Student's Showcase
Heriot-Watt University

Cricket Guru

Data Magic off the Pitch

Name: Aman Manish Valera

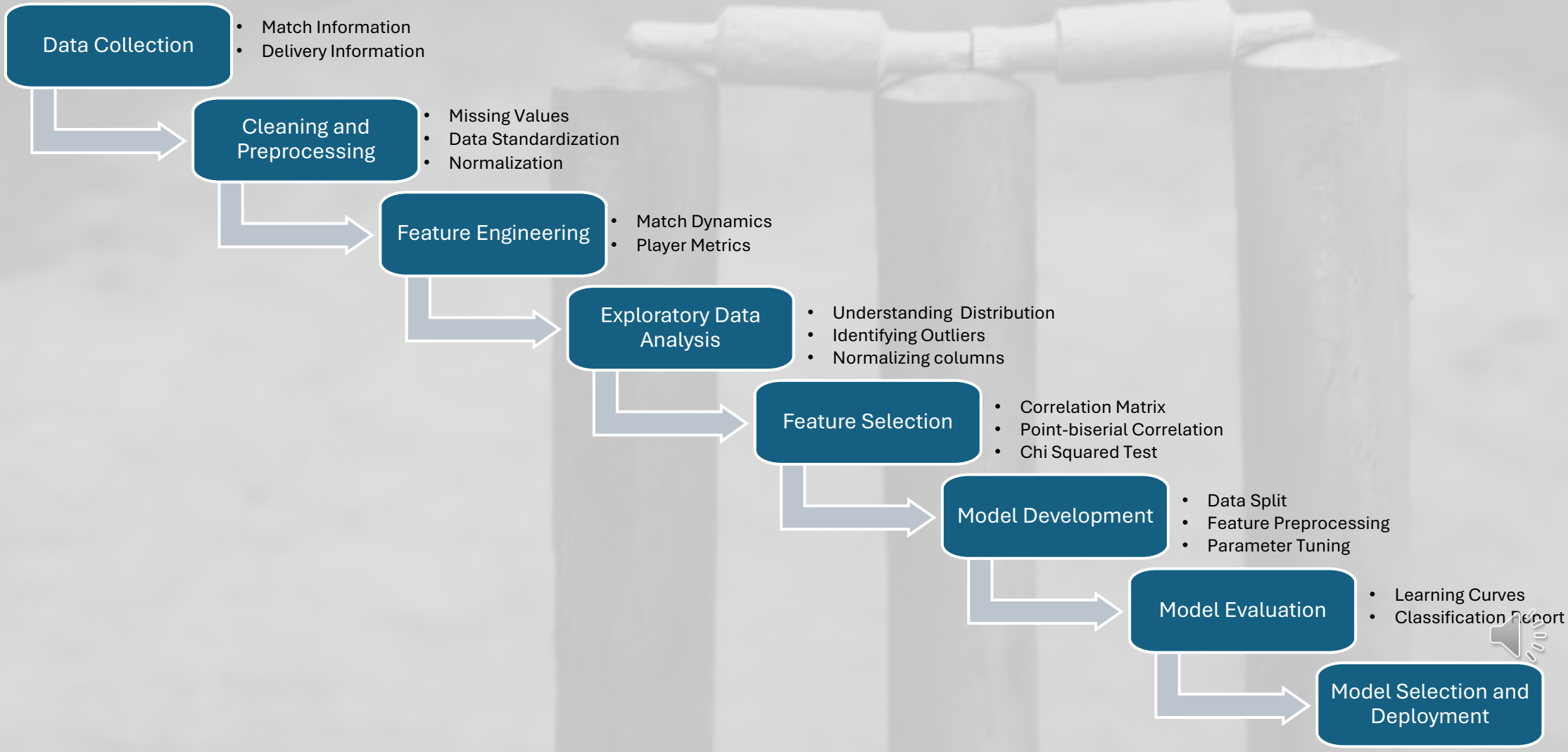
Introduction

- Cricket is often dubbed as a religion in India, and Indian Premier League (IPL) has a massive fan base.
- Providing fans with insights and inspiring more projects in the field of cricket analytics.
- With around 61 million players on fantasy gaming platforms, the IPL 2023 generated revenue of over 335 million USD

Objectives

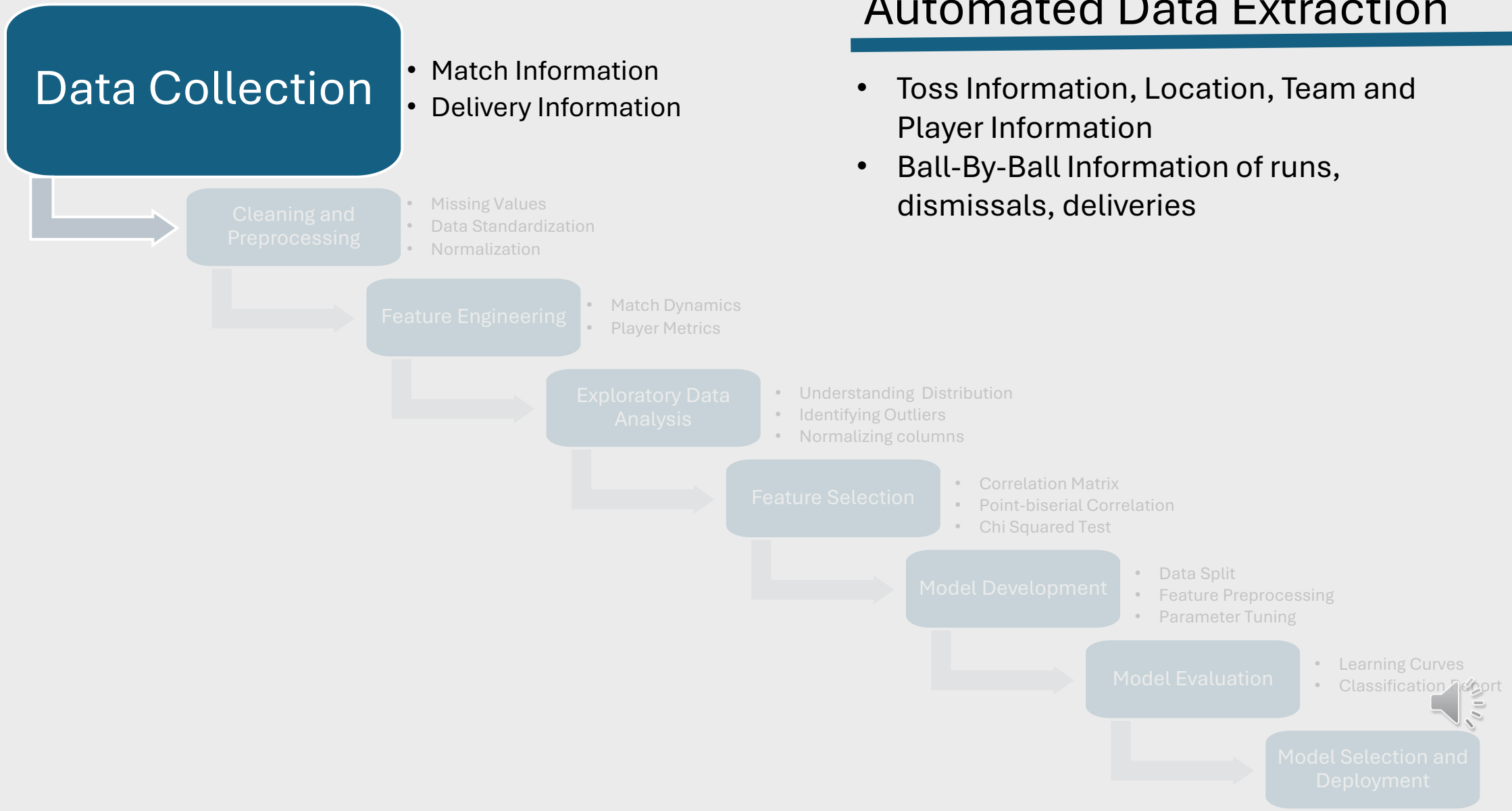
- Predicting winner of an IPL match at a given point of time in real-time using machine learning based on historical data.
- To analyse how much impact various factors have on the match outcomes.

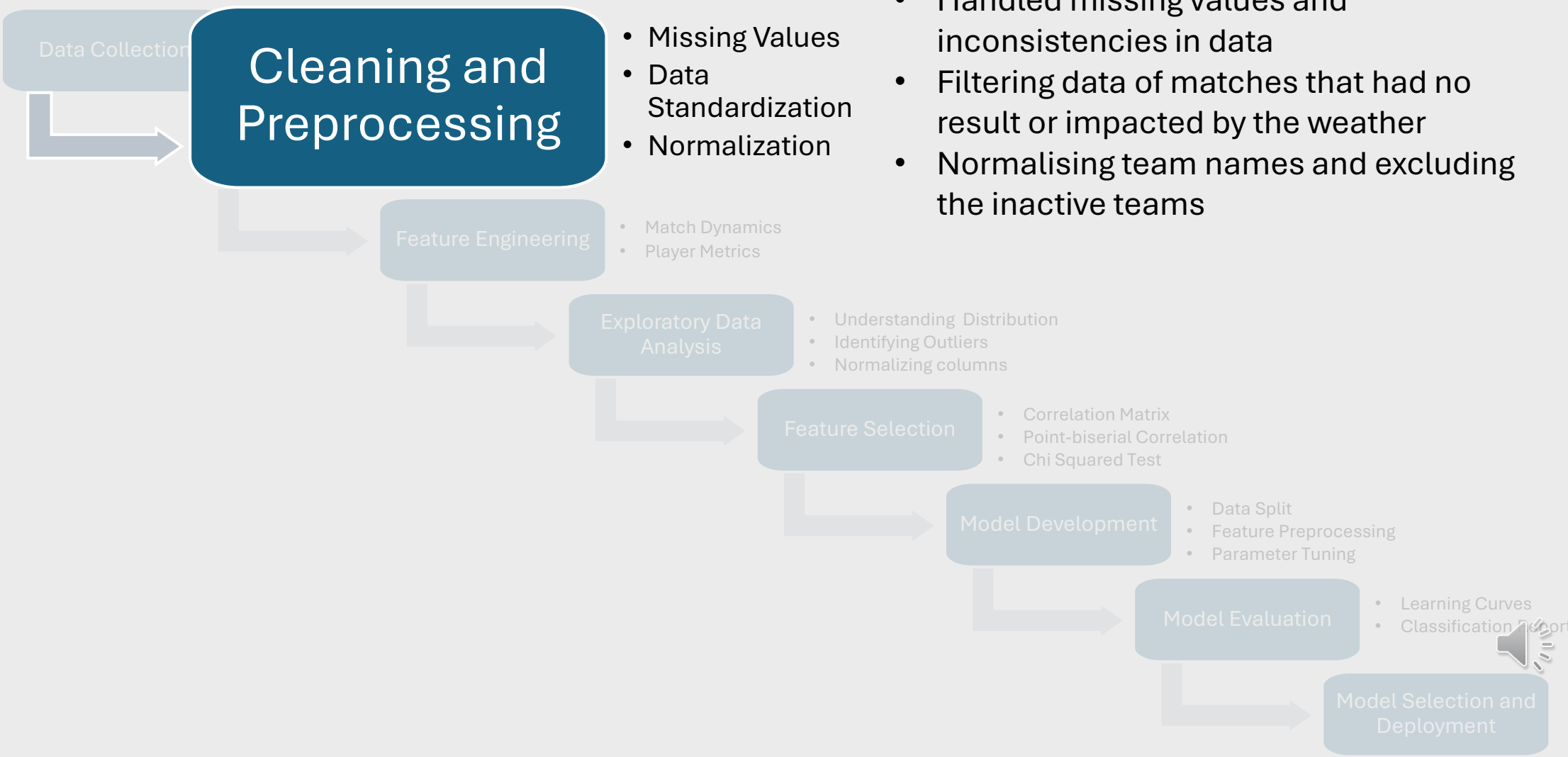
Methodology



Automated Data Extraction

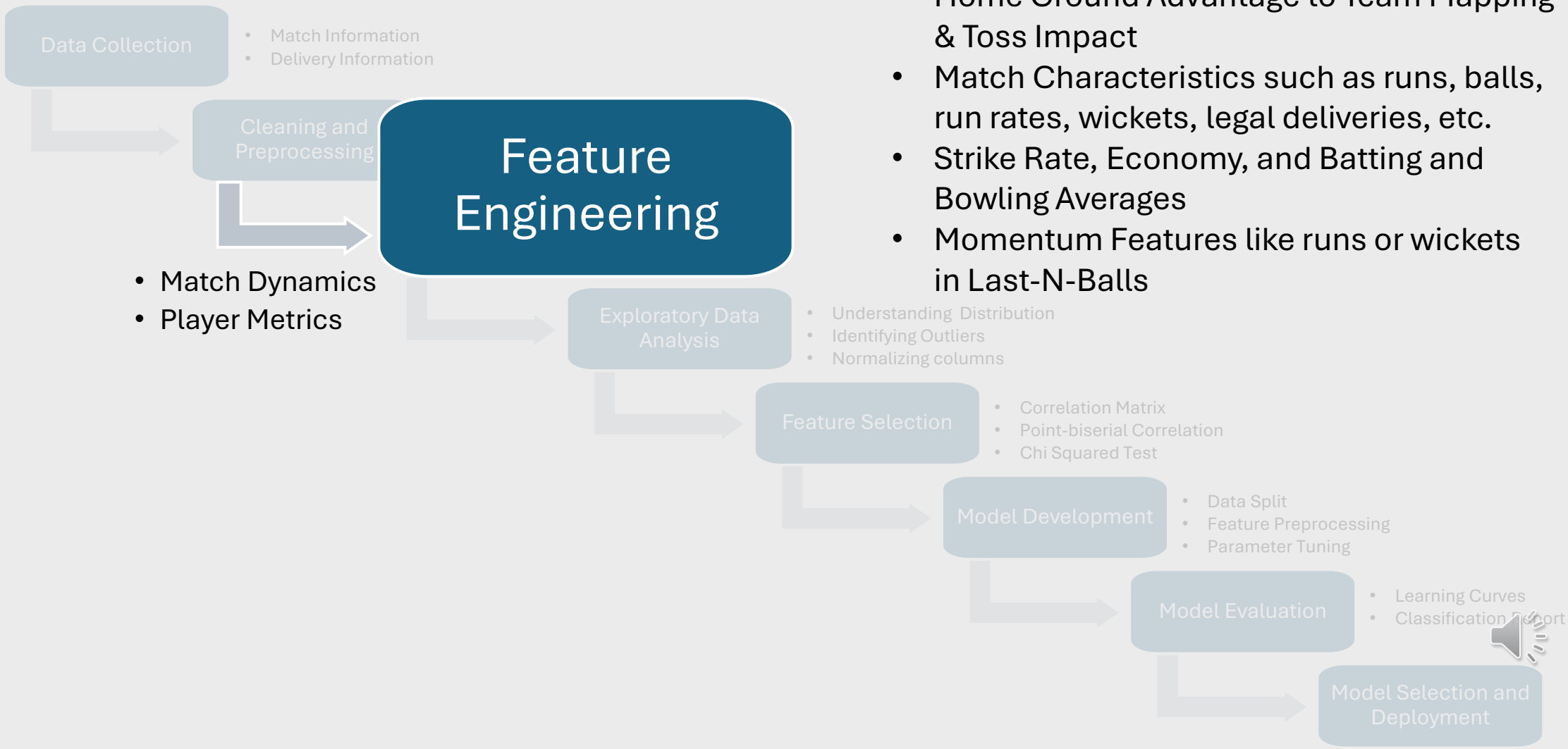
- Toss Information, Location, Team and Player Information
- Ball-By-Ball Information of runs, dismissals, deliveries





Preprocessing

- Handled missing values and inconsistencies in data
- Filtering data of matches that had no result or impacted by the weather
- Normalising team names and excluding the inactive teams



Feature Engineering

- Home Ground Advantage to Team Mapping & Toss Impact
- Match Characteristics such as runs, balls, run rates, wickets, legal deliveries, etc.
- Strike Rate, Economy, and Batting and Bowling Averages
- Momentum Features like runs or wickets in Last-N-Balls

EDA

- Excluded matches with rare Targets
- Normalizing Team, Venue, City and Player Names
- Capped Values for Outliers

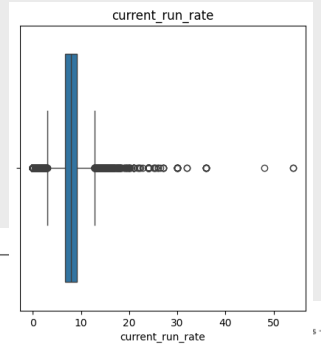
Data Collection
- Match Information
- Delivery Information

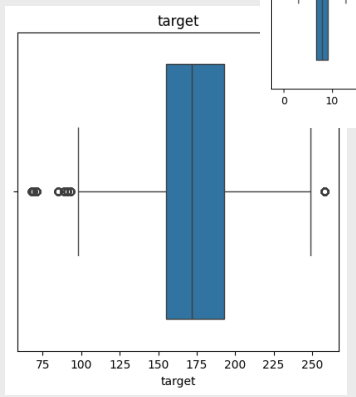
Cleaning and Preprocessing
- Missing Values
- Data Standardization
- Normalization

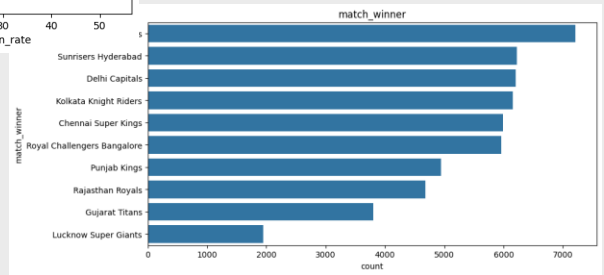
Feature Engineeri

Exploratory Data Analysis

- Understanding Distribution
- Identifying Outliers
- Normalizing columns







Feature Selection
- Correlation Matrix
- Point-biserial Correlation
- Chi Squared Test

Model Development
- Data Split
- Feature Preprocessing
- Parameter Tuning

Model Evaluation
- Learning Curves
- Classification Report

Model Selection and Deployment

Feature Selection

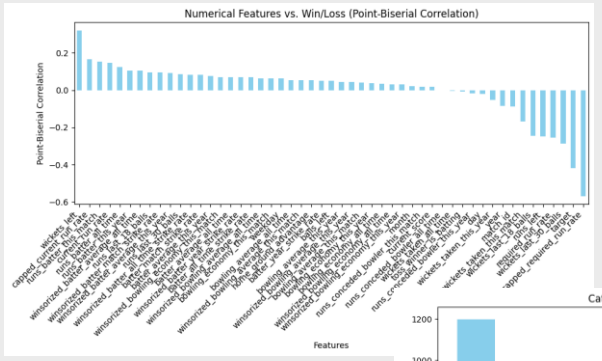
- Match Environment Features such as Venue, Batting Team, Toss Winner
- Match characteristics and Momentum Features
- Batsman's match and all-time, and bowler's match statistics

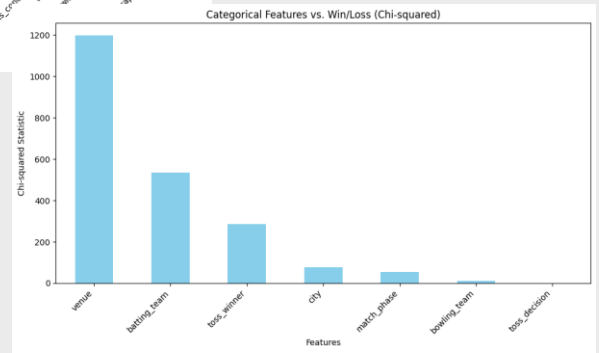
Data Collection
- Match Information
- Delivery Information

Cleaning and Preprocessing
- Missing Values
- Data Standardization
- Normalization

Feature Engineering
- Match Dynamics
- Player Metrics

Exploratory Data Analysis





Feature Selection

- Correlation Matrix
- Point-biserial Correlation
- Chi Squared Test

Model Development
- Data Split
- Feature Preprocessing
- Parameter Tuning

Model Evaluation
- Learning Curves
- Classification Report

Model Selection and Deployment

Data Modelling

- Used ‘TimeSeriesSplit’ for splitting the dataset for training and testing
- Feature Encoding and Scaling
- Model Pipeline and parameter grid construction

Data Collection
- Match Information
- Delivery Information

Cleaning and Preprocessing
- Missing Values
- Data Standardization
- Normalization

Feature Engineering
- Match Dynamics
- Player Metrics

Exploratory Data Analysis
- Understanding Distribution
- Identifying Outliers
- Normalizing columns

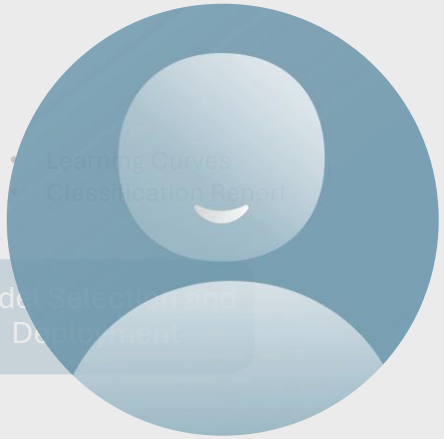
Feature Selectio

Model Development

- Data Split
- Feature Preprocessing
- Parameter Tuning

Model Evaluation
- Learning Curves
- Classification Report

Model Selection and Deployment



Evaluation Techniques

- Plotted Learning Curves to understand how well models learn with the amount of data and spot overfitting
- Classification Report to have thorough evaluation of all accuracies

Data Collection
- Match Information
- Delivery Information

Cleaning and Preprocessing
- Missing Values
- Data Standardization
- Normalization

Feature Engineering
- Match Dynamics
- Player Metrics

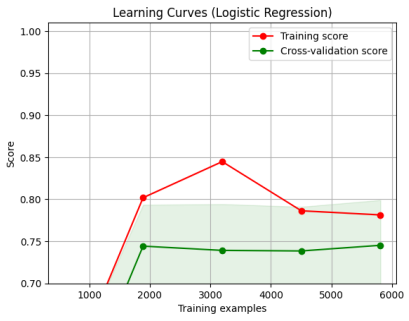
Exploratory Data Analysis
- Understanding Distribution
- Identifying Outliers
- Normalizing columns

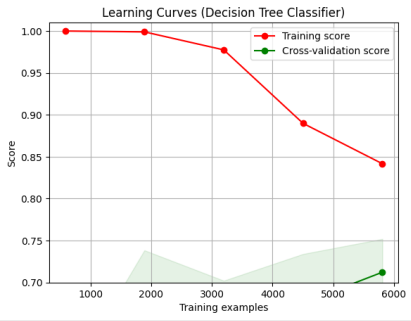
Feature Selection
- Correlation Matrix
- Point-biserial Correlation
- Chi Squared Test

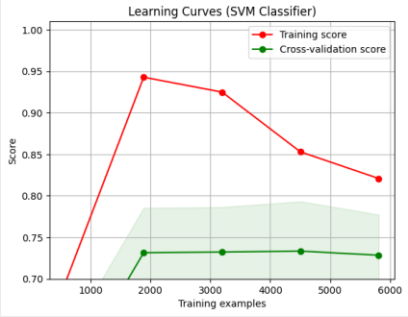
Model Development

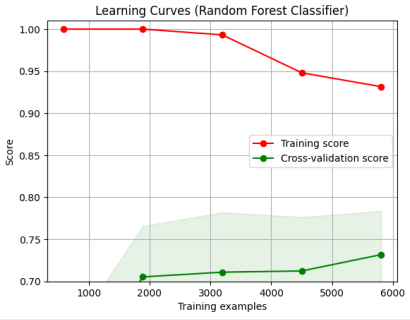
Model Evaluation
- Learning Curves
- Classification Report

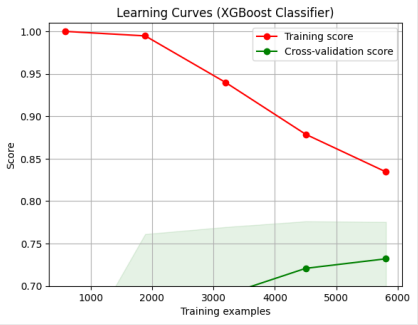
Model Selection and Deployment
- Match Progression Evaluation
- Streamlit for web-application











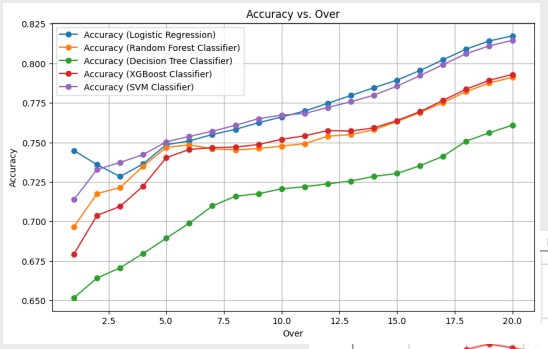
Model Selection & Deployment

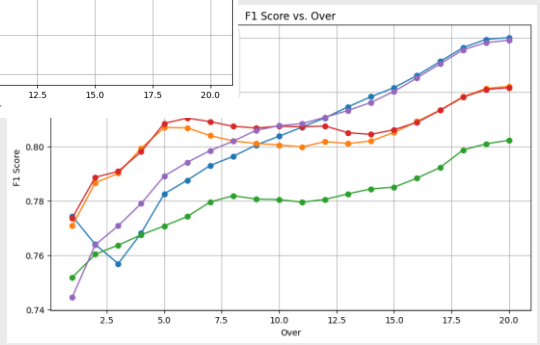
- Microscopic view to accuracy and f1 scores during the match
- Streamlined the Model using its pickle to create a user-friendly interface for real-time inputs for predictions

Data Collection
- Match Information
- Delivery Information

Cleaning and Preprocessing
- Missing Values
- Data Standardization
- Normalization

Feature Engineering
- Match Dynamics
- Player Metrics





Exploratory Data Analysis
- Understanding Distribution
- Identifying Outliers
- Normalizing columns

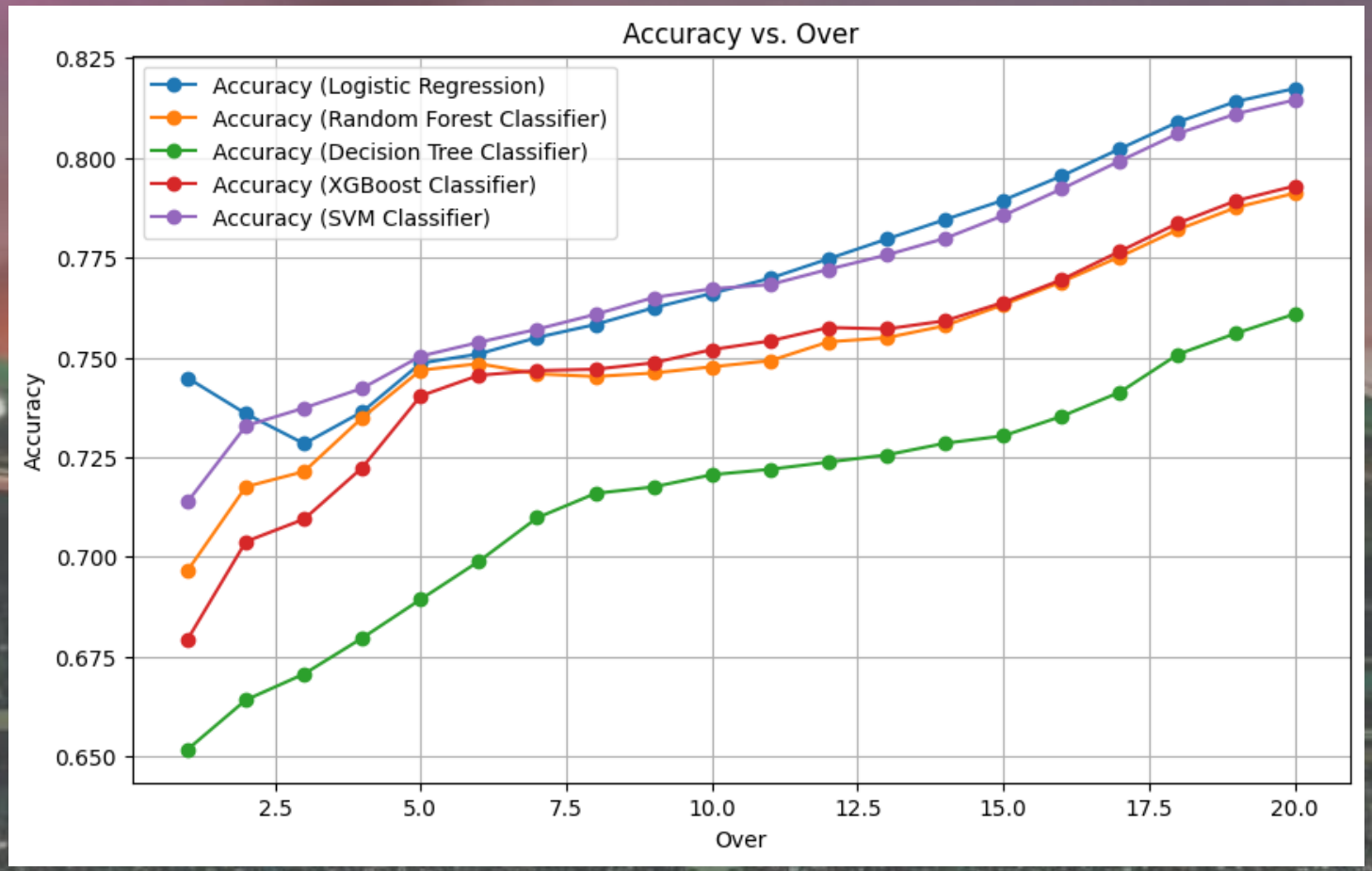
Feature Selection
- Correlation Matrix
- Point-biserial Correlation
- Chi Squared Test

Model Development
- Data Split
- Feature Preprocessing
- Parameter Tuning

Model Evaluation
- Match Progression Evaluation
- Streamlit for web-application

Model Selection and Deployment

Results



- We can see the importance of factors such as venue, toss, etc.
- Provided cricket fans with insights
- Successfully demonstrated how machine learning can be used to predict match outcomes

Future Scope

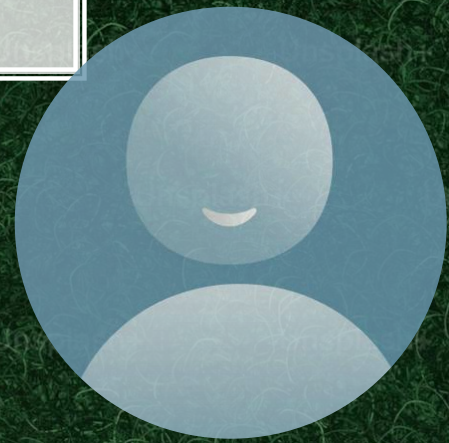
- Web-scrapping or use of APIs to collect live information and integrating other datasets
- Analysing over various types of dataset, such as image, player performance, etc.
- Predicting various different aspects such as injuries, score, player performance, etc.

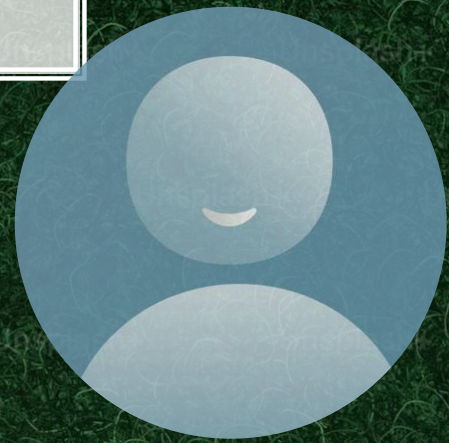
Moments

Special Thanks To

Dr. Chengjia Wang for supervising this project

Q&A?



Q&A?

Let us Check The App