Introduction

Winning and Score Predictor (WASP) is a calculation tool used in cricket to predict the scores and possible results of a limited over match format. For the team batting first, it gives the prediction of the final total. For the team batting second, it predicts the probability of the chasing team winning. Cricket is a game of unpredictability, especially in the context of the T20 league like IPL(Indian Premier League) where every run and every wicket can significantly impact the result of the match. In this project we have combined different machine learning algorithms to predict the score likely to be achieved by the team batting first and to find the winning probability for the team batting second.

Background: Cricket is famous for its unpredictability and becomes even more interesting in leagues such as the Indian Premier League (IPL), where matches unfold with dynamic and unforeseeable outcomes. The T20 format is known for its brevity and emphasis on aggressive gameplay, amplifying the uncertainty and making it challenging to predict results accurately. Every run scored and each wicket taken possesses the potential to reshape the course of the game. In this scenario, it becomes clear that we require tools capable of harnessing the power of predictive modeling. These tools aim to unravel the complexities and improve our comprehension of match dynamics within the constantly evolving world of cricket.

Significance: The introduction of machine learning algorithms in sports analytics, especially in the context of cricket, holds a huge significance. The Winning and Score Predictor (WASP) emerges as a valuable computational tool designed explicitly for cricket.It addresses the overview of limited-over match formats by providing predictions for both teams involved. For the team that bats first, it predicts how many runs they might score, giving an idea of their likely total. Meanwhile, for the team batting second, WASP calculates the chance of them successfully chasing the target, helping us understand how the game might unfold.

In T20 leagues like IPL where games are often close and the competition is tough, using machine learning algorithms helps us make predictions using match information. By employing various machine learning algorithms the project aims to predict the final score for the team batting first and determine the winning probability for the team batting second. Using these prediction models not only helps us understand the details of the game better but also assists fans, experts, and analysts in making smarter decisions, making the whole experience of watching and analyzing cricket matches even better.

Project Overview:

The scope of this project is carefully defined to focus on the predictive modeling aspects related to cricket outcomes, specifically in the context of limited-over match formats, with a primary emphasis on the Indian Premier League (IPL). The project aims to develop and implement a Winning and Score Predictor (WASP) as a computational tool to forecast scores and winning probabilities.

Target Audience: The Winning and Score Predictor (WASP) tool is designed for a diverse range of stakeholders within the cricket ecosystem. The primary beneficiaries include:

- **Teams and Coaches:** Cricket teams and their coaching staff can leverage WASP to make informed strategic decisions during matches. Predicted scores and winning probabilities provide valuable insights into game dynamics, assisting teams in formulating effective strategies for both batting and bowling.
- **Sports Analysts and Commentators:** Sports analysts and commentators can use WASP to enhance their pre-match and in-game analyses. The tool's predictions give us numbers to talk about, helping us understand how the game might go and adding more information for better commentary.
- **Bettors and Fantasy Team Platforms:** WASP can be a helpful tool for bettors and fantasy team platforms. It uses data-driven predictions to guide decision-making. The tool's forecasts can help in placing smarter bets and picking fantasy cricket teams by considering expected player performances.
- **Cricket Enthusiasts and Viewers:** Cricket fans and viewers get a more exciting and enriched experience with WASP. Predicted scores and winning probabilities add an analytical layer to the excitement of the game, allowing fans to follow the match with a deeper understanding of potential outcomes.

Project Structure

1) **Data Collection:**
   The Wasp model relies on two datasets, which contain ball-by-ball data and match results for each game from the 2008 season to the 2019 season.
   - **Ball by ball data:**
     This dataset provides a detailed account of each ball bowled in every IPL match. It includes information such as runs scored, wickets taken, type of delivery, and additional contextual details
   - **Match overview data:**
     This dataset provides a comprehensive information about the match, focusing on details like team names, venue, date, and the final outcome (win/loss).
2) **Preprocessing:**
   To prepare the data for prediction, we divide the data into three sections: the match situation after 5th over, 10th over and 15th over. This includes handling missing values, encoding categorical variables, scaling numerical features, and transforming the data into a format suitable for machine learning algorithms.
   Ball by Ball dataset
   Match overview dataset
3) **Feature Engineering:**
   We extracted the useful features from both the datasets like, player form, team performances, current match situations, match venues and results. These features are further used in different machine learning algorithms to predict the outcome.

4) **Score Prediction(First Innings):**
   The first element of the WASP model focuses on predicting the score for the first innings. Using different machine learning algorithms, the model uses different features to predict the expected total runs a team is likely to score in their initial innings.
5) **Winning Probability Prediction(Second Innings):**
   The second element of the WASP model focuses on predicting the probability for the second innings. Using various machine learning algorithms, the model utilizes different features to predict the likelihood of a team winning the match at that point in time.

System Framework

In the context of a cricket match, we have extracted different features for both elements of the WASP model i.e. for the first innings score prediction and for the second innings winning probability. Below we showcase the key features.

First Innings(score prediction):

X:

1. **Match_id:** it as an identifier in model features to ensure that the model distinguishes between different matches
2. **City:** Venue of the match
3. **Bowling team:** Name of team bowling first
4. **Batting team:** Name of team batting first
5. **Batsman:** denotes the name of the batsman who is currently at crease
6. **Non-Striker:** denotes the name of the batsman who is at the non-striker end
7. **Bowler:** denotes the name of the bowler who is currently bowling at that point in time
8. **Batsman_runs:** denotes the run of batsman who is at the crease
9. **Bowler_wicket:** denotes the current wickets tally of the bowler who is bowling
10. **Current_score:** current score of batting team
11. **Balls_bowled:** number of balls bowled at that point in time
12. **Wickets:** Number of wicket in hand of batting team

Y: Target: Final score of the team batting first

Second Innings(winning probability):

X:

1. **Match_id:** it as an identifier in model features to ensure that the model distinguishes between different matches
2. **City:** Venue of the match
3. **Bowling team:** Name of team bowling first
4. **Batting team:** Name of team batting first
5. **Batsman:** denotes the name of the batsman who is currently at crease
6. **Non-Striker:** denotes the name of the batsman who is at the non-striker end

7. **Bowler:** denotes the name of the bowler who is currently bowling at that point in time
8. **Batsman_runs:** denotes the run of batsman who is at the crease
9. **Bowler_wicket:** denotes the current wickets tally of the bowler who is bowling
10. **Current_score:** current score of batting team
11. **Balls_left:** number of balls left at that point in time
12. **Wickets:** Number of wicket in hand of batting team
13. **Crr**: current run rate of batting team
14. **Rrr:** required run rate for the batting team
15. **Target:** Target to achieve by the batting team

Y: Result: result of the match

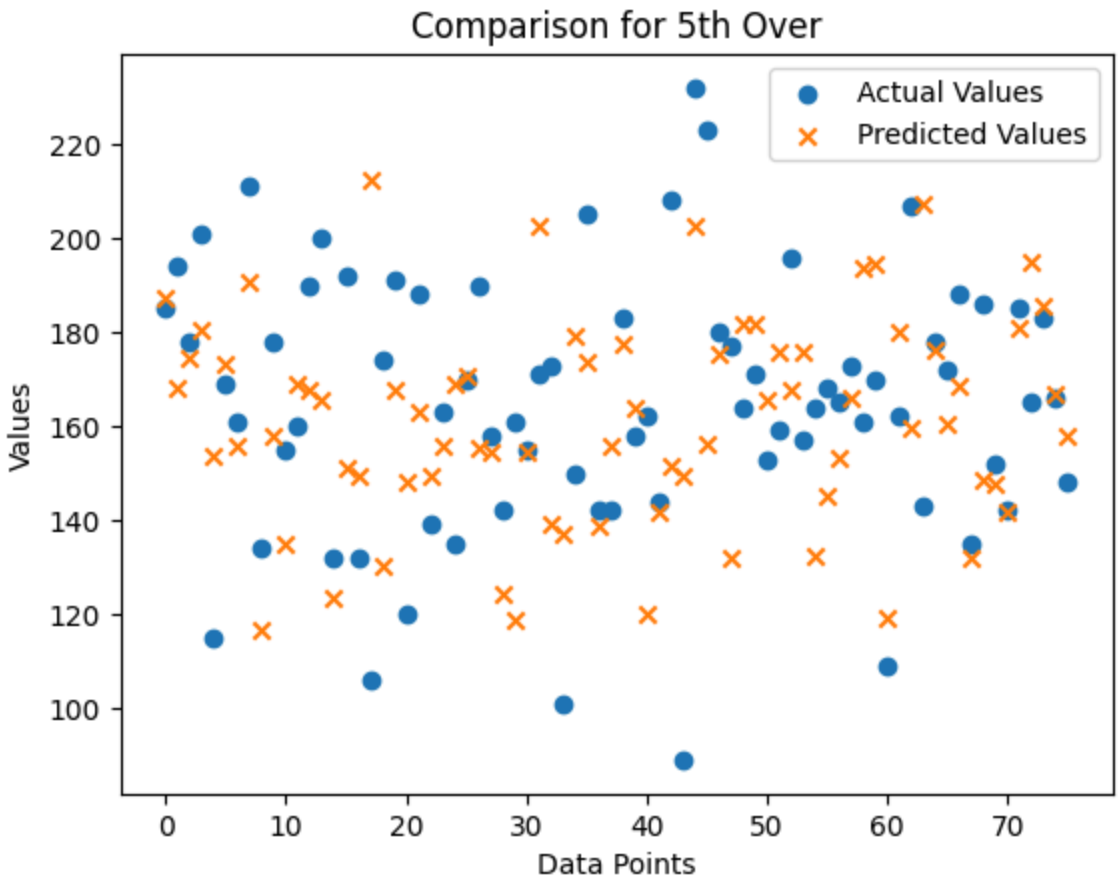
Implementation

We have trained the WASP model in three sections: the match situation after 5th over, 10th over and 15th over for both the innings. We have used different machine learning algorithms to predict the result and then we have ensemble all the algorithms for all the three sections of the match.

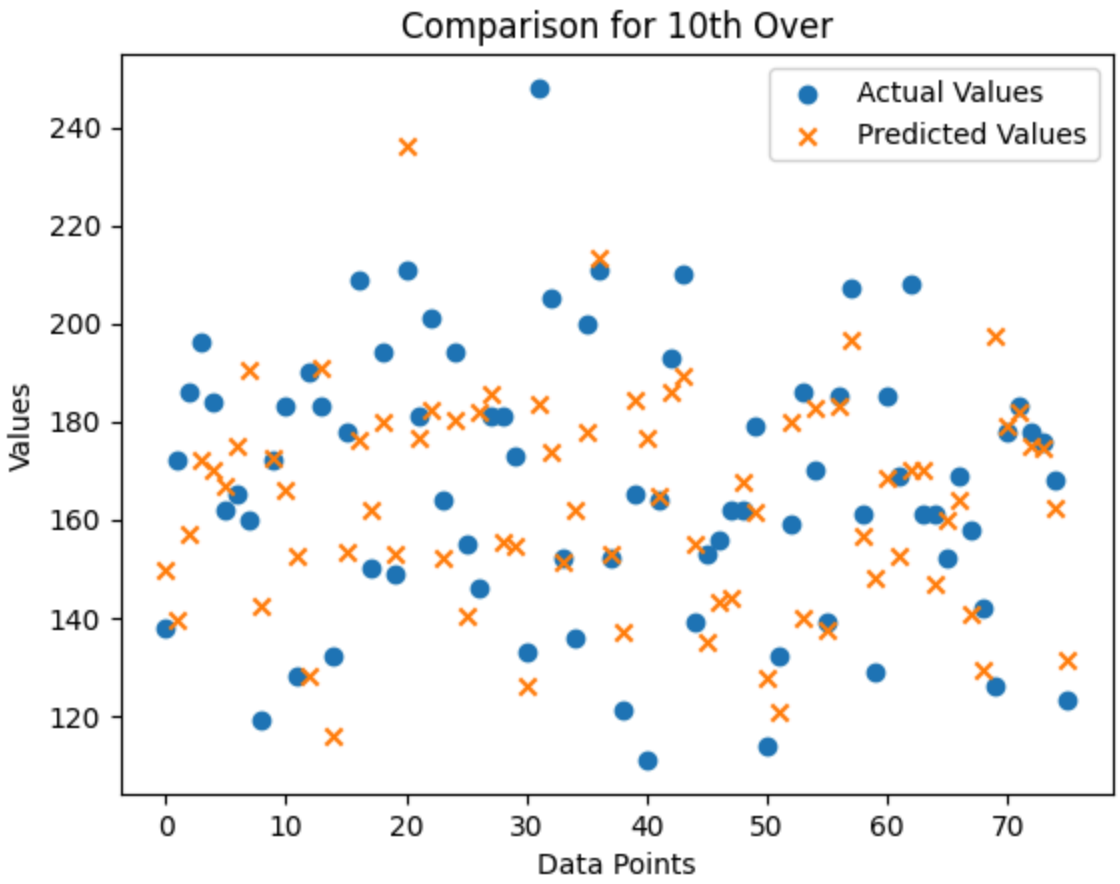
- **First Innings(Score Prediction):**
  To predict the final score for the team batting first, we employed a regression model. In addition to regression, we also applied decision trees, random forests, and XGBoost separately for each of the three sections in the first innings for each match from the IPL 2008 season to the IPL 2019 season. After applying these algorithms, we combined these algorithms using ensemble techniques to enhance the overall accuracy of the model.
- **Second Innings(Winning Probability):**
  To predict the winning probability of the team batting second we used a Classification model. In addition to regression, we also applied decision trees, random forests, and XGBoost separately for each of the three sections in the first innings for each match from the IPL 2008 season to the IPL 2019 season. After applying these algorithms, we combined these algorithms using ensemble techniques to enhance the overall accuracy of the model.

Results

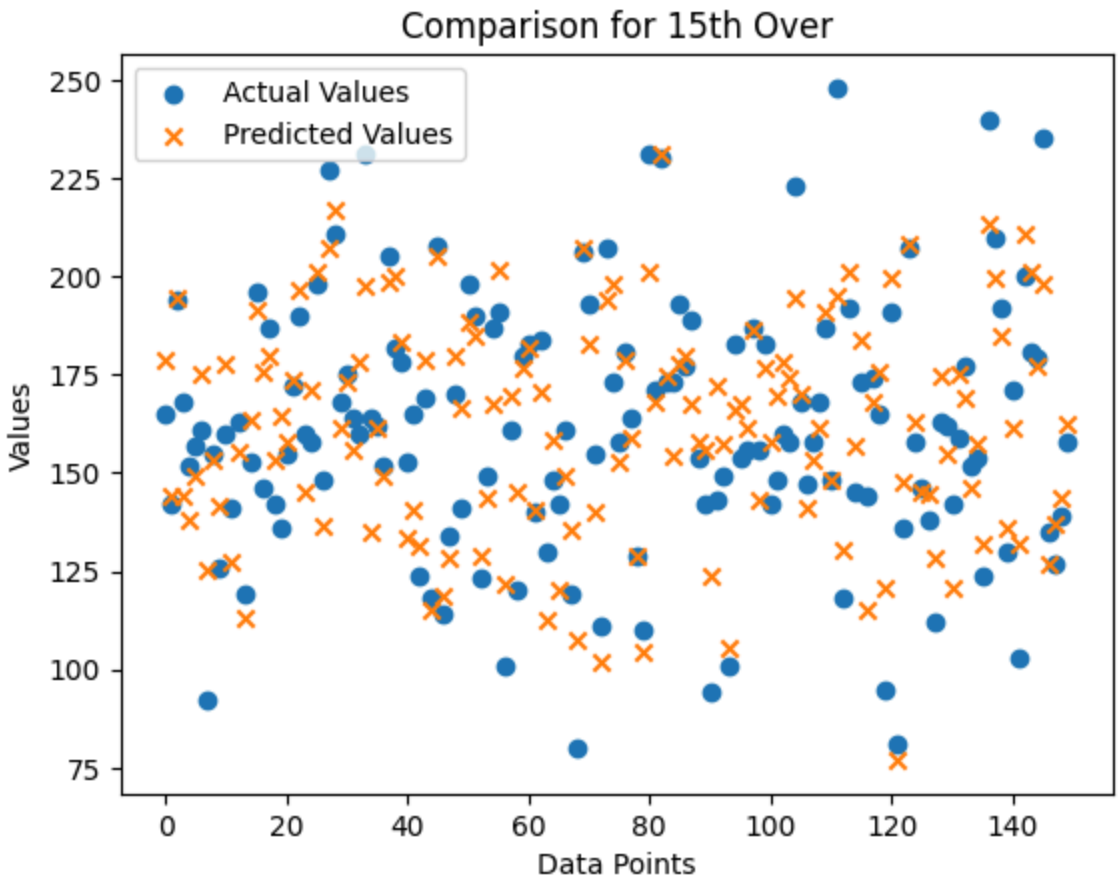
- **First Innings(Score Prediction):**
  R^2 value for 5th over: -12.242594706358778 %



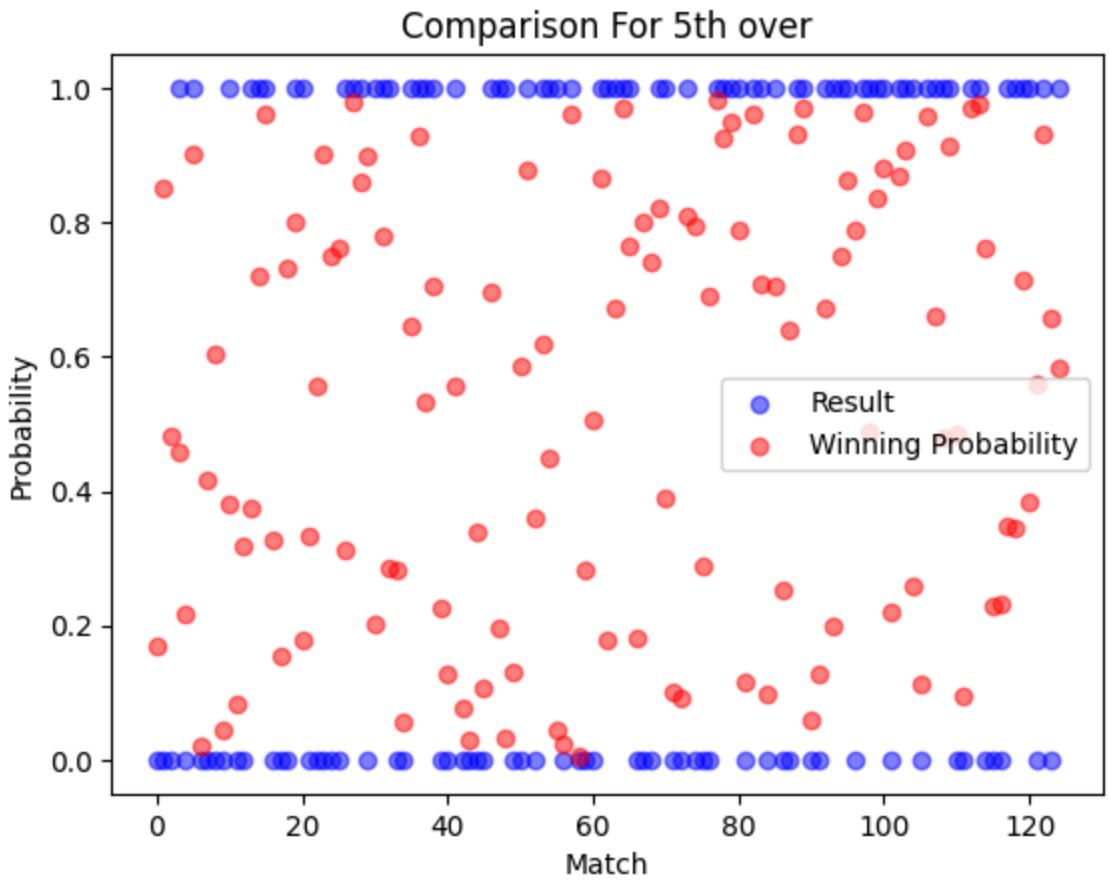
R^2 value for 10th over: 27.67288293556417 %



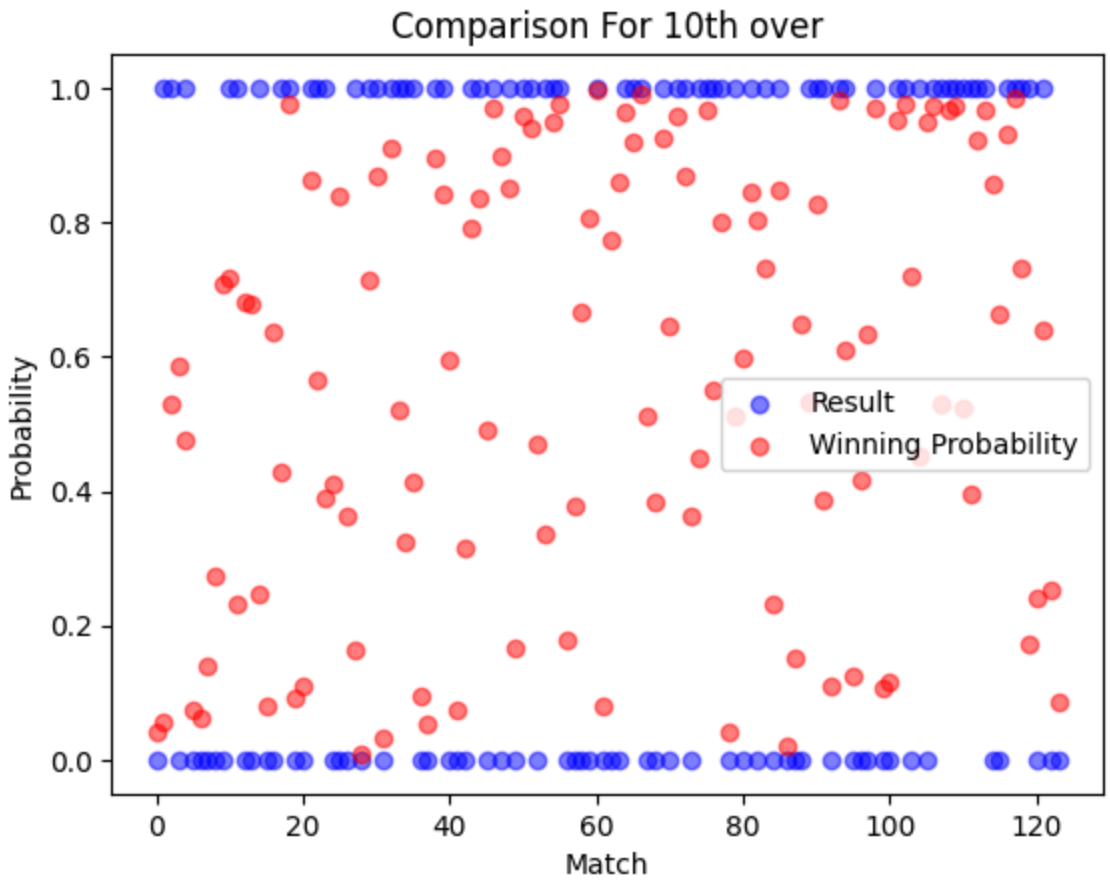
R^2 value for 15th over: 78.50548125012746 %



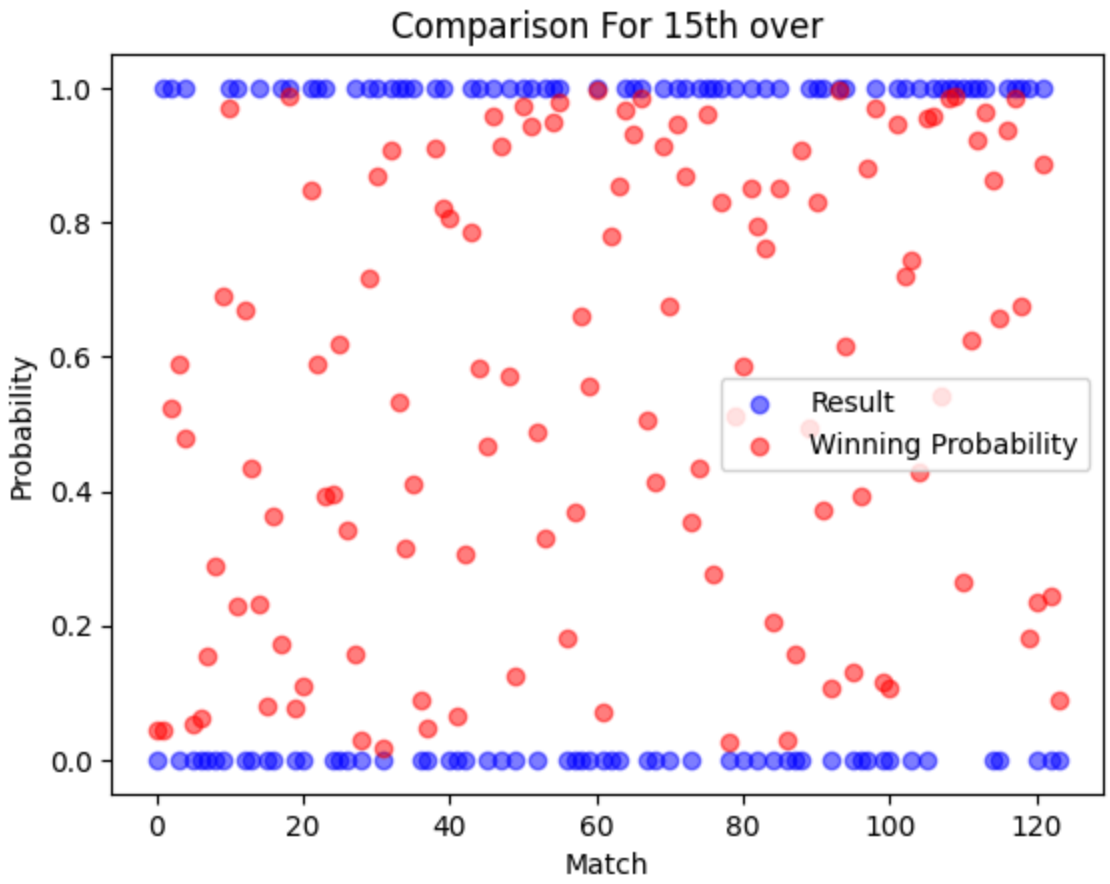
- **Second Innings(Winning Probability):**
  Accuracy Score for 5th over: 68.80 %



Accuracy Score for 10th over: 70.19354838709677 %



Accuracy Score for 15th over: 71.7741935483871 %



Accuracy Score for 15th over: 71.7741935483871 %

Case Study:

Let's consider an example of a match between Sunrisers Hyderabad and Royal Challengers Bangalore, which was played in the IPL 2017 season. We will see how our model predicts the score for the first innings and the winning probability for the team batting second. In this match, Sunrisers Hyderabad is playing in their home stadium.

Royal Challengers Bangalore won the toss, and they decided to field first.

1st Innings:

Batting Team: Sunrisers Hyderabad
Bowling Team: Royal Challengers Bangalore

By the end of 5th Over:

Batsman	Non-Strike	Bowler	Batsm	Bowler_	Current_	Overs	Wlckets

	r		an_run	Wickets	score		
MC Henriques	S Dhawan	S Aravind	17	0	42	5(30 balls)	1

Predicted Score by WASP: 154

By the end of 10th Over:

Batsman	Non-Striker	Bowler	Batsman_run	Bowler_Wickets	Current_score	Overs	WIckets
S Dhawan	MC Henriques	S Aravind	36	0	88	10(60 balls)	1

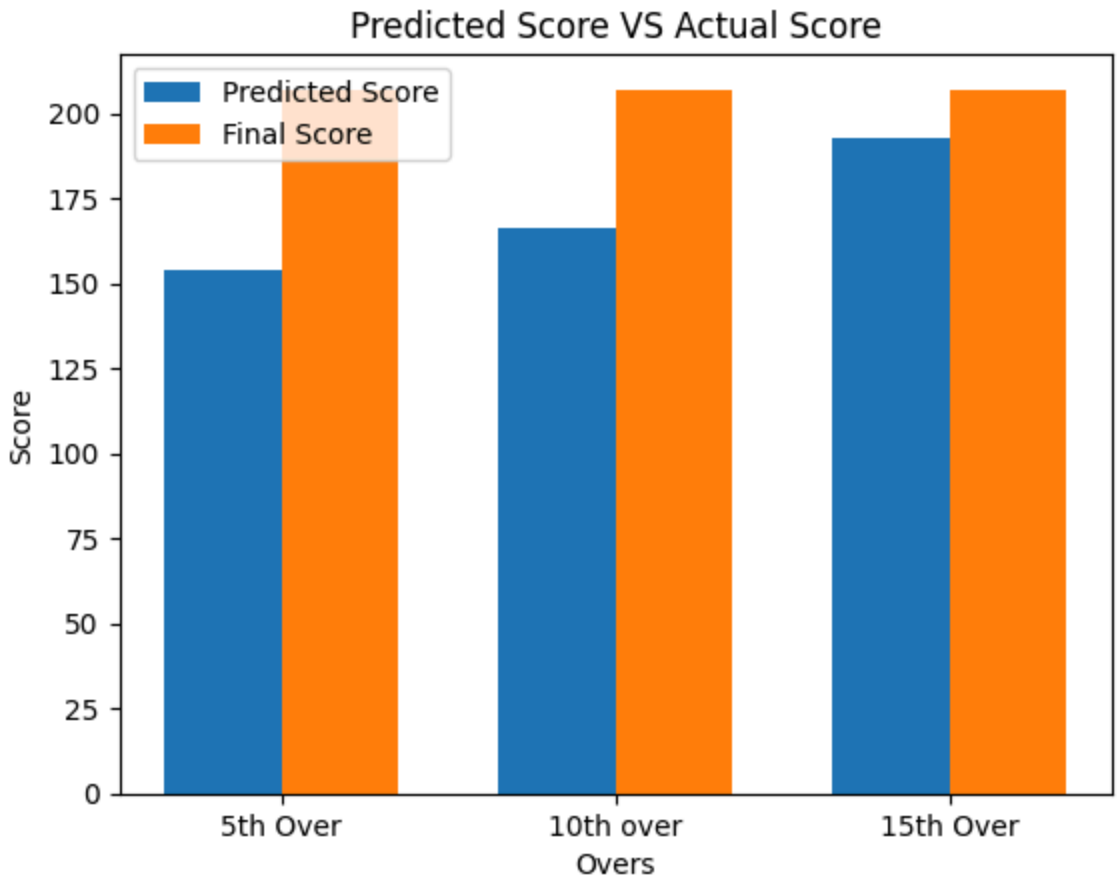
Predicted Score by WASP: 166

By the end of 15th Over:

Batsman	Non-Striker	Bowler	Batsman_run	Bowler_Wickets	Current_score	Overs	WIckets
MC Henriques	Yuvraj Singh	S Aravind	52	0	151	15(90 balls)	2

Predicted Score by WASP: 193

FINAL SCORE: 207



2nd Innings:

Batting Team: Royal Challengers Bangalore
Bowling Team: Sunrisers Hyderabad

By the end of 5th Over:

Batsman	Bowler	Batsman_Runs	bowler_wicket	current_score	balls_left	crr	rrr	wickets	target
C Gayle	B Cutting	22	0	44	90	8.8	10.8666	0	207

Winning Prediction by WASP:
Royal Challengers Bangalore: 17.5%
Sunrisers Hyderabad: 82.5%

By the end of 10th Over:

Batsman	Bowler	Batsman_Runs	bowler_wicket	current_score	balls_left	crr	rrr	wickets	target
T Head	MC Henriques	19	0	94	60	9.4	11.3	2	207

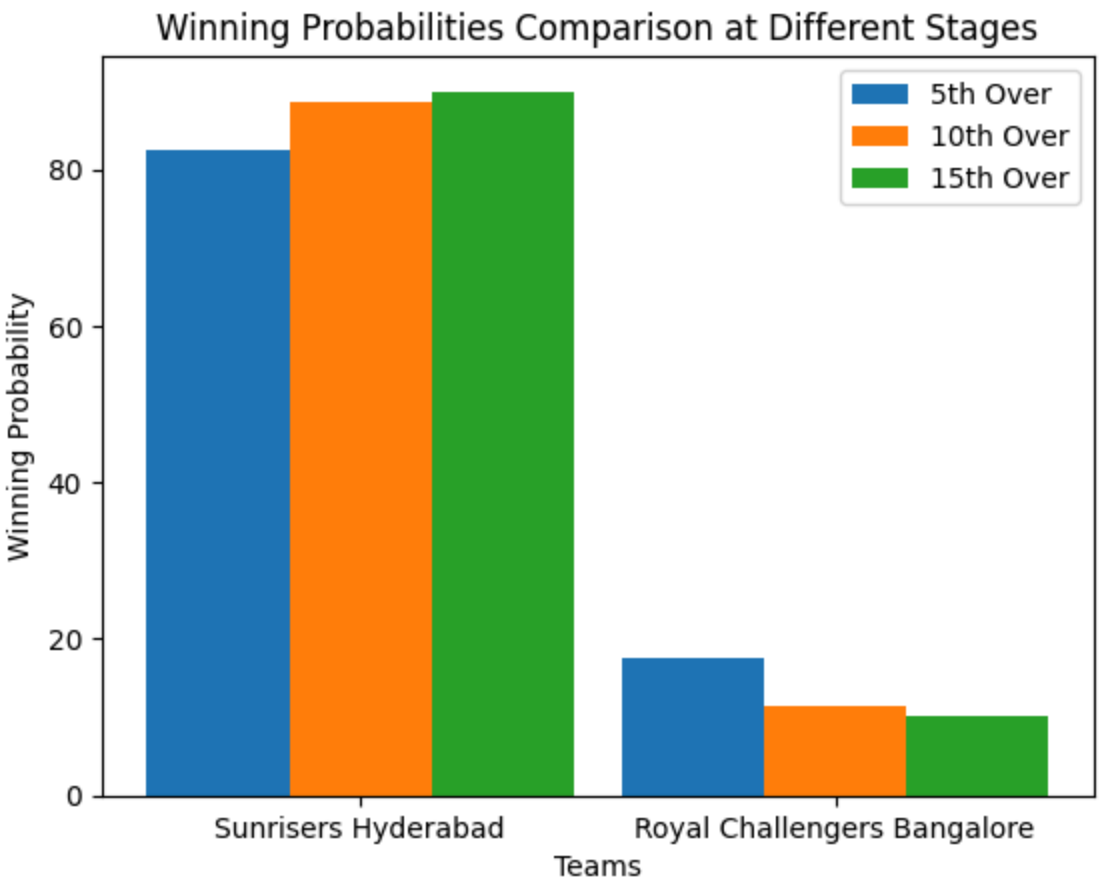
Winning Prediction by WASP:
Royal Challengers Bangalore: 11.45%
Sunrisers Hyderabad: 88.55%

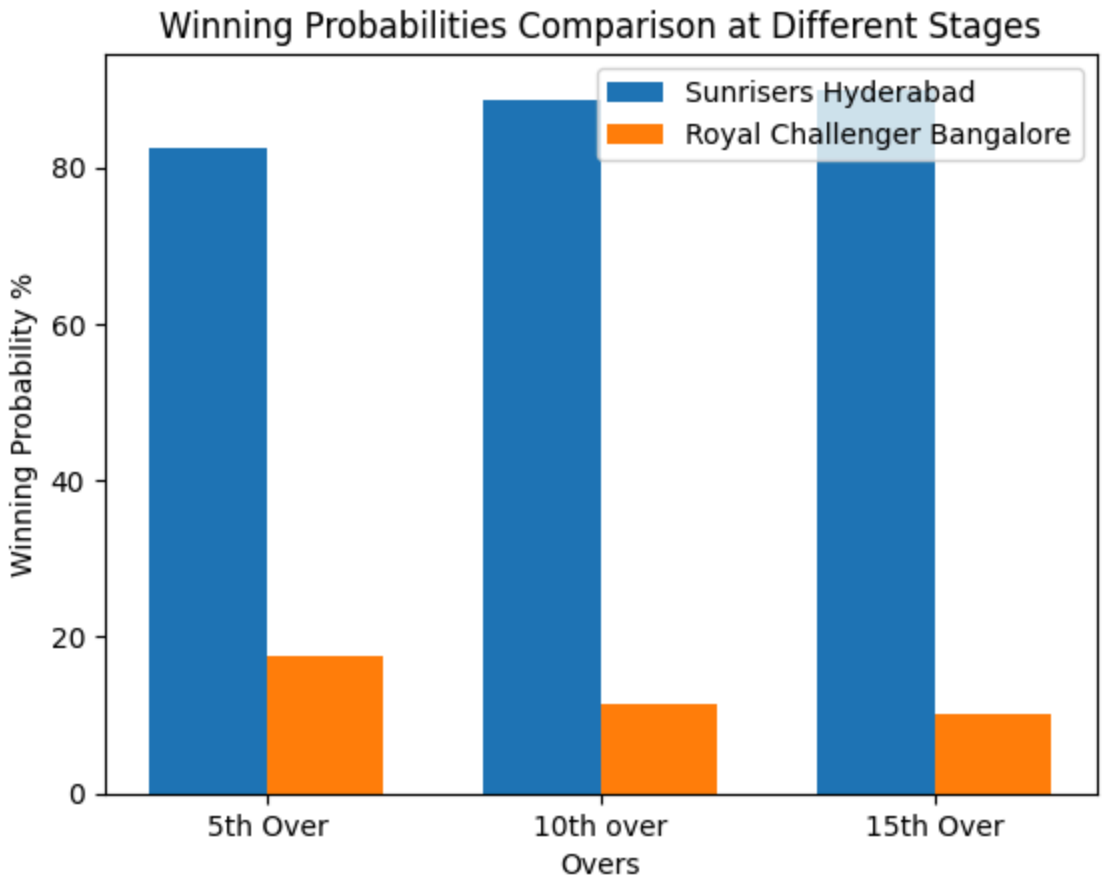
By the end of 15th Over:

Batsman	Bowler	Batsman_Runs	bowler_wicket	current_score	balls_left	crr	rrr	wickets	target
S Watson	Rashid Khan	13	2	139	30	9.2	13.6	5	207

Winning Prediction by WASP:
Royal Challengers Bangalore: 10.035
Sunrisers Hyderabad: 89.97%

FINAL Result: Sunrisers Hyderabad Won By 35 Runs


Winning Probabilities Comparison at Different Stages
5th Over
10th Over
15th Over
Winning Probability
0
20
40
60
80
Sunrisers Hyderabad
Royal Challengers Bangalore
Teams



Future Work:

- **Expand to Other Cricket Formats:** This WASP model currently only consists of IPL matches. We can expand it to include international matches, including Test matches and One Day Internationals (ODIs).
- **Model Optimization:** We can continuously optimize the current models to improve the accuracy. This includes experiments with different match situations, parameters and features. We could also experiment with different ensemble methods to find optimal predictions.
- **User Interface Development:** We could create a user interface or dashboard that allows users, including teams, analysts, and fans, to interact with and visualize the predictions in a user-friendly manner.