Signature of the Student:

Name of the Student: Deep Sharma and Niharika

Course: BCA (2nd semester)

Session: 2024-27

Supervisor's Signature:

Supervisor's Name: Dr. Pankaj Agarwal

ACKNOWLEDGEMENT

INDEX

S.No.	TOPIC
1	Introduction
2	Objective
3	Tools/Environment
4	System study and Analysis
5	Limitations
6	Future Applications of the Project
7	Bibliography
8	References

1. INTRODUCTION

The game of cricket is played in various formats, i.e., One Day International, T20 and Test Matches. The Indian Premier League (IPL) is a Twenty-20 cricket tournament league established with the objective of promoting cricket in India and thereby nurturing young and talented players. The league is an annual event where teams representing different Indian cities compete against each other. It was started by the Board of Control for Cricket in India (BCCI) and has now become a giant, remunerative cricket venture. The teams for IPL are selected by means of an auction. Players' auctions are not a new phenomenon in the sports world. However, in India, selection of a team from a pool of available players by means of auctioning of players was done in Indian Premier League (IPL) for the first time. Due to the involvement of money, team spirit, city loyalty and a massive fan following, the outcome of matches is very important for all stake holders. This, in turn, is dependent on the complex rules governing the game, luck of the team (Toss),the ability of players and their performances on a given day. Various other natural parameters, such as the historical data related to players, play an integral role in predicting the outcome of a cricket match. A way of predicting the outcome of matches between various teams can aid in the team selection process. However, the varied parameters involved present significant challenges in predicting accurate results of a game. Moreover; the accuracy of a prediction depends on the size of data used for the same. The tool presented in this paper can be used to evaluate the performance of players. This tool provides a visualisation of players' performances. Using IPL T-20 variables related to statistics of batsmen and bowlers, a number of apt variables have been identified that have elucidative power over auction values. Further, several predictive models are also built for predicting the result of a match, based on each player's past performance as well as some match related data. The developed models can help decision makers during the IPL matches to evaluate the strength of a team against another.

2. OBJECTIVES

Predict Winning Probability

Utilize a pre-trained machine learning model to accurately predict the winning probability of the batting and bowling teams based on real-time match data.

Provide Historical Performance Analysis

Display detailed historical performance data for the selected teams.

Analyze trends and patterns in team performance over the years to highlight their strengths and consistency.

Offer Head-to-Head Statistics

Provide comprehensive head-to-head statistics between the selected teams.

Highlight past encounters, outcomes, and key performances to give users deeper insights into the teams' historical rivalries.

Enhance Viewer Engagement

Provide a platform for cricket enthusiasts to analyze matches beyond basic statistics.

Foster a more analytical and informed viewing experience, making the game more enjoyable and interesting.

Visualize Data Effectively

Employ data visualization tools such as Plotly and Matplotlib to create informative and interactive charts.

Visualize team performance data and head-to-head statistics in a clear and engaging manner.

PURPOSE:

The **IPL Victory Predictor** project serves several key purposes, aimed at enhancing the experience and engagement of cricket fans, analysts, and fantasy sports enthusiasts:

Enhance Fan Engagement:

1. **Interactive Analysis**: Provide cricket fans with an interactive platform to analyze IPL matches in real-time, adding a layer of depth and understanding to their viewing experience.

2. **Informed Viewing**: Enable fans to make more informed predictions about match outcomes, increasing their engagement and excitement.

Data-Driven Insights:

1. **Predictive Analytics**: Utilize machine learning to deliver accurate predictions on match outcomes based on current match conditions, historical data, and key variables.

2. **Historical Trends**: Offer insights into team performance over the years, helping users understand trends and patterns in team success and strategies.

Support for Sports Analysts:

1. **Performance Evaluation**: Assist sports analysts in evaluating team performance and strategies using comprehensive data and visualizations.

2. **Strategy Formulation**: Provide a data-backed foundation for formulating match strategies and understanding opposition strengths and weaknesses.

Fantasy Sports Optimization:

1. **Decision Making**: Aid fantasy sports players in making better decisions by providing data-driven insights and predictions, potentially improving their fantasy team performance.

2. **Player Performance Analysis**: Analyze historical and current player performance to optimize fantasy sports selections.

Educational Tool:

1. **Learning Resource**: Serve as an educational resource for those interested in learning about machine learning applications in sports analytics.

3. HARDWARE & SOFTWARE REQUIREMENTS

The software is designed to be light-weighted so that it doesn't be a burden on the machine running it. This system is being build keeping in mind the general availability of hardware and software compatibility. Here are the minimum hardware and software requirement for face recognition system for attendance.

• 3.1 HARDWARE REQUIREMENT:

Processor   : Intel Pentium 4 or above Hard Disk Utilization   :
100 MB or above   2
Input Devices   : Keyboard or Mouse

RAM   : 2 GB or above

• 3.2 SOFTWARE REQUIREMENT:

Operating System   : macOS, Window 8, 8.1, 10, 11, etc.

• 3.3 TECHNOLOGY USED:

Code Compiler   : Jupyter Notebook

□

4. SYSTEM STUDY AND ANALYSIS

System analysis is conducted for the purpose of studying a system or its parts in order to identify its objectives. It's a problem, solving technique that improves the system and ensures that all the components of the system work efficiently to accomplish their purpose.

4.1. Data Collection

Data collection is the process of gathering and measuring information from countless different sources. In order to use the data, we collect to develop practical machine learning solutions. Collecting data allows you to capture a record of past events so that we can use data analysis to find recurring patterns. From those patterns, you build predictive models using machine learning algorithms that look for trends and predict future changes. The Indian Premier League's official website is the principal basis of data for this project. The data was web scrapped from the website and kept in the appropriate format using a python library called beautiful soup. The dataset has the columns regarding match-number, IPL season year, the place where match has been held and the stadium name, the match winner details, participating Predicting Results of IPL Matches using Machine Learning teams, the margin of winning and the umpire details, player of the match etc. Indian Premier League was only 11 years old, which is why, after the pre-processing, only 577 matches were available. Here, some of the columns may contain null values and some of the attributes may not be required for match winner prediction which is discussed in data preprocessing.

4.2 Data Preprocessing

4.2.1 Data cleaning

There are some null values in the dataset in the columns such as winner, city, venue etc. Due to the presence of these null values, the classification cannot be done accurately. So, we tried to replace the null values in different columns with dummy values.

4.2.2 Choosing Required Attributes

This step is the main part where we can eliminate some columns of the dataset that are not useful for the estimation of match winning team. This is estimated using feature importance. The considered attributes have the following feature importance.

4.3 Data Visualization

• The data which has been collected is used for visualizing for the better understanding of the information

☐

• Matplotlib Library is used here for visualizing the graphs

• The data visualization is necessary to understand the solution in a better way. The below graphs were drawn based up on the previous seasons of the IPL matches.

2

Model Development and Evaluation

Here, we have developed a generic model and applied all classification methods. The data is split into training data and test data, we train the model using certain features and use it to predict the testing data, then we calculate the performance of the system. The various classification models used are: Logistic Regression, Gaussian Naïve Bayes Classifier, KNN (K Nearest Neighbor) algorithm, Support Vector Machines, Gradient Boost Algorithm, Decision Trees and Random Forest Classifier. Among these methods the Random Forest and Decision tree has given good results.

☐

SOFTWARE REQUIREMENT SPECIFICATION (SRS)

A System Requirement Specification (SRS) is basically an organization's understanding of a customer or potential client's system requirements and dependencies at a particular point prior to any actual design or development work. The information gathered during the analysis is translated into a document that defines a set of requirements. It gives the brief description of the services that the system should provide and also the constraints under which, the system should operate. Generally, SRS is a document that completely describes what the proposed software should do without describing how the software will do it. It's a two-way insurance policy that assures that both the client and the organization understand the other's requirements from that perspective at a given point in time. SRS document itself states in precise and explicit language those functions and capabilities a software system (i.e., a software application, an ecommerce website and so on) must provide, as well as states any required constraints by which the system must abide. SRS also functions as a blueprint for completing a project with as little cost growth as possible. SRS is often referred to as the "parent" document because all subsequent project management documents, such as design specifications, statements of work, software architecture specifications, testing and validation plans, and documentation plans, are related to it. Requirement is a condition or capability to which the system must conform. Requirement Management is a systematic approach towards eliciting, organizing and documenting the requirements of the system clearly along with the applicable attributes. The elusive difficulties of requirements are not always obvious and can come from any number of sources.

1. Introduction

1.1 Purpose

The purpose of this document is to outline the requirements for the IPL Victory Predictor application, which predicts the outcome of IPL matches using machine learning.

1.2 Scope

The application will predict the winning probability of an IPL team based on live match data and historical performance. Users can input match details to generate predictions and view historical performance of teams.

2. Overall Description

2.1 Product Functions

User Input: Collect match details from users such as teams, venue, target score, current score, overs completed, and wickets down.

Prediction: Calculate and display the winning probability for the batting team using a pre-trained machine learning model.

☐

Historical Analysis: Display historical performance of the selected teams and the outcomes of the last 5 matches between them.

2.2 User Characteristics

General Users: Individuals interested in predicting IPL match outcomes.

Cricket Analysts: Analysts who evaluate team performances and strategies.

2

3. Functional Requirements

3.1 User Input

The system shall provide dropdowns for selecting the batting and bowling teams.

The system shall provide input fields for entering the match venue, target score, current score, overs completed, and wickets down.

3.2 Data Processing

The system shall calculate runs left, balls left, wickets remaining, current run rate (CRR), and required run rate (RRR) based on user inputs.

3.3 Prediction

The system shall use a pre-trained machine learning model to predict the winning probability for the batting team.

The system shall display the predicted winning probability as a percentage for both teams.

3.4 Historical Performance

The system shall display the historical performance of the selected teams.

The system shall retrieve and display the results of the last 5 matches between the selected teams.

4. Non-Functional Requirements

4.1 Performance

The system shall generate predictions within 5 seconds of receiving user inputs.

4.2 Usability

2

The user interface shall be intuitive and easy to navigate.

☐

The system shall provide clear instructions and feedback to users.

4.3 Reliability

The system shall provide accurate predictions based on the pre-trained model. The system shall be available 99% of the time during IPL matches.

4.4 Security

The system shall process user data securely and ensure data privacy.

4.5 Maintainability

The system shall be easy to update with new data and models.

☐

USE CASE DIAGRAM

A use case diagram is a visual representation in Unified Modeling Language (UML) that depicts a system's functionalities from the perspective of its users (actors). It illustrates the various ways users can interact with a system and the different scenarios or use cases in which the system responds.
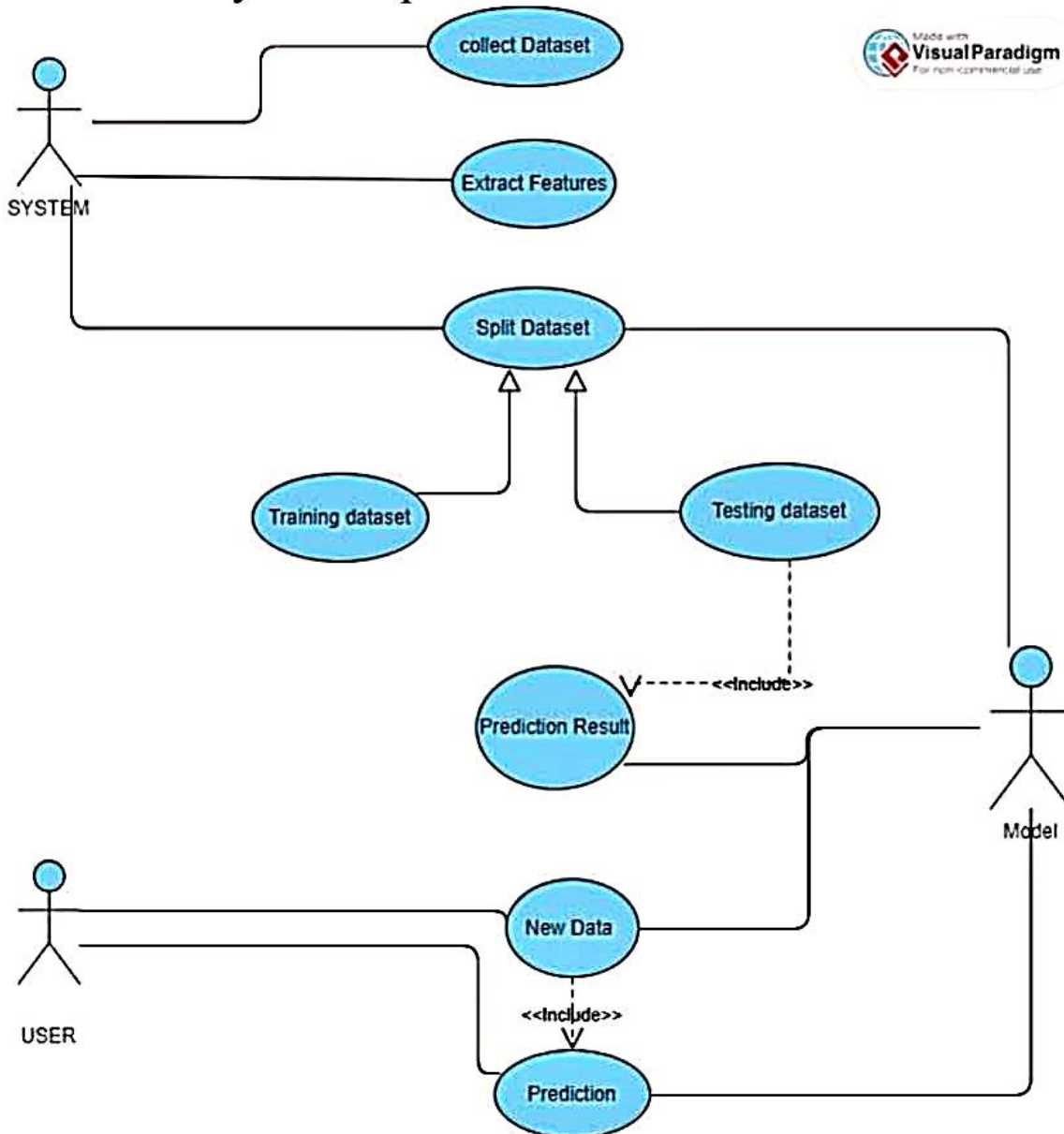

Fig 4.1

SYSTEM FLOW DIAGRAM

A System Flow Diagram (also known as a Context Diagram) provides a high-level overview of how a system interacts with external entities. It illustrates the flow of data between the system and its environment, showcasing major processes or functions. For the sentiment analysis GUI system, we can create a simplified System Flow

This System Flow Diagram offers a simplified overview of the major components and data flow within the sentiment analysis GUI system. It focuses on how information moves between the user and the system's core functionality.
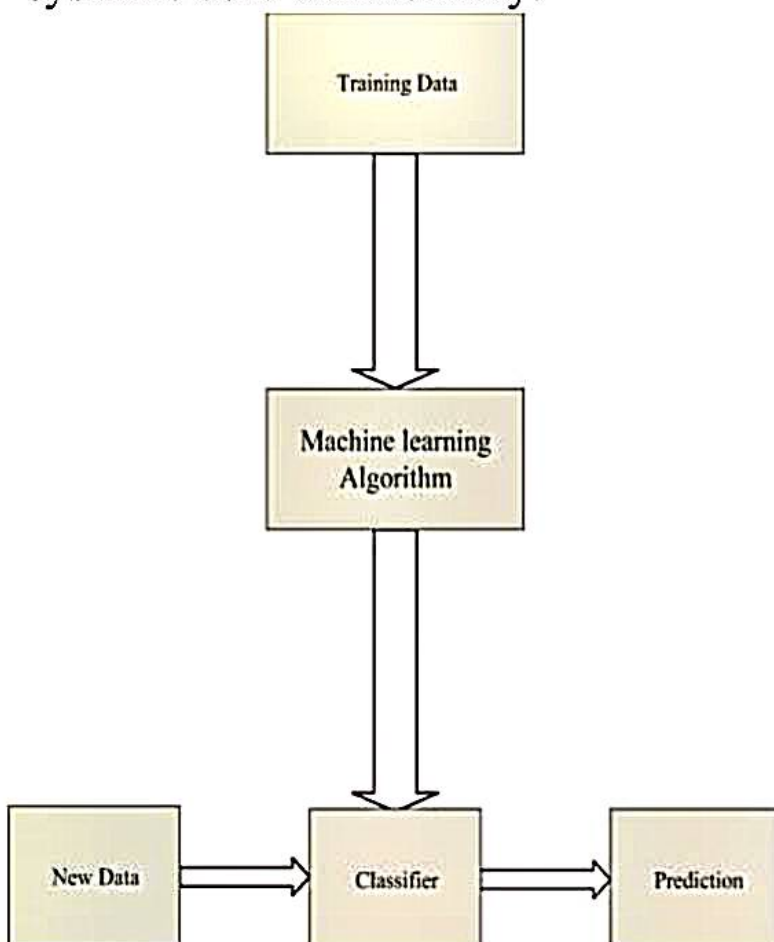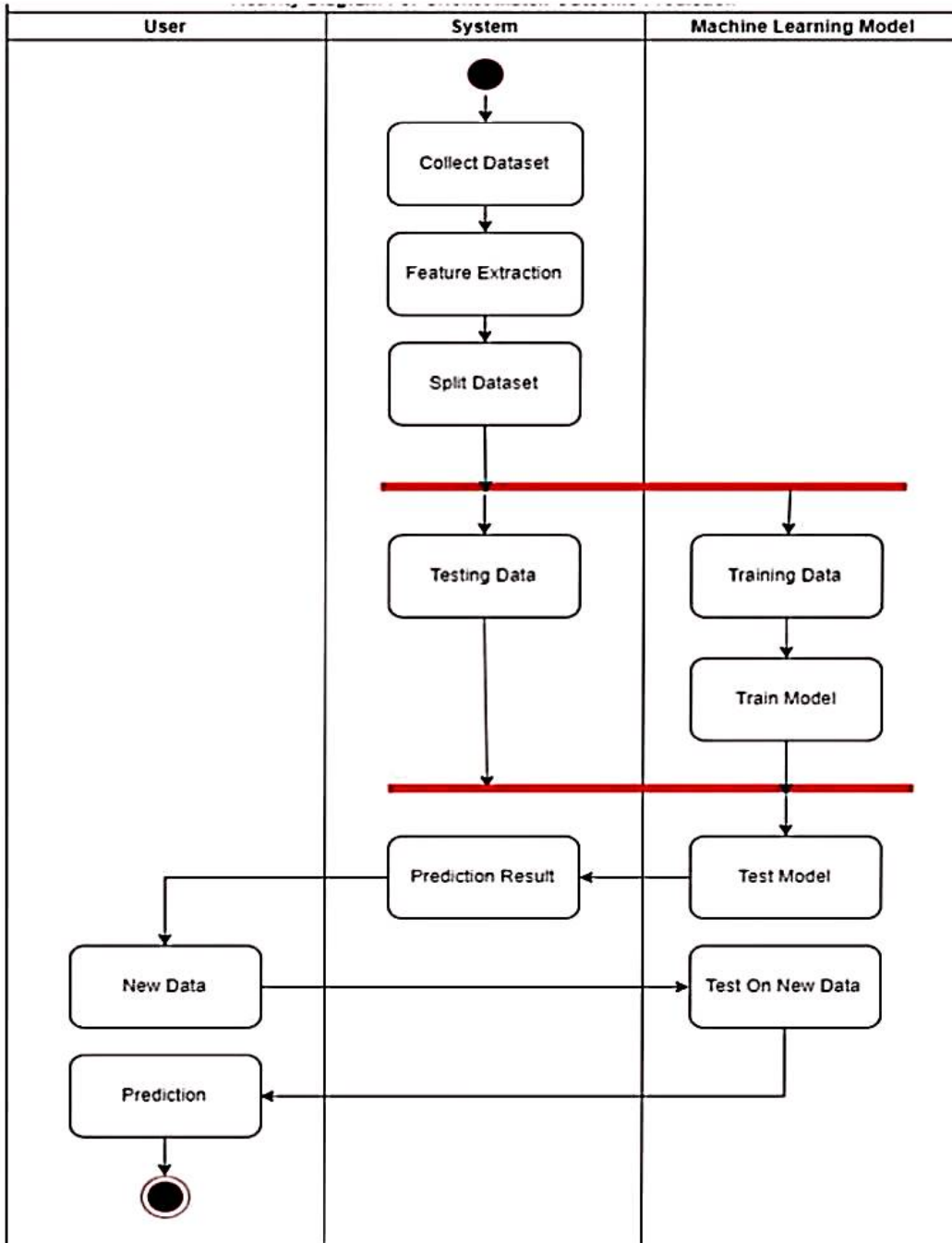

Fig 4.2

ACTIVITY DIAGRAM

User	System	Machine Learning Model

Fig 4.3

FLOW CHART

Flow Chart

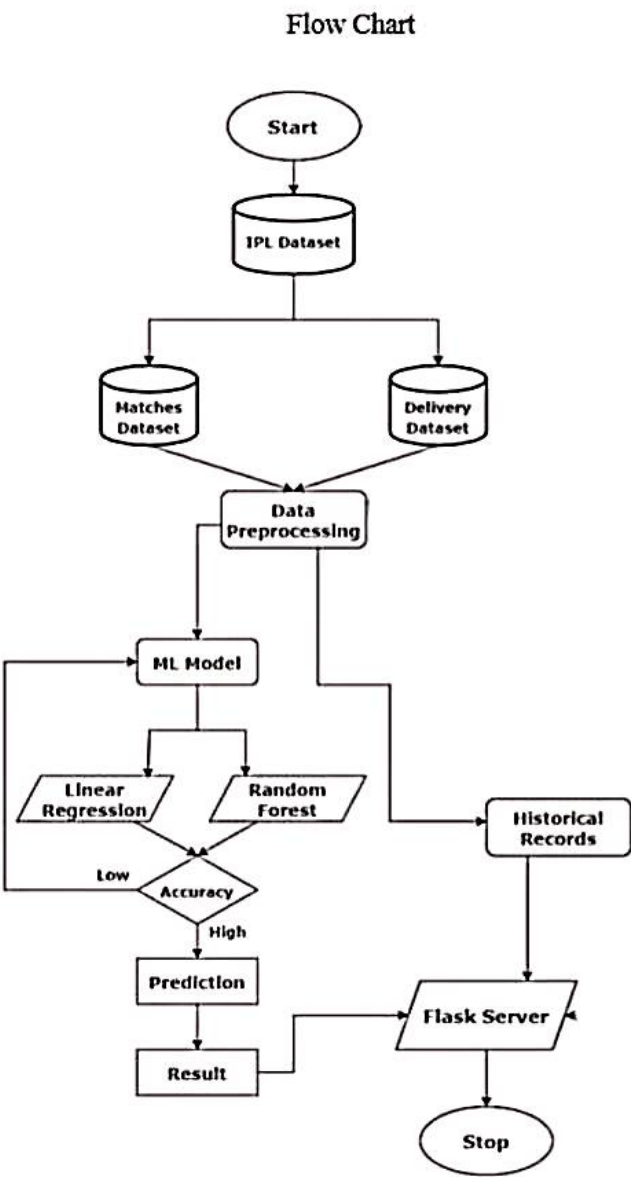Fig 4.4

Data Dictionary

The data dictionary provides a detailed description of each data element used in the IPL Victory Predictor application. This includes the name, type, description, and possible values or ranges for each element. 2

Element Name	Type	Description	Possible Values / Range
batting_team	String	The team currently batting	'Sunrisers Hyderabad', 'Mumbai Indians', 'Kolkata Knight Riders', 'Royal Challengers Bangalore', 'Kings XI Punjab', 'Chennai Super Kings', 'Rajasthan Royals', 'Delhi Capitals'
bowling_team	String	The team currently bowling	'Sunrisers Hyderabad', 'Mumbai Indians', 'Kolkata Knight Riders', 'Royal Challengers Bangalore', 'Kings XI Punjab', 'Chennai Super Kings', 'Rajasthan Royals', 'Delhi Capitals'
city	String	The city where the match is being played	'Bangalore', 'Hyderabad', 'Kolkata', 'Mumbai', 'Visakhapatnam', 'Indore', 'Durban', 'Chandigarh', 'Delhi', 'Dharamsala', 'Ahmedabad', 'Chennai', 'Ranchi', 'Nagpur', 'Mohali', 'Pune', 'Bengaluru', 'Jaipur', 'Port Elizabeth', 'Centurion', 'Raipur', 'Sharjah', 'Cuttack', 'Johannesburg', 'Cape Town', 'East London', 'Abu Dhabi', 'Kimberley', 'Bloemfontein'
target	Integer	The target score set by the first batting team	0 - 500
score	Integer	The current score of the batting team	0 - 500
overs	Float	The number of overs completed	0.0 - 20.0
wickets	Integer	The number of wickets down	0 - 10
runs_left	Integer	The number of runs left to reach	0 - 500

☐

Element Name	Type	Description	Possible Values / Range	
		the target		
balls_left	Integer	The number of balls left to play	0 - 120	
wickets_remaining	Integer	The number of wickets remaining 2	0 - 10	
crr	Float	Current Run Rate (runs scored per over)	0.0 - 36.0	
rrr	Float	Required Run Rate (runs needed per over)	0.0 - 36.0	
winning_probability	Float	The predicted probability of the batting team winning	0.0 - 100.0	
historical_performance	DataFrame	Past performance data of the selected teams	Historical match records including date, venue, winner, and player of the match	

5.
LIMITATIONS

☐

The model relies heavily on historical match data, which may not always be indicative of future performance, especially in a dynamic and evolving sport like cricket.

Changes in team composition, player form, and strategies over time may not be fully captured by past data.

Limited Scope

The model is specifically trained on IPL match data and may not perform as accurately for other leagues, formats, or international matches without additional training and adaptation.

Different playing conditions, rules, and team dynamics in other formats or leagues may affect the model's accuracy.

Technical Limitations

The application is built using Streamlit, which may have limitations in terms of scalability, performance, and user interface customization compared to more robust web development frameworks.

Handling large datasets and complex computations may pose performance challenges, especially with a growing user base.

☐

6. FUTURE APPLICATIONS

Expansion to Other Leagues and Formats:

Extend the predictive model to other cricket leagues such as the Big Bash League (BBL), Caribbean Premier League (CPL), and international T20, ODI, and Test matches.

Adapt the model to accommodate the nuances of different formats and playing conditions.

Real-time In-match Updates:

Integrate real-time data feeds to update predictions dynamically as the match progresses.

Provide ball-by-ball predictions and insights, giving users an up-to-the-minute assessment of the game's state.

Player Performance Prediction:

Develop models to predict individual player performances, such as expected runs scored, wickets taken, or player efficiency in different match situations.

Offer fantasy sports players valuable insights into player form and expected contributions.

Advanced Analytics and Insights:

Implement advanced analytics to identify key moments and turning points in matches.

Use machine learning to detect patterns and strategies employed by teams, providing deeper tactical insights.

Fan Interaction and Engagement:

Create interactive features that allow fans to simulate different match scenarios and outcomes based on hypothetical situations.

Develop social features where fans can share predictions, discuss match scenarios, and compete with friends in prediction leagues.

Broadcast and Media Integration:

Collaborate with sports broadcasters and media outlets to integrate predictive analytics into live broadcasts.

Provide commentators and analysts with data-driven insights to enhance their commentary and pre/post- match analysis

7. BIBLIOGRAPHY

WEBSITE REFERRED
- https://www.youtube.com
- https://www.kaggle.com
- https://geeksforgeeks.org
- https://github.com
- www.google.co.in

BOOKS REFERRED

- Let Us Python by Mark Lutz

- Let us Python Programming by John Zelle

2

8. REFERENCES

T. A. Severini, Analytic methods in sports: Using mathematics and statistics to understand data from baseball, football, basketball, and other sports. Chapman and Hall/CRC, 2014.

R. Rein and D. Memmert, "Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science," SpringerPlus, vol. 5, no. 1, p. 1410, 2016

Veppur Sankaranarayanan, Vignesh and Sattar, Junaed and Lakshmanan," Auto-play: A Data Mining Approach to ODI Cricket

Tim B. SWARTZ, Paramjit S Gill and S. Muthukumarana,"Modelling and simulation for one-day cricket", Canadian Journal of Statistics, 2009, Vol 37, No 2, pp-143-160