SPORTS RESULTS PREDICTION

Project report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology

In

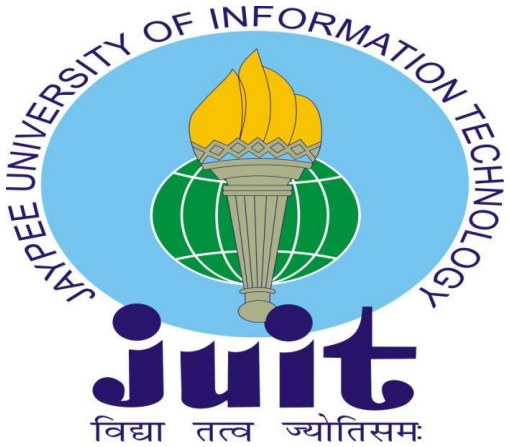
Computer Science and Engineering/Information Technology

By

AISHWARY KUMAR TIWARI (191442)
UJJAWAL TOMAR (191527)

UNDER THE SUPERVISION OF

Dr. Alok kumar
and
Dr. Himanshu Jindal
to



Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology Waknaghat, Solan-173234, Himachal Pradesh

Candidate's Declaration

We hereby declare that the work presented in this report entitled **" Sports Results Prediction "** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology Waknaghat is an authentic record of our own work carried out over a period from August 2022 to May 2023 under the supervision of D**r Alok Kumar** (Assistant Professor(S.G.) in Electronics and Communication Department) and Co-Supervisor **Himanshu Jinda**l (Assistant Professor (SG) in CSE & IT Department).
We also authenticate that We have carried out the above mentioned project work under the proficiency stream Data Science.
The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Aishwary Kumar Tiwari, 191442 Ujjawal Tomar, 191527

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

(Supervisor Signature)
Dr. Alok Kumar
Assistant Professor
(SG)
ECE Department
Dated:

(Co-Supervisor Signature)
Dr. Himanshu Jindal
Assistant Professor (SG)
(CSE & IT) Department
Dated:

PLAGIARISM CERTIFICATE

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: .....

Type of Document (Tick): PhD Thesis | M.Tech Dissertation/ Report | B.Tech Project Report | Paper

Name: _____ **Department:** _____ **Enrolment No** _____

Contact No. _____ **E-mail.** _____

Name of the Supervisor: _____

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): _____

_____

_____

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

- Total No. of Pages =
- Total No. of Preliminary pages =
- Total No. of pages accommodate bibliography/references =

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at .....(%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor) **Signature of HOD**

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

<table>
<tr><th>Copy Received on</th><th>Excluded</th><th>Similarity Index (%)</th><th colspan="2">Generated Plagiarism Report Details (Title, Abstract & Chapters)</th></tr>
<tr><td rowspan="2"></td><td rowspan="4">• All Preliminary Pages
• Bibliography/Images/Quotes
• 14 Words String</td><td rowspan="2"></td><td>Word Counts</td><td></td></tr>
<tr><td rowspan="2">Report Generated on</td><td>Character Counts</td><td></td></tr>
<tr><td rowspan="2"></td><td rowspan="2">Submission ID</td><td>Total Pages Scanned</td><td></td></tr>
<tr><td>File Size</td><td></td></tr>
</table>

Checked by
Name & Signature **Librarian**

.....

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File) through the supervisor at plagcheck.juit@gmail.com

Acknowledgement

With the guidance and assistance of numerous well-wishers, an endeavor over a lengthy period of time can be effective. We would like to take this time to let everyone know how much we appreciate them.

We wish to express our sincere thanks and gratitude to our project guide, Dr. Alok Kumar, Assistant Professor (S.G.), Department of Electronics and Communication Engineering, and Dr. Himanshu Jindal, Assistant Professor (SG), Department of Computer Science & Engineering/Information Technology at Jaypee University of Information Technology (JUIT), for the stimulating discussions, in analyzing problems associated with our project work, and for guiding us throughout the project. Project meetings were highly informative. We express our warm and sincere thanks for the encouragement, untiring guidance, and confidence she has shown in us. We are immensely indebted to her for her valuable guidance throughout our project.

We would like to express our heartiest gratitude to Dr. Alok Kumar, Assistant Professor (S.G.), Department of Electronics and Communication Engineering, and Dr. Himanshu Jindal, Assistant Professor (SG), Department of Computer Science & Engineering/Information Technology at Jaypee University of Information Technology (JUIT). We would also generously welcome each one of those individuals who have helped us straightforwardly or in a round about way in making this project a win. In this unique situation, we might want to thank the various staff individuals, both educating and non-instructing, which have developed their convenient help and facilitated our undertaking. Finally, we must acknowledge with due respect the constant support and patience of our parents.

Aishwary Kumar Tiwari (191442)
Ujjawal Tomar (191527)

TABLE OF CONTENT

LIST OF FIGURES

LIST OF GRAPHS

LIST OF TABLES

Abstract

Prediction is totally based on the past data , like how the team had perform in past years. Sports in current time generate sufficient substantial analytical data about every sportperson, different sport team, and seasons and games. Teamsports science was once thought to be owned by professionals like as coaches, managers, and analysts. Sports Recently, companies have discovered that the accessibility of science they have in their data and wish to utilise that science through the application of various data mining methods. sports analytics supports managers and coaches in a variety of ways, including forecasting the outcomes, player performance, identifying the talent, and Analyzing the game's tactics.

Prediction aids clubs and managers in making the best choices to win leagues and tournaments. The current study demonstrates that past research on data mining systems to forecast outcomes and assess the benefits and drawbacks of each technology. All sports have successfully used prediction. Although this application has been highly limited in many ways. It is crucial to investigate how machine learning is used in these situations and determine whether doing so can improve the analysis's outcomes.

This report intends to provide a solution that will help to create predictions more accurate and precise than the preceding methods by utilising datasets that are more exact and machine learning. The idea of prediction is used by many businesses because prediction is at the core of remarkable scientific areas. In this study, machine learning, a field of intelligent systems, will be applied to offer answers to the profession

CHAPTER-1
INTRODUCTION

1.1) Introduction

Prediction is a statement of an uncertain event which uses past data, analyzes it and predicts the future. This application focuses on predicting game result of the premier league on the basis of Back Propagation Algorithm. The input parameters chosen were Home Teams, Away Teams, Home Team Goals, Away Team Goals and Goal Differences.

Principles of Machine Learning are used for developing the prediction system. There are two types of machine learning namely supervised machine learning and unsupervised machine learning. In supervised machine learning we must train the machine by providing huge data sets and the outcomes. In this project, we use various classification algorithms such as Naïve Bayes, Support Vector Machine, K-Means also Logistic Regression for regression analysis and Decision Tree algorithms for making effective decisions feature selection.

Around the world, cricket is played in a lot of different nations. Numerous domestic and international cricket competitions are now taking place in various nations. Cricket is a sport in which two teams each have 11 players.Every team , Either a win, a defeat, or a draw is the outcome. Cricket is a game that cannot be played in the rain, but occasionally due to terrible weather the game is also called off. The game is also incredibly unpredictable because one team always gains momentum over the other at every stage of play. When the game is particularly tight, the outcome is frequently decided by the final ball of the game.

There is a lot of excitement among the audience to participate in some of these unforeseen game's scenarios. either before the game or during the game, forecasts. To win money, many spectators also participate in betting games. In

light of all of these potential outcomes, this research intends to investigate the issue of anticipating game outcomes prior to the commencement of the game using the statistics and data available from the data set. These are the various prediction methods. The performance of the players and the team can both be taken into account when making the prediction.

There are many unforeseen events that might occur during a cricket match, such as games being cancelled due to rain, a key player becoming hurt before the match, players switching sides, etc. Sometimes a crucial player is hurt in the course of the game and is unable to continue playing. To some extent, each of these variables does influence the prediction. The report goes into the process I used to forecast the outcome of the game. The process entails using attribute selection algorithms to narrow the list of qualities to just the most crucial ones, then using data mining algorithms to those attributes.The game prediction problem I'm working on does not factor in player performance, nor does it factor in team prior success to a significant degree, nor does it factor in other variables like the winner of the toss, the outcome of the toss, home crowd support, etc.

1.2) Factors

Like all other games, the outcome of cricket may be predicted. The best qualities or elements that affect the result of a match must be identified. A cricket match's outcome is more influenced by in-game and pre-game factors. The outcome of a match is mostly influenced by pre-game factors like the pitch, team strength, weather, location, etc. and in-game factors like run rate, total run, strike rate, wickets in hand, etc. The characteristics that determine a cricket match's outcome are listed below.

1.2.1 Pitch

Unlike other sports, the shape and size of cricket stadiums can vary, save for the The inner circle and pitch have respective sizes of 30 yards and 22 yards. respectively. Pitch and outfield changes can significantly impact bowling. Hitting and. The ball's bounce, seam movement, and spin are all dependent on the the pitch's nature. It depends on how wet the field is. The wetter the the slower it will play at that pitch. Those balls will dry out if there's any possibility. It won't dramatically alter, but it will become easier the drier it is. Green fields usually become less difficult to bat on. Wickets might become much more dry or wet.

1.2.2 Toss

Cricket analysts assert that a side will undoubtedly benefit if it the toss is won. This may not be the match's deciding factor, but it would give the team the freedom to decide "what they want."

1.2.3 Team Strength

To win a game, the team's strengths should be evenly distributed. Another deciding aspect is team captaincy. Records from the past: Match results can be predicted using team performances from the past. History of games played at those places, team performance there, performance versus that particular opposition there, and experience at that particular stadium.

1.2.4 Home Ground Advantage

This is another attribute which determines the winner in the match. If you are playing in the home ground condition everything would be in your hands like

climatic factor, pitch nature and major role is played by the home crowd. Home team gets better motivation.

1.3) Problem Statement

Cricket first appeared in England in the sixteenth century. Cricket is a team sport with a wide range of rules, rulesets, and lengths. Twenty20 is one of three types of cricket that are officially recognised by the International Cricket Council (ICC). There are 20 total overs in that arrangement, and each inning is played by two teams. Twenty20 cricket has gained a lot of popularity since it is played so swiftly and with such enthusiasm.Numerous events are held annually both domestically and internationally. In the business world, the ability to predict player performance is highly coveted. This has led to several assessments of both individual and team performance as well as game forecasts for various types of games.

With millions of admirers worldwide, sports have always been a popular form of entertainment. With the popularity of sports betting and fantasy leagues, it is more important than ever to predict the outcomes of sporting events accurately. Sports data from previous games is analysed using statistical models and algorithms to forecast outcomes for upcoming games. This big project report seeks to offer a complete technique for predicting sports results that can be used to a number of sports and yield accurate results.

Sports outcome prediction is a difficult problem that necessitates the examination of enormous amounts of data from multiple sources. The challenging challenges include identifying the most important data points and developing a model that can effectively utilise this data to generate accurate predictions. This necessitates both a solid understanding of the relevant sport as well as the ability to gather and analyse data in a methodical, efficient manner.

To develop an effective sports results prediction system, we need to address several key challenges. These include:

1. **Data Collection:** Any system that predicts sports outcomes must collect data with accuracy and thoroughness if it is to be successful. For this, it is necessary to identify pertinent data sources, such as match histories, player statistics, weather reports, and injury reports, as well as to create procedures for gathering and organising this data in an ordered way.

2. **Feature Selection:** Once the data has been acquired, the following stage is to identify the crucial factors that are most likely to affect the result of the game. The most important factors must be selected, such as team and player performance, injuries, and weather conditions, and processing and analysis algorithms must be developed for this data.

3. **Model Development:** To create a precise predictive model, advanced statistical techniques and machine learning algorithms must be applied. The model must be able to learn from prior data and recognise trends and patterns in order to make accurate predictions. The model needs to be adaptable enough to take into account a wide range of sports and variables.

4. **Performance Evaluation:** Finally, the accuracy of the model needs to be evaluated against real-world results. This involves testing the model on a range of different games and comparing its predictions to the actual outcomes of the game. The performance of the model needs to be rigorously tested to ensure that it can deliver accurate and reliable results.

1.4) Objective

- The first objective is to pinpoint the key elements that are most likely to have an impact on the game's outcome once the data has been gathered. Choosing the most pertinent variables, such as team and player performance, injuries, and weather conditions, and creating processing and analysis algorithms for this data are required.

- The second objective is to create a precise predictive model utilising machine learning algorithms and advanced statistical methods. In order to make precise predictions, the model must be able to learn from past data and spot trends and patterns. The model must to be flexible enough to accommodate various sports and a variety of different variables.

- The third objective is to compare the model's performance to outcomes from the real world. To do this, test the model on a variety of games and compare the results to what was really predicted. To ensure that the model can produce findings that are precise and trustworthy, its performance must be thoroughly tested.

- The creation of a user-friendly system that can be quickly included into current sports betting and analysis systems is the fourth objective. The system must be scalable, flexible enough to accommodate multiple sports, and capable of providing real-time, accurate, and dependable results.

- The ultimate goal is to make sure that a fair and ethical development process is followed while creating a system for predicting sports results. This entails creating procedures for data collection and analysis that are open and responsible.

1.5) Methodology

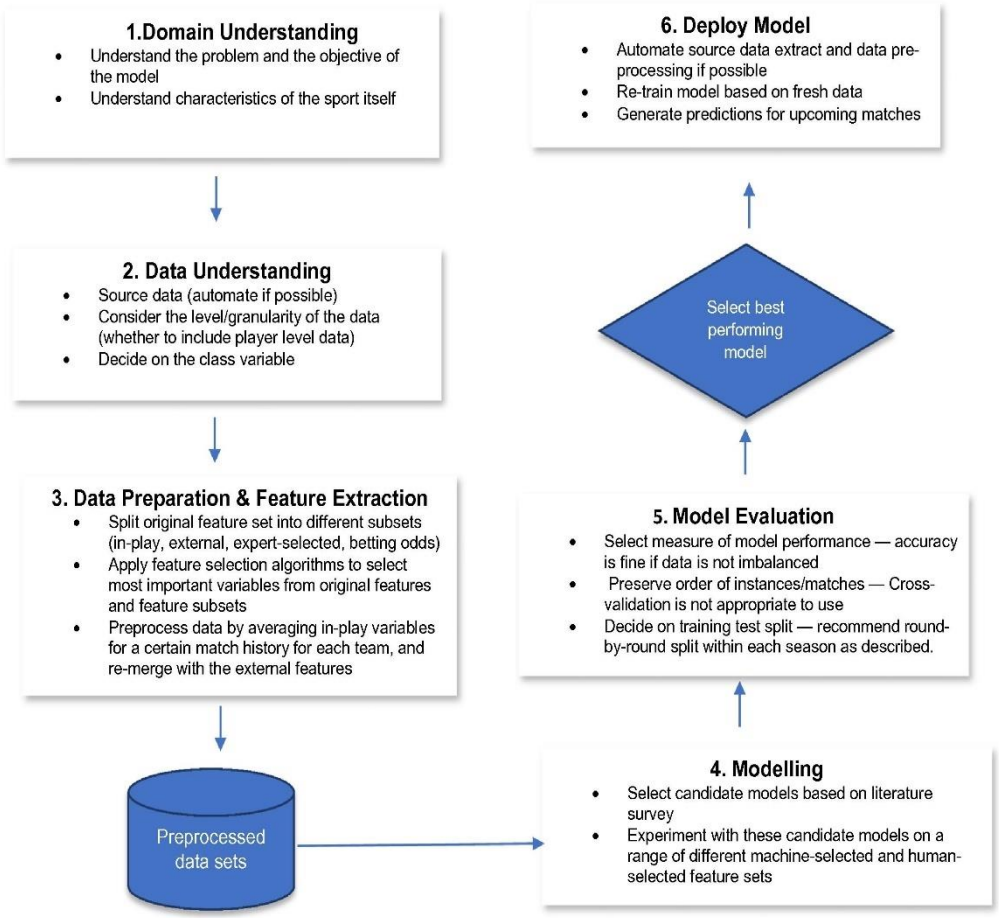


Figure - 1.1 Flowchart of our sports result prediction web app

1.5.1 Domain understanding - Understanding the problem, the purpose of the modelling, and the unique features of the sport itself are all examples of domain understanding. To do this, we must have some knowledge of the sport's rules and the potential contributing elements to match results. This could be discovered through personal experience with the sport, research into the body of existing literature, or by speaking with specialists in the field.Clarity is also required with regard to the model's goal. The goal could be to forecast outcomes in order to compete with professional predictions, participate in online competitions, or ultimately utilise the model's predictions to place wagers on

sporting events. The matches that will be wagered on must also be taken into account if the prediction model is to be utilised for betting. For instance, there will probably be a betting odds threshold where, despite the model's prediction of a win for that team, the odds are so low that the return does not justify placing any bets at all on that game.

1.5.2 Data understanding -

It is frequently possible to gather data for sports forecasting online from publicly accessible sources. Prior research has automated the data collection process by creating scripts that load online data into databases after automatically extracting it. Some studies have also developed an end-user interface that allows users to enter information about an upcoming game, after which the prediction is generated.One thing to take into account is the data's level. Training data from prior studies has been been at the match/team level. Player-level data, which includes information on the players who participated in each of the matches, is another option. It is typical for player level data to be separated into a separate data set that must be transposed and linked with match level data in order to include certain player statistics as attributes for each match.

The advantage of including player level data is that we can examine whether particular players' actions or presence are significant for the performance of the team in terms of winning or losing.It's also important to take into account how the class variable is defined. The majority of earlier research has approached the sport prediction problem as a classification problem with two or three classes of values (home win, away win) or (home win, draw, away win). Other earlier research has also approached the problem as a numeric prediction problem, using regression techniques to predict the points margin (home points minus away points) and then making a win-loss prediction based on the predicted points margin. In the end, the scientists discovered that approaching

the issue as a classification produced better results, but this does not imply that it would be true for all sports or data sets.

1.5.3 Data preparation and feature extraction -

Creating feature subsets : There are numerous subsets of features that can be found in sport result data. Divide the features into match-related and standings features, for instance, or compare the effectiveness of a feature set that only includes betting odds with one that also includes public data elements. To explore the utility of expert opinion for feature selection, we can also use a separate expert-selected feature set versus their own feature set. Of course, certain feature sets are chosen using feature selection methods, either using the entire feature set or just a portion of it. Ideally, we should evaluate the candidate classification models of several different feature selection approaches.

As a result, we only have the average of these characteristics for a specific number of previous games between these clubs. For instance, before to the contest, we would be aware of the average number of passes made by both sides every game, but we would not be aware of the number of passes made during the match until after it had already started. As a result, only historical average data for these traits may be utilised to forecast a future game. Therefore, before being re-merged with the external features, match-related features should go through a separate averaging process.

Figure - 1.2 match-related data should be averaged over a predetermined number of games.

Data set - The keggle was used to gather the data. The website contains information on every game played in an international event between 2008 and 2016. It is an ODI format for competitions. It indicates that each team has a maximum of 50 overs to bat or bowl. This website's data set was in when it was downloaded. format CSV. I read the using the Python Pandas library. CSV file, then copy the data to a new file. The . The CSV file had
information about 5219 separate entries, and the python code essentially duplicated the crucial information from all 5219 files before combining them into one file. information from the. More irrelevant data, such as gender, date, umpire information, etc., can be found in the CSV file.

1.5.4 Modelling -

Choosing which candidate models will be utilised in the experiment is the first step in the modelling process. This would entail a review of earlier material and the identification of widely used predictive models that have had success in the past. Then, on each feature subset and subsets chosen by feature selection techniques, each model may be tested. The ideal classifier and feature selection technique will be found through experimentation with these various feature selection methods and classification models.

1.5.5 Sport prediction model evaluation -

Measuring model performance : We would classify match results into home wins, away wins, and draws (if the sport has draws) in order to evaluate model performance, and we would then look at the number of matches that the model had correctly identified using a conventional classification matrix. Although a significant degree of imbalance in the class values for the dataset is unlikely, one is expected to find a little skew in favour of home victories given the frequently seen home advantage phenomena. Classification accuracy is a reasonable evaluation metric in this situation. ROC curve analysis might be preferable in situations where the data is seriously unbalanced.

Training and testing : For the sport prediction problem, it is crucial to maintain the order of the training data so that only previous matches are used to predict upcoming matches. For the sport outcome prediction problem, cross-validation is an inappropriate method of dividing the data into training and testing because it typically entails rearranging the sequence of the occurrences. The order of the instances should be retained in a held-out training test split. Software for machine learning, like WEKA, offers the choice to maintain the order of the instances.

Selecting an adequate training-test split is necessary. This may depend on the quantity of data the researcher has available—whether they have data from numerous seasons or just one season. Professional sports tournaments are typically formatted into rounds, with teams playing games throughout the weekend. In every round, teams typically play one game, unless they have a "bye." When only one season of data is available, it is necessary to decide how many rounds will be used to train the model and how many rounds will be used to test the model.

Another alternative to round-by-round prediction is to refresh the training data set after each match is finished. In this situation, all previous games up to the

present game will serve as training data, and the next game will serve as training data (only having that one record as the training data). This is similar to cross-validation with an order-preserved leave-one-out.

Except in situations where teams play more than one match during a single competition round, this match-by-match approach is probably not required.

1.5.6 Model Deployment -

The process should ideally be automated such that new round data is downloaded from the internet and added to the match database (or otherwise added to the database manually by the end-user). The model is then retrained using the fresh training data and new matches are predicted after adjusting the training data and test data. The end user is then provided with predictions. Additionally, the learning model in the proposed architecture might be live and dynamically receiving input data

both before the game starts (external features) and during play (match features). Additionally, it should be incremental in that the training data set is updated often, causing the classifier to adapt to reflect changes in the learning environment.

1.6 Motivation

There are many instances in daily life where we require such data analysis, some of which are listed below and call for the immediate use of such software.

For instance, there are users that place bets using the software DREAM 11, and in those cases, the software can be helpful because it provides information on all players, enabling the user to make an effective team choice.

It can also be used by commentators or analysts to illustrate the standing and accomplishments of players on a particular field. Some youtubers find it useful to examine some players' gameplay in their videos

1.7 Scope

- It can also be utilised on a smaller scale, such as in interschool competitions and district-level competitions, to evaluate the performances of different teams and thereby enhance competition.

- It mostly serves as a tool for analysing which specific field a team has a higher winning rate at.

- On applications like DREAM 11, where users select their teams and play against one another for prizes, it can be utilised to predict the teams. It will assist the person in assembling a group of players who can work well together under the given circumstances.

- It can be used to forecast a team's total winning odds in a specific match on a specific surface.

- The team manager can use it to analyse the team and create one that has a higher chance of success.

CHAPTER-2
LITERATURE SURVEY

---

[1] This research paper the most important aim of analyzing sports data is to upgrade team achievement and enhances the probability of getting victory in the match.They cannot anticipate receiving clear, properly prepared data from Kaggle. Consequently, data pre-processing is so important that they cannot emphasise how crucial that is. It is the most crucial stage because just cleaning the data before feeding it to your models can take up 40% to 70% of the entire procedure. They used three scripts to scrape information from the Crickbuzz website, including team rankings as of May 2008, information on the 2008 IPL Twenty-Twenty schedule, and information on each team's record in prior IPL Cup games.The data must now be built into machine learning models using the data that has been processed and cleansed. They compared their models with Scikit Learn's.Their findings were almost identical to those of the Scikit Implementation.

They claim that the two subfields of supervised learning in the realm of machine learning for predicting the winner of the IPL-2020 are classification and regression. These are the fundamental concepts that one should become familiar with before beginning their machine learning adventure. The only way to study and master these subjects is through projects.

In their work, they achieve an accuracy of approximately 90%, which is higher than the maximum prediction accuracy of 85%. They plan to forecast the results of international one-day games, test matches, and 20-20 games using this methodology. This model can also be used to forecast the results of other sports, such as baseball, rugby, football, tennis, and more. According to the study, there are a lot of factors that affect how an IPL match turns out.

The host team, away team, stadium, toss-winner, and many other things are the primary variables that fundamentally affect any IPL match. This somewhat aided in the strength calculation. For the IPL data set, various ML approaches were applied, which aided in their research. The data collection includes all IPL games played during the last six years, from 2014 to 2019. The created models were used to predict IPL match outcomes. T20 cricket offers the potential for change because only a few balls can completely alter the course of the match. When the IPL began 12 years ago, there were many fewer games played than in fifty - fifty and test cricket matches.

As a result, constructing ML for game outcome prediction with a precession of 75% at this time is highly successful.

[2] With an emphasis on the application of neural networks (NN), the paper offers a comprehensive analysis of the literature on machine learning for sport outcome prediction. This use of NN is a relatively new study area, even though there have been many studies that have looked at sports outcome prediction in the statistics and operations research literature. The authors look at the difficulties in predicting sporting results using intelligent models and suggest a CRISP-DM-style framework (SRP-CRISP-DM) based on the six phases of the conventional CRISP-DM framework. This study aims to support academics, researchers, students, sports fans, team management, and bookies in their efforts to tackle the challenging challenge of sport outcome prediction utilising NN-based methods. The authors hope that this paper will be useful for future research in this area.

Predicting a target variable from fresh, unexpected data is a common machine learning assignment. Building a classification model using training data that can forecast the value of the target variable or class for test data is the goal of classification. Since the target variable directs the training process, this is

referred to as supervised learning. Applications for classification models include loan approval, medical diagnosis, and email screening, among others.

Artificial neural networks (ANNs) are a well-liked approach for forecasting sports outcomes when it comes to machine learning. ANNs are made up of neurons, which are interconnected components that use inputs to produce outputs. Due to the hidden neurons' ability to alter the weights that contribute to the final output due to their non-linearity, ANNs are a useful tool for classification tasks. Throughout the training phase, the weights attached to the interrelated components are continuously adjusted to achieve high anticipated accuracy. The classification model's attributes from the training dataset are used to build the ANN model. As a result of this approach, the model may overfit and perform well on the training data but poorly on the test data.

[3] The purpose of this work is to explore the application of data mining and machine learning in sports. With the growth of data science and sport analytics, the importance of these techniques for forecasting performance and results has increased. The paper addresses the analysis strategies that have already been used in the literature in order to explain the data collecting and analysis prediction processes and to identify the performance-influencing variables. The study makes a reliable machine learning tool recommendation for data mining analysis. Through this inquiry, the authors hope to highlight the crucial part that technology plays in forecasting sporting success and provide suggestions for future research in this field.

[4] The research uses data analytics methods from the field of machine learning to create predictions about cricket matches that have been impacted by rain. Because of a multitude of factors, such as team selection, batting, bowling, fielding, and target revision, these types of games are challenging to predict.

In order to circumvent this issue, the authors propose a mathematical model based on historical match outcomes and employ support vector machines (SVM) to create predictive models. They also develop a piece of software called the Deep Mayo Predictor, which anticipates cricket scores and player performance using statistical analysis, data mining, Bayesian prediction techniques, and parameter-based filtering. The authors also employ clustering methods like KNN and MLP and supervised methods like SVM and Naive Bayes to efficiently categorise data. They also look at Kaluarachchi's work, which categorises match-influencing factors such home field advantage, the day/night effect, the toss, and batting first using Bayesian classifiers in machine learning. This effort led to the development of a software programme called CricAI, which can be used in the real world to alter crucial factors and increase winning chances.

[5] The study studied a sizable dataset of football matches in order to estimate the outcomes of the most watched football leagues in Europe and the rest of the world. The authors' three classification algorithms—Nave Bayes, Support Vector Machine, and Logistic Regression—performed differently before and after normalisation. Goals scored, goals conceded, corner kicks, red and yellow cards, as well as other attributes of offensive and defensive teams, were some of the study's prediction factors. The researchers found that Multinomial Naive Bayes and SVM were two machine learning algorithms that performed better without normalisation. But logistic regression performed the best when normalisation was applied.

The authors also mentioned that the accuracy of the forecasts was enhanced with the addition of new data variables, such as player or team form.

Below is a table which contain summary of different literature review -

[Ref No.]	Author(s)	Published By (IEEE,Elsevier,Springer)	Pros and cons
[1]	Rushikesh Bhor,Ajay Jagdale, Shubham Pisal, Rohit Korke	Predicting Match Winner Using Machine Learning.	Instead of utilising the time-consuming traditional method, the best team may be chosen using data mining techniques. Accuracy is low as compared to other.
[2]	RoryP. Bunker, Fadi Thabtah	A machine learning framework for sport result prediction, Auckland University of Technology, Auckland,New Zealand	It will some more variables, such as player information, the halftime score, the final score, and many more, increasing the prediction's accuracy. It was trained only one model with accuracy of 80.035
[3]	Said Lotfi, Mohamed Rebbouj	Machine Learning for sport results prediction using Algorithms , International Journal of Information Technology and Applied Sciences	The Indian Premier League's batting statistics, which contain fuzzy data, are categorised into the proper groups using the clustering principle.

[4]	G. Sudhamathy and G. Raja Meenakshi	Prediction on ipl data using machine learning techniques in r package	The application of machine learning and data mining techniques, which have not been applied in this sector but have produced positive results in other domains, can increase prediction accuracy. Researchers are advised to get information from reliable leagues so that there would be an opportunity for comparisons among various studies.
[5]	Ragini Singla,Dr. Amardeep Singh	Sport Prediction using Machine Learning	The only two variables that affect a team's ability to win are the home team and team ranking.
[6]	SKevin Andrews, KLakshmi Narayanan,K Balasubadra,M S Josephine	Analysis on Sports Data Match Result Prediction Using Machine Learning Libraries	There isn't much data to support the claim that a human agent can predict match outcomes more accurately than a machine learning approach.

Table 2.1 : Summary of different literature review

CHAPTER-3
SYSTEM DEVELOPMENT

3.1) System Configuration

This project can be run on common hardware. To accomplish the project, we used an Intel I5 CPU with 8 GB of RAM, a 2 GB Nvidia graphics processor, and 2 cores that run at 1.7 GHz and 2.1 GHz, respectively. In the test phase, which comes after the training phase and lasts for roughly 10-15 minutes, predictions may be made and accuracy can be determined in a matter of seconds.

3.2 Analysis

The ubiquity of machine learning techniques and rapid access to the Internet have fueled sports betting and research. Compared to other sports, football is thought to be the most popular game in 200 different countries.

Soccer is a fascinating subject for investigation because it is thought to be significantly more diversified and complex. The creation of prediction systems involves the application of numerous procedures and techniques. We forecast the result of match between a team from the Premier League and the host team.

The estimates are supported by a wide range of noteworthy data points from prior Premier League campaigns. These crucial qualities could affect how well a team performs in a contest.

The "Sports Internet" has undergone significant changes, one of which is the inclusion of the financial success of various professional athletes in the codex. The company has made headway toward ambitious growth.

Activities and pressing issues, like comparing athletes, were hardly ever completed in the senior years in a timely and orderly manner. With the help of

the Internet, the periodical assessment of these conundrums was simple to implement and to monitor. In the end, we can say that as the sport is growing day by day ,craze of sport is going at next level and this is also letting to beneficial in betting industry as many models are designed nowdays to predict the result of sport . so nowdays fatasy website owners are gaining very much profit from betting with the help of internet.

3.3 Requirement analysis

3.3.1 Functional requirement -

- Forecasts the outcome of a match between two teams.
- Compares the outcomes of the trained data.

3.3.2 Non - functional requirement -

- The basis for prediction is the team name.
- The outcome of the game -- whether a team wins , loses , or ties -- will be indicated.

The basic functionality of this application is that it predicts the result of the game between two teams. The results are then compared with the data that are being trained . The non-functional requirement includes displaying of the input field and a predict result button, then a result will be shown comparing the data with the trained data.

As we see below Figure 3.1 , it is use case diagram of our web app , it basically takes input as 2 teams and give the predicted result.

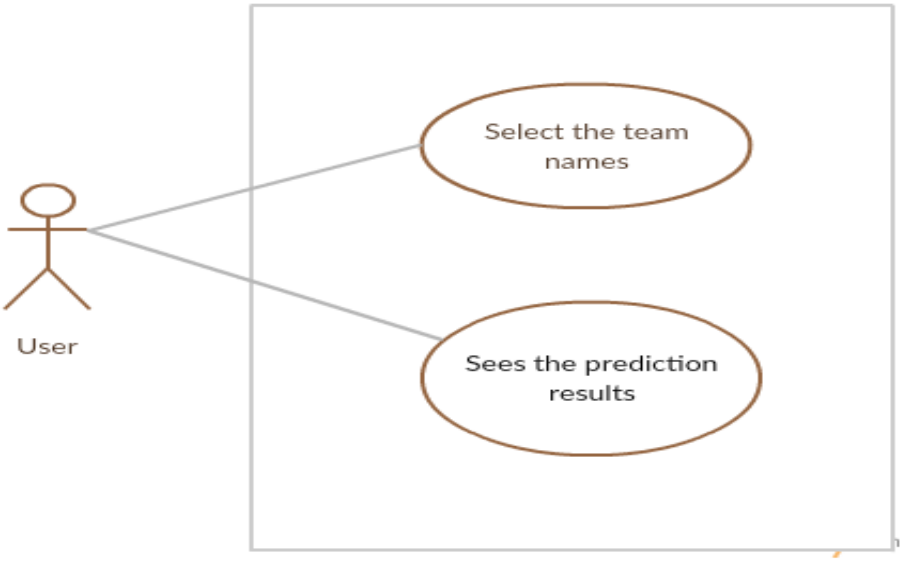


Figure 3.1 - Use case diagram of Sports Results Prediction

3.4) Feasibility analysis -

3.4.1) Technical Feasibility -

The Random forest method is the foundation of the programme Sport Result Prediction. To display the content for the user, it makes use of the various ml concept like Random Forest,SVM, gradient descent and also some backend language like Flask and Html ,css as a front end. Since all of the technology required for Sport Result Prediction is readily accessible and open source, it is technically doable.

3.4.2 Operational feasibility

The user interface of Sport Result Prediction is straight forward and friendly, making it simple to use. User can easily visit the web app input the query and see the result . Due to the availability of the internet, the user can see the outcome by entering the team names, which makes it practical and more suitable , it is also very fast so that user cant face any problem .

3.4.3 Schedule feasibility

The Sport Result Prediction application has been given a time limit of 47 days, as seen in figure 2.2 it shows activity network diagram for sport result prediction as indicated below:

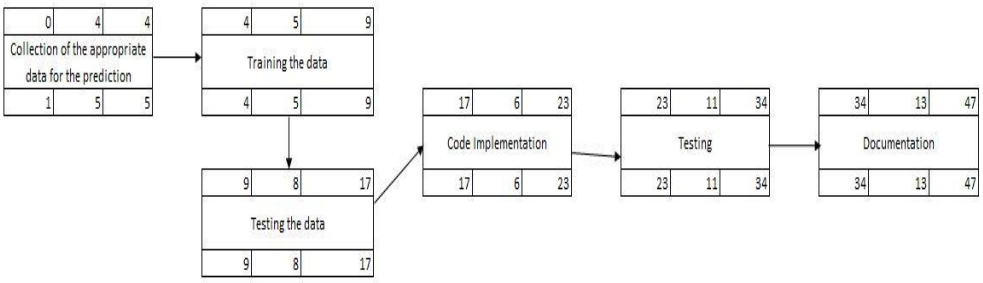


Figure 3.2 Activity network diagram for Sport Result Prediction

3.5 System Architecture

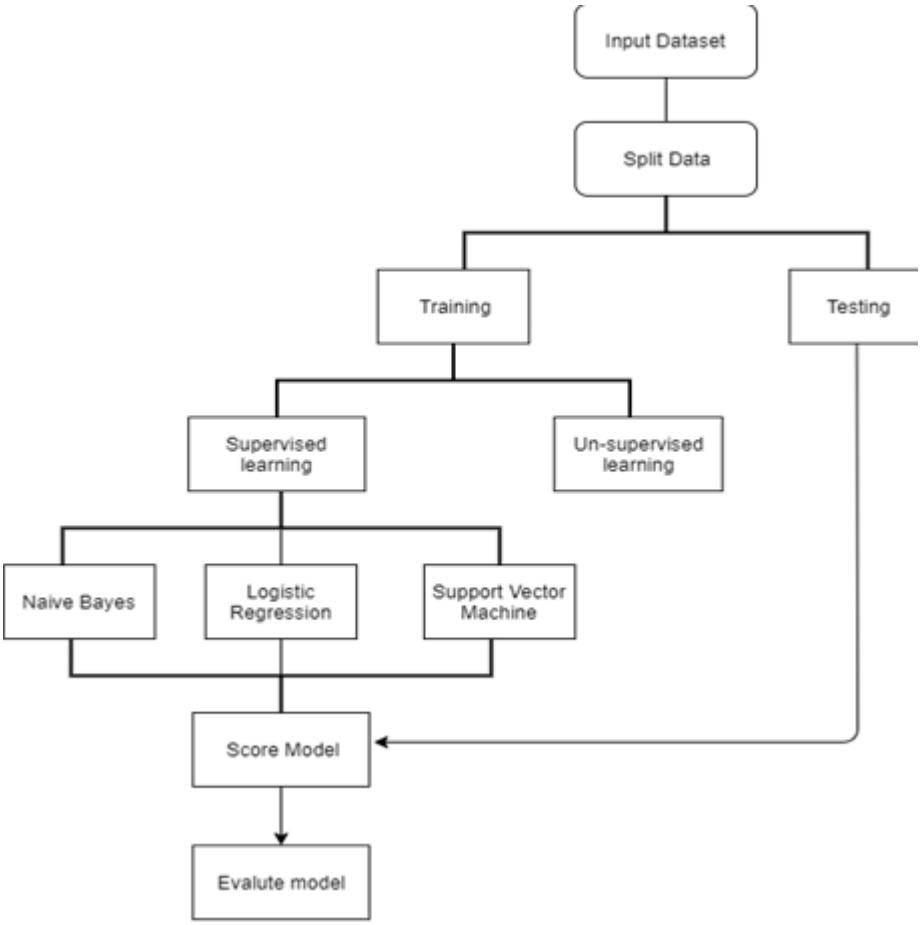


Figure - 3.3 Architecture of Sport Result Prediction

.

3.6 Design of Project

Our web app contains three models , which predict 3 things -

1 - winning team based on previous data set

2 - winning percentage in second innings after 5 overs based on target , wickets fallen,and runs scored in previous overs

3 - second inning run prediction

3.7 System Design –

3.7.1 Sequence Diagram

The sequence diagram for the sport Result Prediction is shown in the figure 3. It provides the application's flow from the user to the system and back again. When a user selects a team from a list of teams with a trained set of data, interprets it, and then gives the user the expected outcome. Moreover, it is admin provides trained and tested results, which are then displayed on the application.

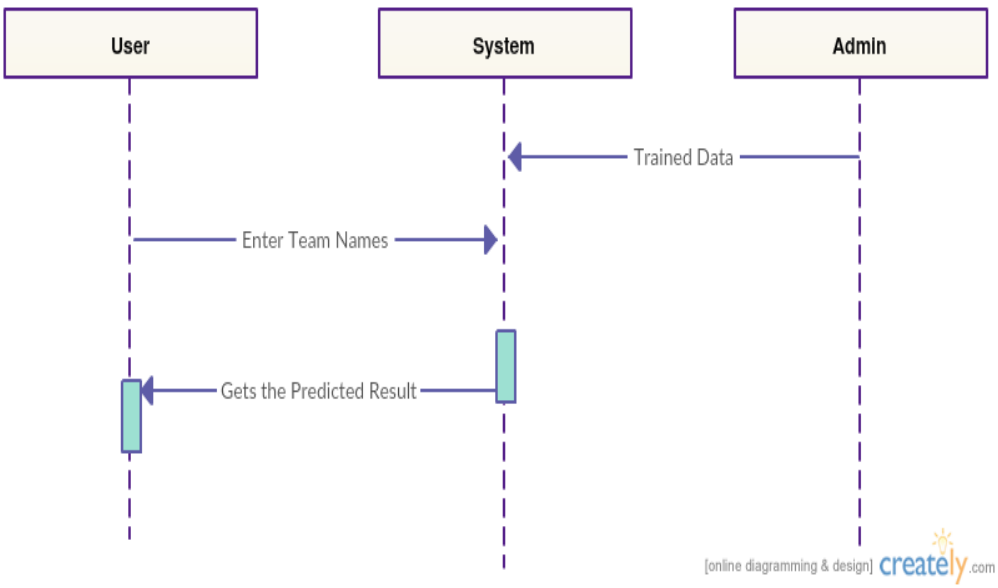


Figure - 3.4 Sequence diagram of Sports Result Prediction

3.7.2 Event Diagram

The sports result prediction event diagram is shown in the above image. It shows how events and a process are related to one another. Admin first Data that is registered for testing purposes is trained, and testing is then carried out. After the test is run, an intuitive user interface is created, and the user joins the team names into the fields and selects the predict button to display the outcome.using the user.

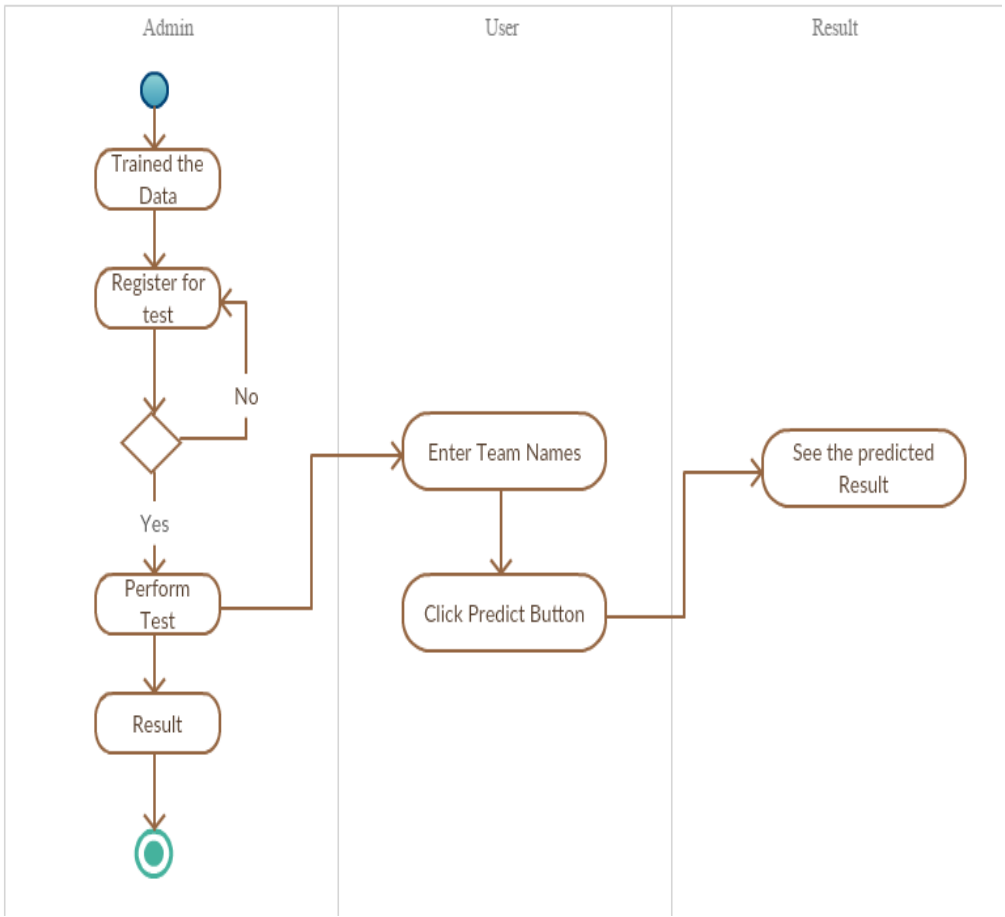


Figure - 3.5 Event diagram of sport result prediction

3.8 Dataset used for building a sports results prediction-

The dataset utilised is the ipl dataset, which can be acquired through Kaggle and includes information on every match from 2008 through 2021.The keggle website is where the information was gathered. Data are on the webpage. about every IPL game (from 2008 until 2021). The tournament format is Twenty-Twenty. It indicates that each team has a maximum of 20 overs to bat or bowl. The website's data set was in.CSV format when it was downloaded. After then after installing pandas and using pd.read() , the.CSV file was read and wrote the content to a newly created file using pandas, the file writer, and the python classes file.

The.CSV file contained information about 5219 separate entries, and the python code essentially duplicated the crucial information from each of the 5219 files before combining them into one file. information from the. The python code replicated just information regarding the team1,team2,place where match was held,city , toss_winner & toss_decision and each ball data .

The CSV file contained additional unnecessary information like gender of players, match date, description about umpires, etc. that was totally eliminated. . data is shown in table below -

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets
0	1	2008	Bangalore	4/18/2008	Kolkata Knight Riders	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Kolkata Knight Riders	140	0
1	2	2008	Chandigarh	4/19/2008	Chennai Super Kings	Kings XI Punjab	Chennai Super Kings	bat	normal	0	Chennai Super Kings	33	0
2	3	2008	Delhi	4/19/2008	Rajasthan Royals	Delhi Daredevils	Rajasthan Royals	bat	normal	0	Delhi Daredevils	0	9
3	4	2008	Mumbai	4/20/2008	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	normal	0	Royal Challengers Bangalore	0	5
4	5	2008	Kolkata	4/20/2008	Deccan Chargers	Kolkata Knight Riders	Deccan Chargers	bat	normal	0	Kolkata Knight Riders	0	5
5	6	2008	Jaipur	4/21/2008	Kings XI Punjab	Rajasthan Royals	Kings XI Punjab	bat	normal	0	Rajasthan Royals	0	6
6	7	2008	Hyderabad	4/22/2008	Deccan Chargers	Delhi Daredevils	Deccan Chargers	bat	normal	0	Delhi Daredevils	0	9
7	8	2008	Chennai	4/23/2008	Chennai Super Kings	Mumbai Indians	Mumbai Indians	field	normal	0	Chennai Super Kings	6	0
8	9	2008	Hyderabad	4/24/2008	Deccan Chargers	Rajasthan Royals	Rajasthan Royals	field	normal	0	Rajasthan Royals	0	

Table - 3.1 Dataset as downloaded from keggle

3.8.1 Data Set Description :

The information from the.CSV files was included in the new merged data set that I created. 5219 unique instances and 5 distinct attributes make up the data set. The data collection includes all of the One Day International seasons from 2008 to 2021. The data set's classifier is the winning team, and the goal of this project is to forecast which side will prevail in the game.The Information contain in data was accurate and it is very usefull in building the model , The following are the attributes in the data set:

- Season
- City
- Team1
- Team2
- Toss winner
- Toss Decision
- Winner
- Win by runs
- Win by wickets
- Venue

3.8.2 Data Cleaning :

In our dataset major work for pre processing of data was data cleaning -

- Missing values - Fill in the missing value manually,also using attribute mean and global constant.

- Noisy Data - using Clustering and Regression.

- Inconsistent data - Manually using external references.

After cleaning the data data becomes accurate without any fault.

Table of dataset after Cleaning -

	team1	team2	city	toss_decision	toss_winner	venue	winner
0	2	3	2	1	3	14	2
1	5	9	6	0	5	22	5
2	6	7	9	0	6	8	7
3	1	3	22	0	1	34	3
4	4	2	21	0	4	7	2
5	9	6	16	0	9	26	6
6	4	7	14	0	4	23	7
7	5	1	7	1	1	15	5
8	4	6	14	1	6	23	6
9	9	1	6	1	1	22	9
10	3	6	2	1	6	14	6
11	2	5	7	0	2	15	5
12	1	4	22	1	4	4	4
13	7	9	6	0	7	22	9
14	5	3	2	0	5	14	5
15	2	1	21	0	2	7	1
16	7	3	9	1	3	8	7
17	4	9	14	1	9	23	9
18	6	2	16	0	6	26	6
19	5	7	7	0	5	15	7

Table - 3.2 Dataset after cleaning

3.8.3 Attribute Selection

The attribute selection procedure required to be used to filter the data set that had been produced after dealing with the missing values. Considering that there were 21 attributes, it was required to pinpoint each crucial attribute that would be beneficial for data mining jobs. The attribute removal procedure was described in detail in the publication by Haghighit. Here, a set of dataset with a

certain amount of allocate is initially chosen, and after that, one by one, each attribute in the set gets erased. The deletion is done based on the results of using the classification algorithm on the collection of attributes.. If accuracy increases after an attribute is removed, the attribute is totally removed; otherwise, the attribute is retained in the data set.

As a result, after going through this elimination procedure, we have a collection of qualities that will give us the classification algorithms' best prediction accuracy. The wrapper approach and the ranker method, the two types of selection algorithms employed by the authors in their investigation, are described in detail in the publication by Trawinski [2]. For the attribute selection stage, I referenced this paper. Before using an attribute to make a prediction during the data mining phase, I wanted to assess its value and rank each attribute to determine their significance.

We now have a group of characteristics that will provide the classification algorithms the best prediction accuracy as a result of this elimination process. The authors of the Trawinski article used two different types of selection algorithms in their analysis: the wrapper technique and the ranker method. For the attribute selection stage, I referenced this paper. Prior to using each attribute to generate predictions during the data mining phase, I planned to assess its value and rank each one according to relevance.

3.9 Development

Our Project application will include the following components:

1. Frontend Design: HTML, CSS, JavaScript

2. Backend: Flask

3. Deployment: localhost

4. Model design using ML and its libraries.

Frontend design -

Our home page UI looks like as shown in figure below , it contains link to cricket page , football page and hockey page , the cricket nav link contains some sub links of 3 different models of cricket prediction .

Figure - 3.6 Home page UI

Backend Flask Code - The backend is designed using micor-web framework of python called flask , it is used to deploy ml model to web . The below diagram illustrate the code structure of our web app , where app.py is file where all user defined api are done .

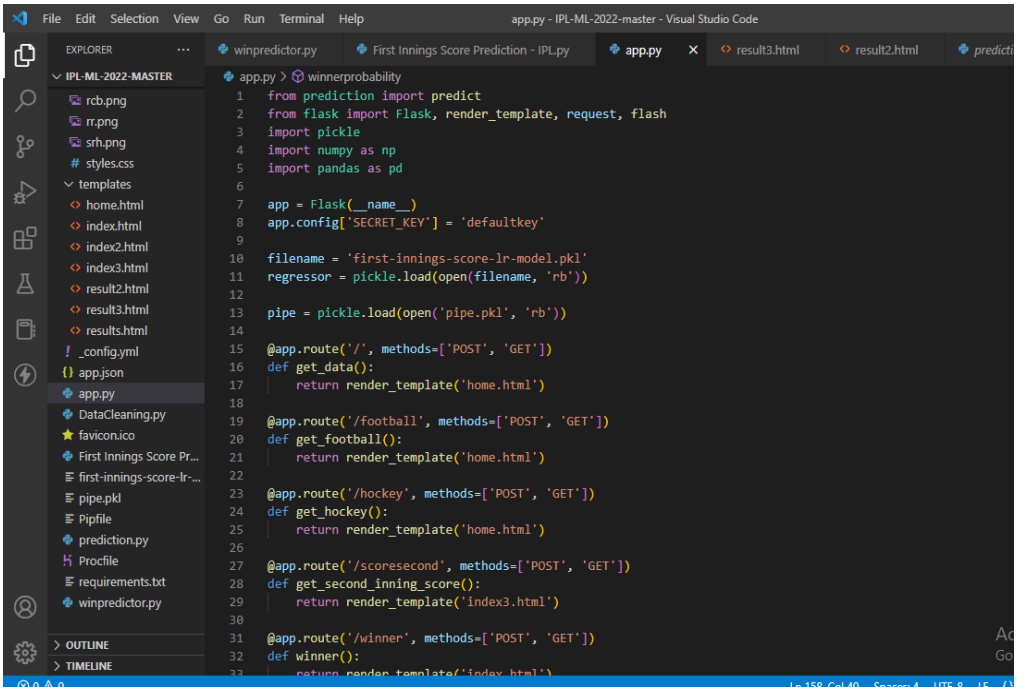


Figure - 3.7 Backend Flask code

CHAPTER-4
PERFORMANCE ANALYSIS

4.1 Analysis -

People are becoming more and more interested in conducting research to enhance learning management systems. Sports organisations must complete a number of duties in a limited amount of time. Analysing a team's performance by hand requires time and raises the chance of error. This process is streamlined by a machine learning system. Based on a range of characteristics, including past performance, socioeconomic conditions like weather, and pitch time, it examines a team's performance and forecasts future outcomes. This paper proposes a machine learning model that employs various techniques to anticipate a team's future grade. Sports analysts can employ this strategy to assess student performance and give teams performance-related advice.

Past performance as well as extra real-world factors are taken into account in order to accurately predict a team's success. Thanks to machine learning, applications may now produce precise predictions without explicit coding. The application of machine learning to prediction, speech recognition, spam filtering, and teaching is common. The Sports Result Prediction System is designed to take input data, train on that data, and produce the required range of outcomes.

Below in a table - 4.1 showing Survey of various models and their purpose -

Survey	Methods	Sports	Purpose
Haghighat et al.(2013)	Machine learning & Knowledge based system	Team sports & Non-team Sports	Match result prediction
Razali et al. (2018)	Bayesian Network	Soccer	Match result prediction
Beat et al. (2019)	Artificial Intelligence	Team sport and fantasy sport	Match result prediction , player investments,injury prediction
Keshtkar Langaroudi and Yamaghani (2019)	Machine learning & knowledge-based system	Team and Non-Team Sports	Match result prediction,tactical behaviour
This review	Machine Learning	Team sports	Match result prediction

Table 4.1 Survey of various models and their purpose

4.2 Implementation of analytical model -

4.2.1 Implementation of Logistic Regression -

Many individuals use the categorization method known as logistic regression to group observations into distinct groups. It is widely used to solve problems like email spam filtering, fraud detection online, and figuring out if a tumour is benign or malignant. Logistic regression uses the logistic sigmoid function to transform its output into a probability value.

In its most basic version, logistic regression predicts the target variable, y, using a collection of input features, X. It is a method of supervised

classification. The target variable in a classification problem can only take discrete values.

Contrary to popular belief, a logistic regression model is not the same as a regression model. When using logistic regression, a programme creates a regression model to determine the likelihood that a certain data entry belongs to the group designated by the number "1". The logistic regression model models the data using the sigmoid function, as opposed to the linear function used in linear regression.

```
[ ] X=final_result.drop(['winner'],axis=1)
    y=final_result['winner']
    X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.30,random_state=42)
```

```
model=LogisticRegression()
model.fit(X_train,y_train)
train_score=model.score(X_train,y_train)
test_score=model.score(X_test,y_test)
print("Traning accuracy: ",train_score)
print("Testing accuracy: ",test_score)
```

```
Traning accuracy:  0.6628352490421456
Testing accuracy:  0.5822222222222222
```

Figure - 4.1 Code of Logistic model design

4.2.2 Implementation of Random Forest Classifier -

Popular machine learning techniques like Random Forest are a part of the supervised learning strategy. It can be used for classification and regression problems in ML. The strategy is based on the concept of ensemble learning, which involves combining different classifiers to tackle a difficult problem and improve the performance of the model.Instead of relying exclusively on one decision tree, the random forest method makes predictions based on the combined input from a number of decision trees on distinct subsets of the input dataset. As a result, the probability that the algorithm would overfit the training set of data is reduced.One of the key characteristics of the random forest method is its ability to train more quickly than other algorithms. Additionally, it can continue to be accurate even in the absence of a sizable amount of data. Furthermore, Random Forest accurately predicts outcomes and performs well even with large datasets.

The below figures 4.3 illustrates about implementation of random forest model and its output -

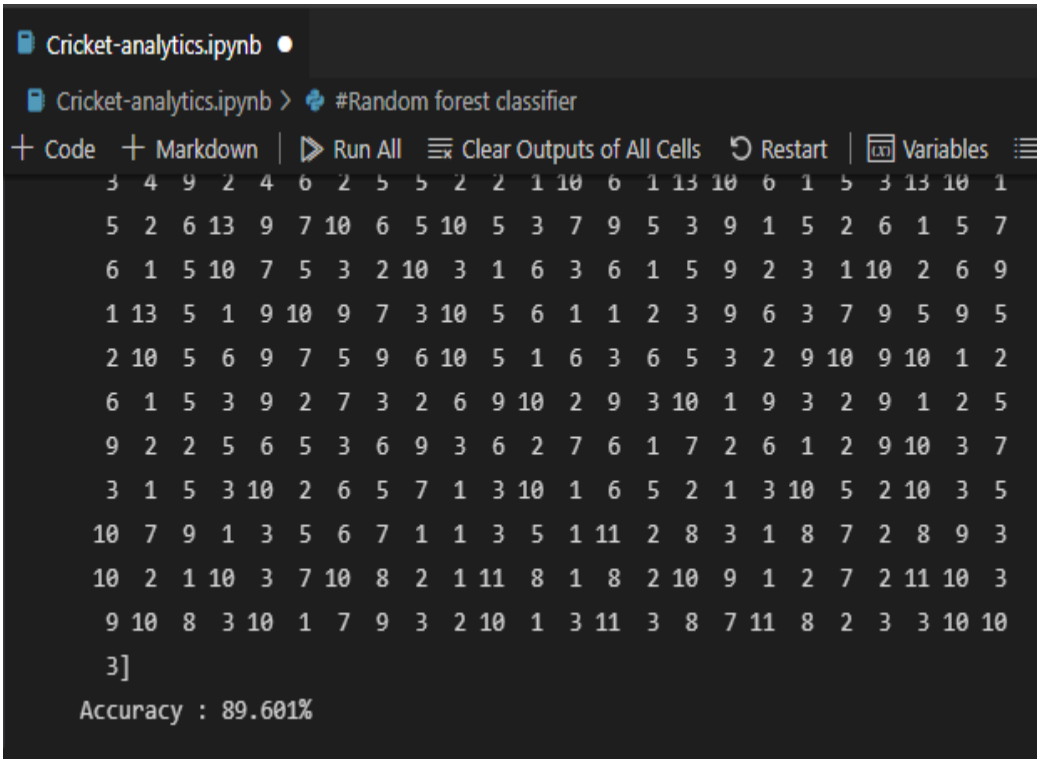
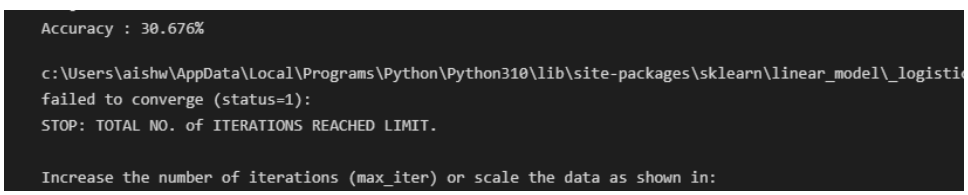


Figure 4.2 - Implementation of Random Forest model and its accuracy

4.3 Results at various stages :

4.3.1 Accuracy of different models -

1 - Logistic regression



2 - Gaussian NB

```
Accuracy : 20.624%

c:\Users\aishw\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\utils\validation.py:111
vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for exa
  y = column_or_1d(y, warn=True)
```

3 - KNeighborsClassifier

```
  9  9  1  9  3  7 10  7  2  1 10  7  1  6  2 10  7  2  1  7  2 10 10  3
  7 10  2  9 11  1  7  1  3  2 10  1  3 11  3  8  7  7  8  2  3  3  2 10
  3]
Accuracy : 62.218%
```

4 - SVM

```
 10  2  1 10  3  7 10  8  2  1 11  7  1  8  7 10  9  1  2  7  2 11 10  3
  9 10  8  3 10  1  7  9  3  2 10  1  3 11  3  8  7 11  8  2  3  3 10 10
  3]
Accuracy : 89.081%
```

5 - Decision Tree

```
  3]
Accuracy : 89.601%
```

4.3.2 result based on Toss -

1 - Graph between toss win and match win vs toss loss and match win as shown in figure 4.4

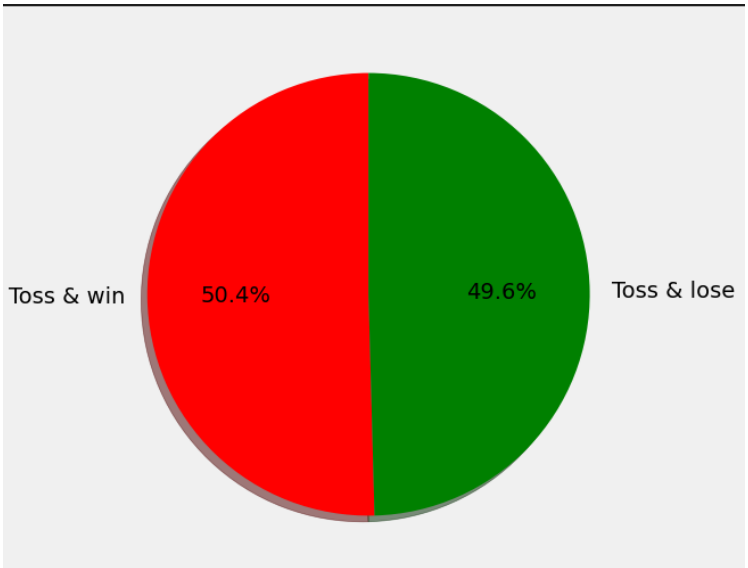
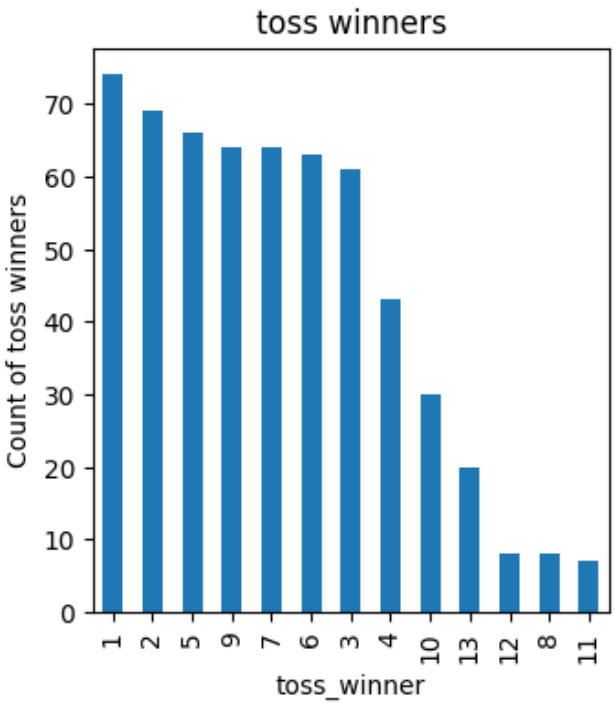
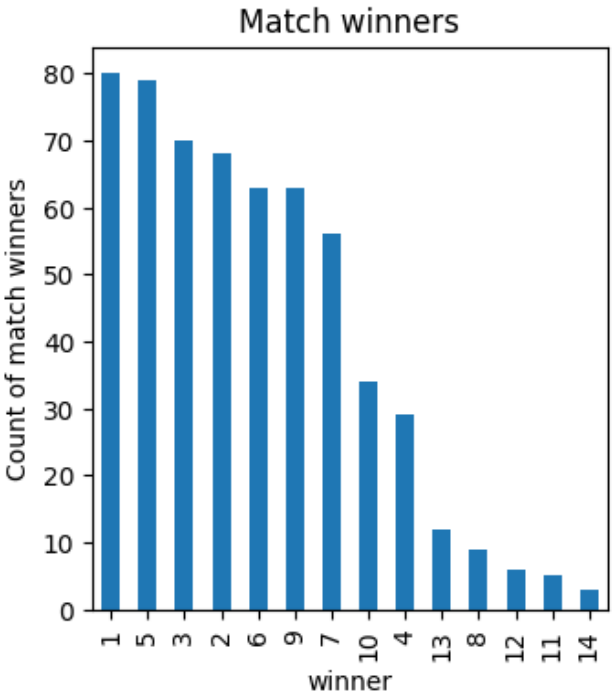


Figure 4.3 - Result based on toss

2 - Graph of toss vs match winners

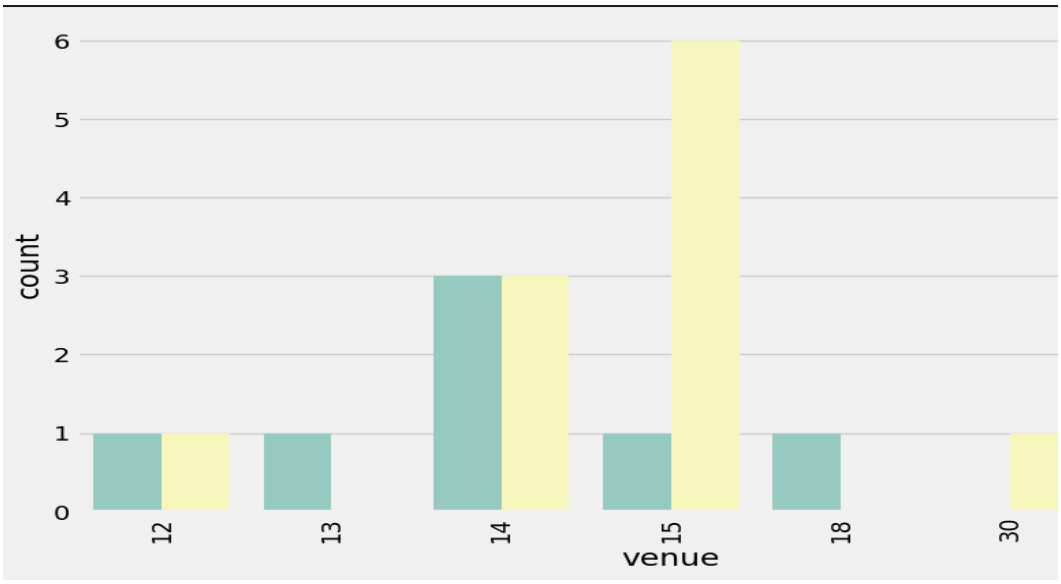


Graph 4.1 - Graph of toss winner



Graph 4.2 - Graph of match winner

4.3.3 Result based on venue



Graph 4.3 - Graph of match winner vs venue

4.4 Output in corresponding to input at various stages :

4.4.1 win prediction :

First we have predicted that who will win match based on inputs team 1 , team 2, venue, toss winning team , toss decision as seen from the figure below .

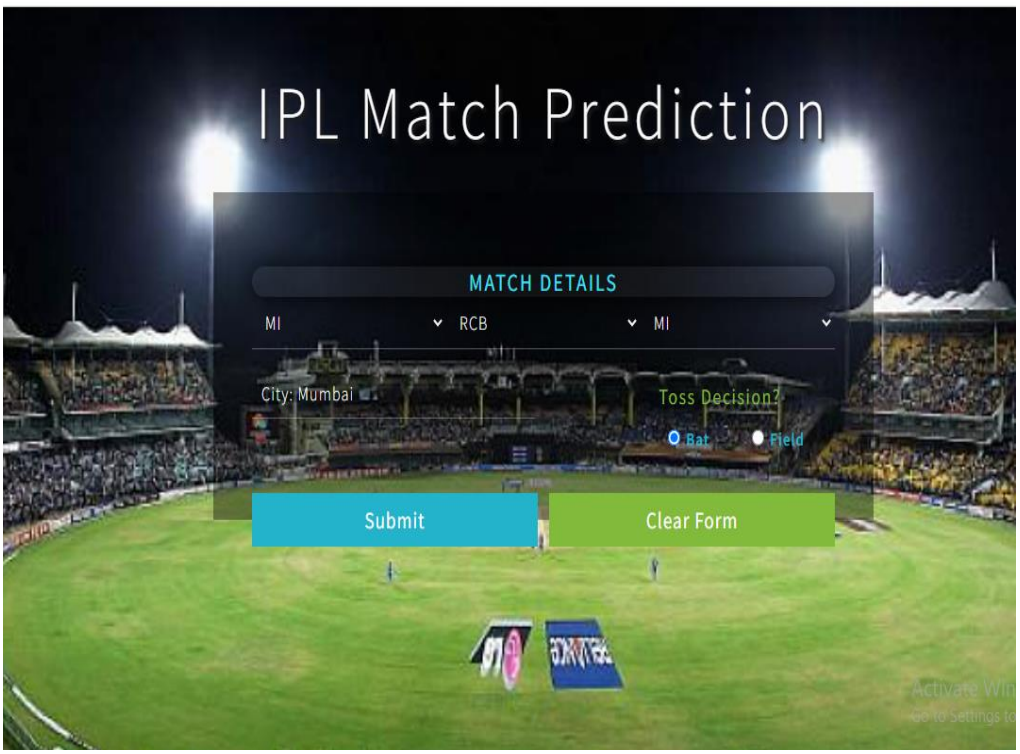


Figure - 4.4 UI of win prediction

After clicking of submit button it will predict who have high chance of winning match with 60% accuracy , we can also click on clearform button to click on the input of UI.output of below input is shown below at diagram

Figure - 4.5 output of win prediction

4.4.2 win percentage :

It calculates the winning percentage in second innings based on previous innings and after 5 overs in second innings , it has a accuracy of around 85%

As seen from the figure below it takes input as

- team 1
- team 2
- Venue
- Target
- Score
- Overs completed,
- Wickets fallen

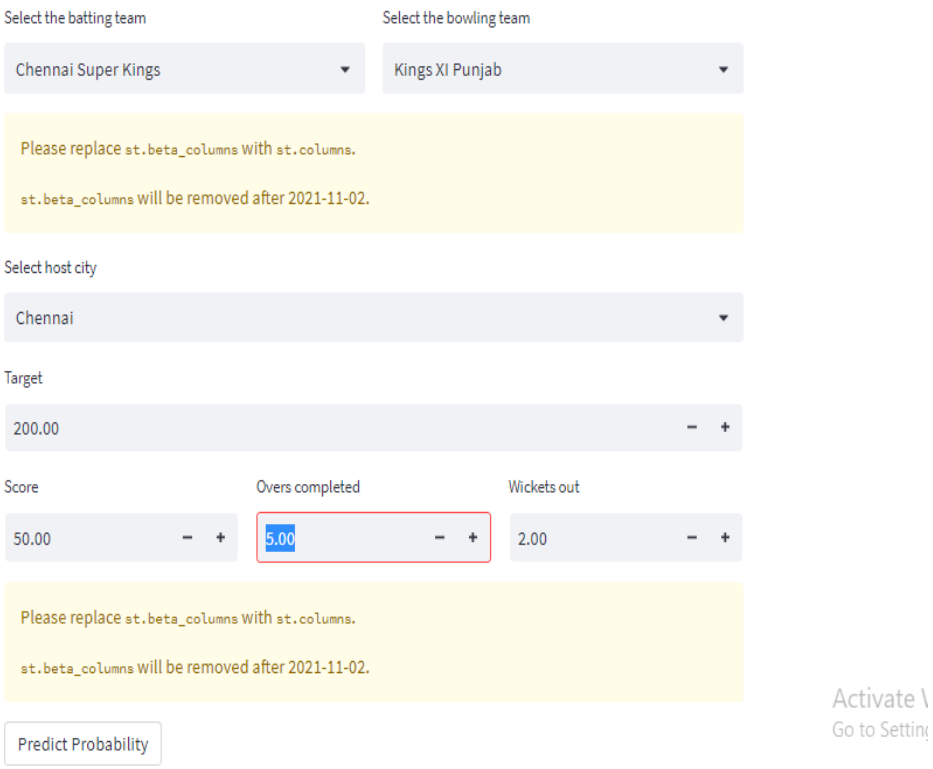


Figure - 4.6 UI of win percentage

It will give win percentage -

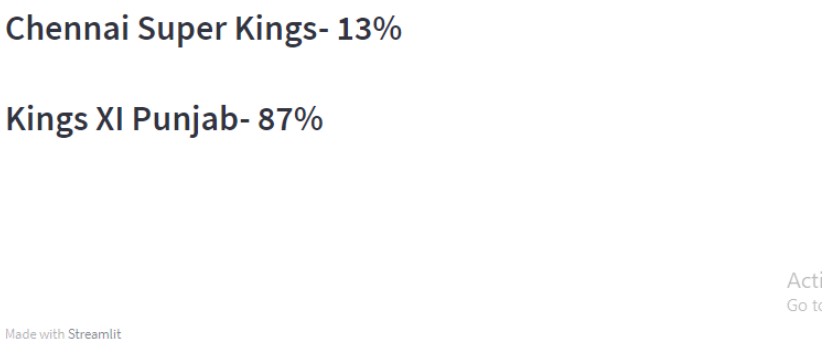


Figure 4.7 - Output of win percentage

4.4.3 First Inning score prediction :

It calculates the range of score after 5 overs that , how much runs will be scored by team after 5 overs till last overs. It calculates with the accuracy of 85% , It takes input as seen from the figure -

- team 1,
- team 2 ,
- overs completed must be greater than 5 ,
- runs scored till now ,
- wickets fallen till now ,
- runs scored in 5 overs,
- wickets fallen in 5 overs

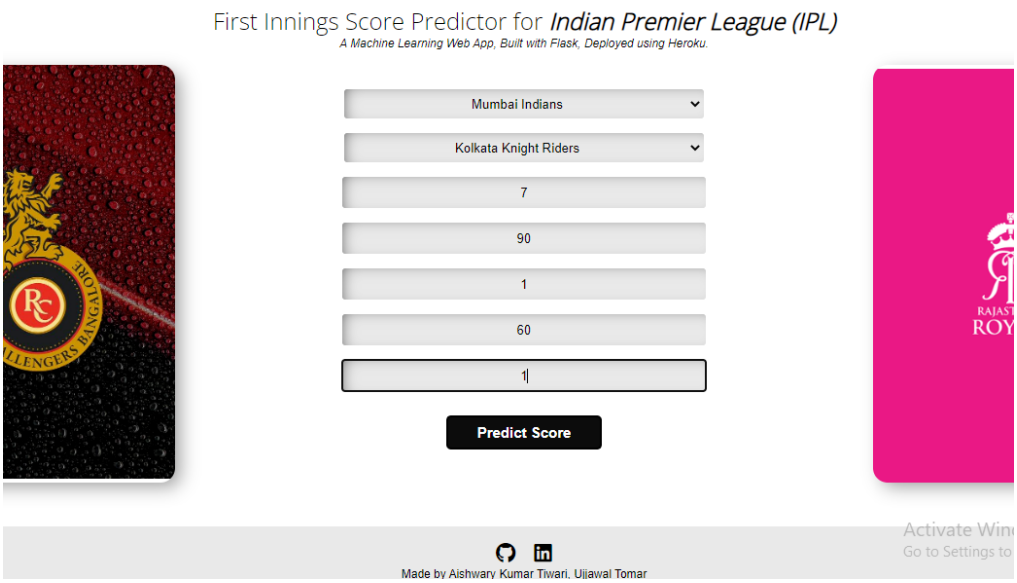


Figure - 4.8 UI of First Inning Score Prediction

Predicted Score Range -

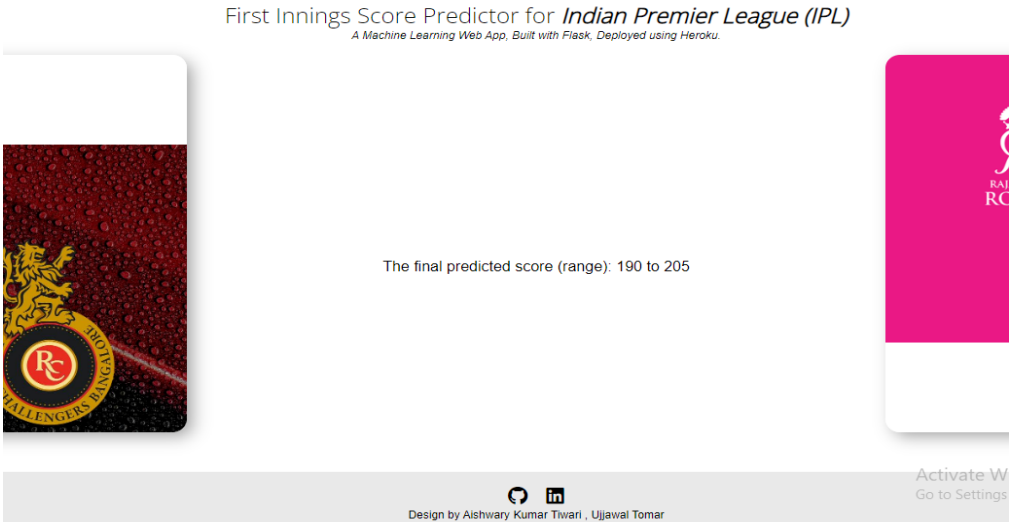


Figure - 4.9 Output of predicted score range

CHAPTER-5
CONCLUSIONS

5.1) Conclusions

This research tries to comprehend the dataset of the IPL data's previous ten-year history. Understanding the four main machine learning algorithms' underlying principles and how they are implemented in R is beneficial. It produces the Model and Training dataset and aids in prediction using the built-in model. The data is categorised by the model, and the outcomes are compared. The metrics accuracy, error rate, precision, recall, sensitivity, and specificity are taken into account. Based on this, Random Forest is chosen as the best algorithm. The primary goal of this study is to examine IPL data and offer its conclusions through graphical representation and comparative analysis. This allows the Indian Premier League and the fans to make judgements on the team's performance and forecast the winner of the trophy that will lead to success in future

With a 60% to 70% accuracy rate, the model that is used to forecast the outcomes of the matches was successfully created. The number of qualities was reduced to By utilising the attribute selection process, the dataset's 21 accessible attributes were reduced to 10 essential ones. Naive Bayes was one of four data mining methods that was used on the model. The random forest algorithm had the highest accuracy (71.08%). The total number of cases in the data set was 574 and no of classes were 11, and even if the accuracy was in the range of 60% and 70%, it was still low.

To find patterns in the data set and carry out an analysis, we require at least 100 occurrences of each type high accuracy forecast rate. For the dataset's 11

classes, a minimum of 1100 instances are required in order to execute prediction with a high rate of accuracy. Since there were 574 occurrences in the data set, accuracy may increase if there were more in the future. This is because a bigger number of examples will provide the model more freedom to draw stronger conclusions and spot more patterns in the data than a smaller number.

5.2 Future Scope

There are some future works that can be done in order to improve this project -

- The data set may include a few external variables, such as player injury, player fatigue, winning streaks with specific teams, overall winning streaks, average runs scored by a team against a specific team in prior matches, etc. Based on these variables, we can attempt to make predictions and monitor how accurately they turn out.

- Instead of relying solely on high level information about the various matches, such as the winner of the toss, the outcome of the toss, the home team, etc., the prediction can also take into account the performance of the players in the team, such as the total number of runs scored by a player in the tournament, the form guide of the player, the number of man of the match awards earned, etc.

- In this we can only predict about cricket , but in future we can also implement model for Football ,Hockey and rugby.

- We can also try new techniques and algorithm to improve accuracy of model.

- We will also design login system so that particular person can login and his history would be saved so that in future when he comes it becomes easier for him

- Write now we are predecting only twenty - twenty format of ipl because it is easy to imlement as it have less dataset , but in future we will be prediciting about different format of different sports , like International One - day and test , world cups , champions trophy and asia cups , same goes for hockey and talking about football we can predict about FIFA, Champions League, EPL, and NFL

- As we are seeing kabaddi is rising in India Pro kabaddi is becoming famous in India and many Indians liking this event so In future we will also predict about Pro Kabaddi by results from their previous database.

- We can also increase factor or variables which can improve the accuracy of the model , and Thus we can better predict with better accuracy of model . We will also take large amount of dataset of players from different leagues so that we can predict with more accuracy.
-

5.3 Applications Contributions

Recently, a variety of machine learning (ML) applications in sports have emerged.

There are studies on anticipating sports-related injuries (Van Eetvelde et al., 2021), implementing

Using forecasts for athletic training and predictive analytics (Rajsp & Fister, 2020)

deciding on the best team configurations (Ishi & Patil, 2021). There are numerous research that

concentrate on "in-play" forecasts, which foretell specific contestant events,

a goal being scored, for instance, by a certain player.

The use of ML to forecast the results of games in team sports, however, is the specific focus of the current survey.

In order to pinpoint the most crucial elements in producing successful outcomes, players, team management, and performance analysts are all interested in the prediction of final match results, in addition to bookies and bettors.

Choosing the optimal set of features for the predictive model is a crucial step in the outcome prediction of sporting events.

The most crucial predicted features can be extracted if the utilised model is interpretable and includes some feature selection mechanisms, or alternatively, if a feature selection approach is used before applying the model. In-play features may highlight areas where teams can modify their tactics or strategies to improve performance, even when some features, such as the match venue, officials, weather, etc., are external to the sport competition.

While match outcome prediction in sports has been discussed in academic literature for some time by the statistics and operations research communities, the use of ML approaches for this purpose is more recent. The first study in this field appears to have been published in 1996 (Purucker, 1996), although research in this field did not pick up until the 2010s, as seen in Figure 1. (b). The inclusion criterion for this review's consideration of research publications from the past three decades in this area is that they had to employ at least one machine learning (ML) approach. To that purpose, knowledge-based systems (such systems based on fuzzy logic and rules) and rat.

For instance, we investigate the possibility that the predictability of games may differ depending on whether they are invasion (time-dependent) sports or striking/fielding (innings-dependent) sports. We also talk about how the frequency and size of goals or points may have an impact on how predictable a match's outcome is. We also discuss some of the key factors that have been used in successful studies that have used machine learning, as well as how those factors have led to higher predictive accuracies

5.4 Maintenance And Support

This web app is a high big app and requires a lot of maintenance to improve user experience. This application will be kept up over time to meet needs in the future.Some maintenance method include:

5.4.1 Adaptive Maintenance

Every year, this application needs to be trained and the data updated to increase the accuracy of the predictions.

5.4.2 Perfective Maintenance

A more user-friendly interface can be created and altered frequently for users. The score line rather than win, lose, or draw might be used to forecast the outcome of the game

REFERENCE

[1] S. Lotfi and M. Rebbouj, "Machine Learning for sport results prediction using algorithms," *International Journal of Information Technology and Applied Sciences (IJITAS)*, vol. 3, no. 3, pp. 148–155, Aug. 2021, doi: https://doi.org/10.52502/ijitas.v3i3.114.

[2] R. P. Bunker and F. Thabtah, "A machine learning framework for sport result prediction," *Applied Computing and Informatics*, vol. 15, no. 1, pp. 27–33, Jan. 2019, doi: https://doi.org/10.1016/j.aci.2017.09.005.

[3] B. Singla, S. Taneja, R. Garg, and P. Nagrath, "Sport Result prediction : A comparative study," *Intelligent Decision Technologies*, pp. 1–14, Mar. 2022, doi: https://doi.org/10.3233/idt-210065.

[4] P. K. Jain, W. Quamer, and R. Pamula, "Sports result prediction using data mining techniques in comparison with base line model," *OPSEARCH*, Aug. 2020, doi: https://doi.org/10.1007/s12597-020-00470-9.

[5] "The use of machine learning in sport outcome prediction: A review," *WIREs Data Mining and Knowledge Discovery*, Jan. 2022, doi: https://doi.org/10.1002/widm.1445.

[6] Q. Zhou, "Sports Achievement Prediction and Influencing Factors Analysis Combined with Deep Learning Model," *Scientific Programming*, vol. 2022, pp. 1–8, Jan. 2022, doi: https://doi.org/10.1155/2022/3547703.

[7] P. K. Jain, W. Quamer, and R. Pamula, "Sports result prediction using data mining techniques in comparison with base line model," *OPSEARCH*, Aug. 2020, doi: https://doi.org/10.1007/s12597-020-00470-9.

[8] N. Patil, B. H. Sequeira, N. N. Gonsalves, and A. A. Singh, "Cricket Team Prediction Using Machine Learning Techniques," *SSRN Electronic Journal*, 2020, doi: https://doi.org/10.2139/ssrn.3572740