

**A
Project Report
On
"IPL Winner Prediction"**



Prepared by
Jethloja Rohan Rajeshbhai (17CE037)

Under the guidance of
Prof. Minal Shah

Submitted to
Charotar University of Science & Technology
Degree of Bachelor of Technology
in Computer Engineering
CE448 : Software Group Project-V
Of 7th Semester of B.Tech

Submitted at



U. & P. U PATEL DEPARTMENT OF COMPUTER ENGINEERING
Faculty of Technology & Engineering, CHARUSAT
Chandubhai S. Patel Institute of Technology
At: Changa, Dist: Anand – 388421
October 2020



CHARUSAT
CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

CERTIFICATE

This is to certify that the report entitled "**IPL Winner Prediction**" is a bonafied work carried out by **Jethloja Rohan Rajeshbhai (17CE037)** under the guidance and supervision of **Prof. Minal Shah** for the subject **Software Group Project - V (CE448)** of 7th Semester of Bachelor of Technology in **Computer Engineering** at Faculty of Technology & Engineering (C.S.P.I.T.) – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

Under the supervision of,

Prof. Minal Shah
Assistant Professor
U. & P. U Patel Dept. of Computer Engg.
C.S.P.I.T., CHARUSAT-Changa.

Dr. Ritesh Patel
Head,
U. & P. U Patel Dept. of Computer Engg.
C.S.P.I.T., CHARUSAT-Changa.

Chandubhai S Patel Institute of Technology (C.S.P.I.T.)
Faculty of Technology & Engineering, CHARUSAT

At: Changa, Ta. Petlad, Dist. Anand, PIN: 388 421. Gujarat

ABSTRACT

This project “IPL Winner Prediction” basically provides solution to predict the winner of the particular match. Cricket is one of the most popular team games in the world. Cricket is the second most watched sport in the world after soccer, and enjoys a multi-million-dollar industry. There is remarkable interest in simulating cricket and more importantly in predicting the outcome of cricket match which is played in three formats namely test match, one day international and T20 match. The complex rules prevailing in the game, along with the various natural parameters affecting the outcome of a cricket match present significant challenges for accurate prediction. Several diverse parameters, including but not limited to cricketing skills and performances, match venues and even weather conditions can significantly affect the outcome of a game. In this project embark on predicting the outcome of a IPL (Indian Premier League) cricket match using a supervised learning approach from a team composition perspective.

ACKNOWLEDGEMENT

If words are considered as a symbol of approval and token of appreciation then let the words play the heralding role expressing our gratitude. The satisfaction that accompanies that the successful completion of any task would be incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all efforts with success. I am grateful to my project coordinator **Prof Minal Shah** for the guidance, inspiration and constructive suggestions that help us in the preparation of this project.

I am really grateful to our Head of Department Dr. **Ritesh Patel**, for providing us with an opportunity to showcase our skills and talent by providing us with all the facilities needed for our project development. I would like to express our sincere gratitude towards our parents & faculties of Charusat University, for their kind co-operation and encouragement which helped us in completion of this project.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped us out with their abilities.

Table of Contents

Abstract	iii
Acknowledgement	iv
List of Figures	vii
List of Tables	viii
Chapter 1 Introduction	1
1.1 Project Summary	1
1.2 Project Purpose	1
1.3 Project Scope	1
1.4 Project Objective	2
1.5 Technology and literature survey	2
Chapter 2 Project Management	4
2.1 Project Planning	4
2.1.1 Project Development Approach and Justification	4
2.1.2 Roles and Responsibilities	5
2.2 Project Scheduling	5
Chapter 3 System Requirements Study	6
3.1 User Characteristics	6
3.1 Hardware and Software Requirements	6
3.3 Assumptions and Dependencies	6
Chapter 4 System Analysis	8
4.1 Study of Current System	8
4.2 Problems and Weaknesses of Current System	8
4.3 Requirements of New System	8
4.3.1 Functional Requirements	8
4.3.2 Non-Functional Requirements	9
4.4 Use Case Diagram	9
4.5 Class Diagram	10
4.6 Flowchart of system	10
4.7 Sequence Diagram	11
Chapter 5 System Design	12
5.1 Design Phases	12
5.1.1 Data Pre-processing	12

5.1.2 Model Generation and its testing	12
5.1.3 GUI Development for better user interaction	13
5.2 User Interface Snapshots	13
Chapter 6 Implementation Planning	15
6.1 Implementation Environment	15
6.2 Program/Modules Specification	15
Chapter 7 System Testing	16
Chapter 8 Conclusion and Discussion	17
8.1 Self Analysis of Project Viabilities	17
8.2 Summary of Project Work	17
8.3 Social Benefit	17
Chapter 9 Bibliography	18

List of Figures

Figure 2.1 Iterative waterfall model	4
Figure 2.2 Work Breakdown Structure	5
Figure 4.1 Use case diagram	9
Figure 4.2 Class diagram	10
Figure 4.3 Flowchart	10
Figure 4.4 Sequence diagram	11
Figure 5.1 Home Page	13
Figure 5.2 Predicted Output Page	14

List of Tables

Table 2.1 Roles and responsibility	5
Table 7.1 Input form test suite	16

CHAPTER 1.0

INTRODUCTION

1.1 PROJECT SUMMARY

This project “IPL Winner Prediction” is a web application which intends to predict the winner of match using various factor affecting cricket. Statistical modeling has been used in sports since decades and has contributed significantly to the success on field. Various natural factors affecting the game, enormous media coverage, and a huge betting market have given strong incentives to model the game from various perspectives. However, the complex rules governing the game, the ability of players and their performances on a given day, and various other natural parameters play an integral role in affecting the final outcome of a cricket match. This presents significant challenges in predicting the accurate results of a game. In this project used toss decision and the venue of the match, along with the relative team strength, and adopt supervised learning algorithms to predict the winner of the match.

1.2 PROJECT PURPOSE

Statistical modeling has been used in sports since decades and has contributed significantly to the success on field. Cricket is one of the most popular sports in the world. This project is used to predict the winner of match using various factor like venue, toss discussion, and previous records of particular team against particular team.

1.3 PROJECT SCOPE

The goal of the project is to predict the result of the match beforehand by taking into consideration the data and the results of the previous matches played between the two teams. The software is quite simple to use and can be used by any person. This project is also useful for management to take decision of team and toss.

1.4 PROJECT OBJECTIVE

- To predict the results of the cricket match by not only taking factors like toss, venue, but also by considering factors like team composition, the batting and bowling averages of each player in the team, the performance of the team in their past few matches, the probability of winning by batting first at a specific venue against a specific team
- To assist the team management for taking decision of toss like bet first or field first.

1.5 TECHNOLOGY AND LITERATURE REVIEW

A Kaluarachchi, SV Aparna (2010).A classification based tool to predict the outcome in cricket.

According to this factors contributing to winning games are imperative, as the ultimate objective in a game is victory. The aim of this study was to identify the factors that characterize the game of cricket, and to investigate the factors that truly influence the result of a game using the data collected from the Champions Trophy cricket tournament. According to the results, this cricket tournament can be characterized using the factors of batting, bowling, and decision-making. Further investigation suggests that the rank of the team and the number of runs they score have the most significant influence on the result of games.

Tejinder Singh, Vishal Singla, Parteek Bhatia (2015).Score and Winning Prediction in Cricket through Data Mining

In this a model has been proposed that has two methods, first predicts the score of first innings not only on the basis of current run rate but also considers number of wickets fallen, venue of the match and batting team. The second method predicts the outcome of the match in the second innings considering the same attributes as of the former method along with the target given to the batting team. These two methods have been implemented using Linear Regression Classifier and Naive Bayes Classifier for first innings and second innings respectively. In both methods, 5 over intervals have been made from 50 overs of the match and at each interval above mentioned attributes have been recorded of all non-curtailed matches played between 2002 and 2014 of every team independently. It has been found in the results that error in Linear Regression classifier is less than Current Run Rate

method in estimating the final score and also accuracy of Naive Bayes in predicting match outcome has been 68%.

Sasank Viswanadha, Kaustubh Sivalenka, Madan Gopal Jhawar, Vikram Pudi(2006). **Dynamic Winner Prediction in Twenty20 Cricket: Based on Relative Team Strengths.** proposes a model to predict the winner at the end of each over in the second innings of an IPL cricket match. Our methodology not only incorporates the dynamically updating game context as the game progresses, but also includes the relative strength between the two teams playing the match. Estimating the relative strength between two teams involves modeling the individual participating players' potentials. To model a player, we use his career as well as recent performance statistics. Using the various dynamic features, we evaluate several supervised learning algorithms to predict the winner of the match. Finally, using the Random Forest Classifier (RFC), we have achieved an accuracy of 65.79% - 84.15% over the course of second innings, with an overall accuracy of 75.68%.

CHAPTER 2.0

PROJECT MANAGEMENT

2.1 PROJECT PLANNING

I have divided the project into individual modules which have unique constraint limited functionalities which will be further explained. The estimated project development period has been 3 months. The project team consists of 1 key members.

2.1.1 Project Development Approach and Justification

For the project development the approach I have made was using Iterative waterfall model. In iterative development, first of all feature code is designed, developed and then tested in repeated cycles. With each iteration additional features can be designed, developed and tested until there is a fully functional product which is ready to use.

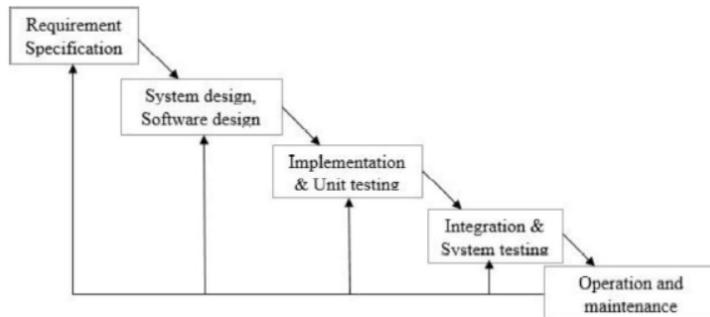


Figure 2.1 - Iterative waterfall model

The project IPL Winner Prediction is used Machine learning algorithms for predict the winner of match using previous data. Front end design is done using HTML, CSS, JavaScript and jquery. I used Flask framework for the actual backend system. For this project I have to use various machine learning algorithm and fand the best algorithm for this project. I needed to run and test that particular module and then respective changes were to be made and new features were to be added. I found Iterative waterfall model to be the best to follow for my project.

2.1.2 Roles and Responsibilities

Student ID	Student Name	Role
17CE037	Rohan Jethloja	Full stack Developer

Table 2.1 Roles and Responsibilities

2.2 PROJECT SCHEDULING

The Project Scheduling will be initially explained by the help of WBS (Work Breakdown Structure). The Work Breakdown Structure is a hierarchical reflection of all the work in the project in terms of deliverables. In order to produce these deliverables work must be done. A typical approach in developing a WBS is to start at the highest level, with the product of the product of the project

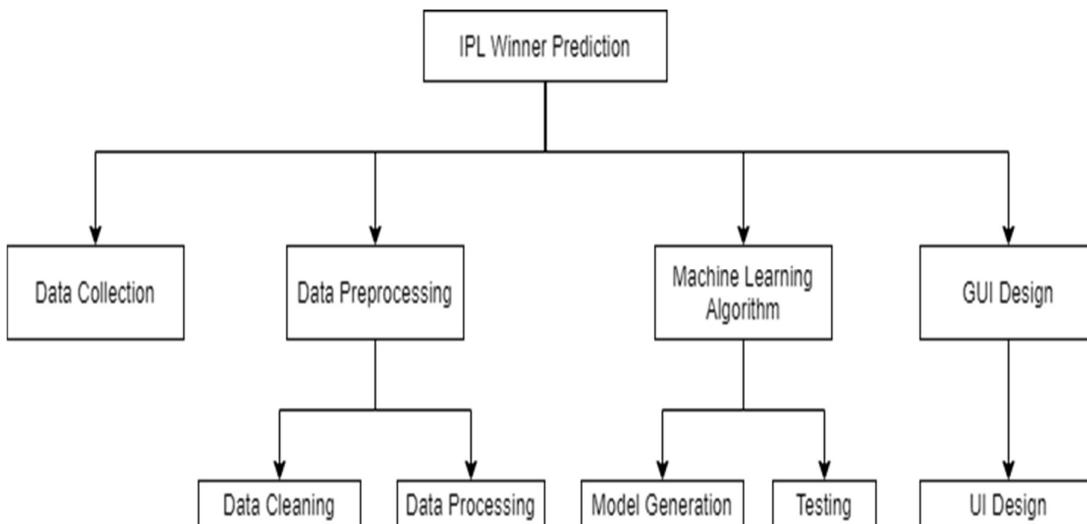


Figure 2.2 – Work Breakdown Structure

CHAPTER 3.0

SYSTEM REQUIREMENTS STUDY

3.1 USER CHARACTERISTICS

This project is useful for predict the winner of IPL cricket match using supervised learning algorithm. User can directly enter the match details like team1, team2, home team, toss win, and toss decision using all this details algorithm predict the winner of match which team has higher chance of winning the match.

3.2 HARDWARE AND SOFTWARE REQUIREMENTS

- Operating System – Windows XP and above
- Hardware Requirements – Pentium IV or more, RAM – 512 or more (Though core i3 or above is recommended if results are required fast)
- Storage Requirements – A hard drive containing the dataset and the software code itself.

3.3 ASSUMPTIONS AND DEPENDENCIES

- Accuracy: all the fields of the input form which is used to take predicting match details are required else prediction may go wrong.
- Performance: the hardware and software requirements mentioned in the operating environment should be present for proper functioning of the software.
- Language: English is kept as the only displayed language.
- It is assumed that the python (3.0) is installed in the system alongside the following libraries -
 - Flask==1.0.2
 - beautifulsoup4==4.6.3
 - pandas==0.23.4
 - Requests==2.20.0
 - scikit_learn==0.19.2
 - gunicorn

- There should be a System administrator whose job is to keep the dataset updated before a user runs the program.
- The system should have sufficient processor speed so as to train the huge amount of dataset on different machine learning algorithms.

CHAPTER 4.0

SYSTEM ANALYSIS

4.1 STUDY OF CURRENT SYSTEM

The current system has been studied before the development began. Firstly, the current functionalities were observed and thoughts were given to their respective extensions. In current system there is some project available which predict the outcome for the particular tournament only.

4.2 PROBLEM AND WEAKNESSES OF CURRENT SYSTEM

The current system is good for some specific tournament only also the prediction accuracy is also low for some time so chance of the prediction goes right is very low. The current system also uses only past result for predicting the winner for particular match. So, there is chance that teams may be change so the outcome will be different.

4.3 REQUIREMENTS OF NEW SYSTEM

Nowadays, Cricket is the most popular sport in India. There is remarkable interest in simulating cricket and more importantly in predicting the outcome of cricket match, we have to use high accuracy algorithm for predict the outcome of the match because it is very useful for the team management and also in the field of fantasy sport, there is so many fantasy cricket league is for this we have to do data mining manually, so using this project the task is done by machine automatically and we get the result in few minutes without any efforts.

4.3.1 Functional Requirements

Requirement 1: Input Form.

I/P: User enter details like home team, away team, toss winner, and toss decision.

O/P: predicted winner between this two team.

Requirement 2: Change of dataset

I/P: After every match dataset need to update with new value.

O/P: Updated data for prediction.

4.3.2 Non-Functional Requirements

Performance Requirements: To generate the result with optimized efficiency and process it faster, a fast processor will be required so as to train the machine learning model quickly.

Safety Requirements: The software is an attempt to predict the result with maximum accuracy possible. Since the result is based on the current player performances, how the teams will play and probabilities from the test and train data, so 100% accuracy is impossible to achieve. Since the possibility of winning a game turns with every moment, which is what this software aims to predict accurately, the results can sometimes be not as the true outcome. The software developing company hold no responsibility if any harm or loss occurs due to the use of software results.

Security Requirements: The System Administrator will be responsible for keeping the database secured for access by only certain individuals. Any malicious link, or advertisements will not be shown and if it does, users are requested to click them at their own risk as they are a part of user's client app.

4.4 USE CASE DIAGRAM

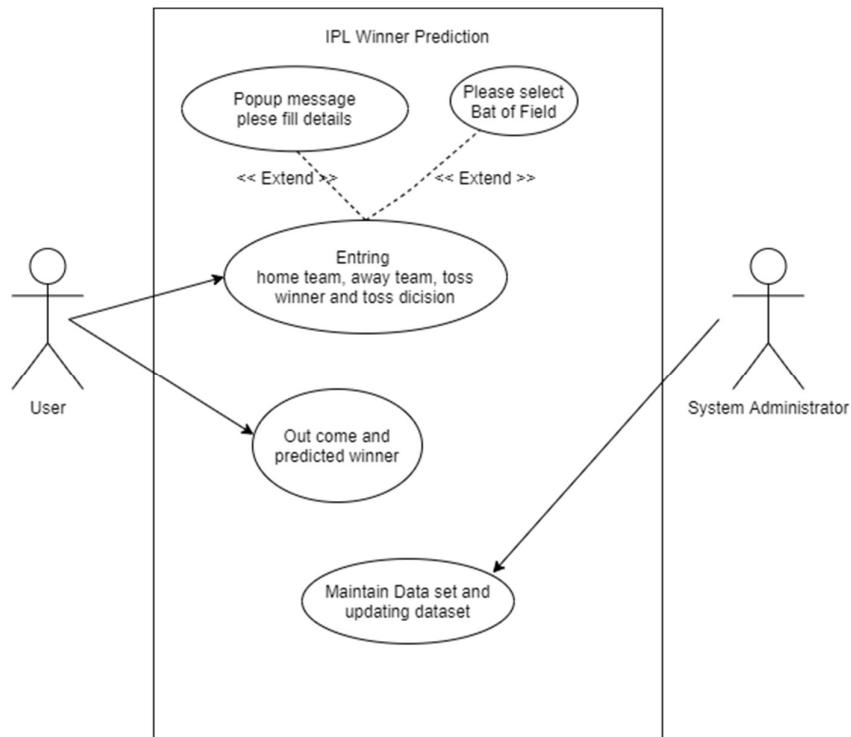


Fig 4.1 Use Case Diagram

4.5 CLASS DIAGRAM

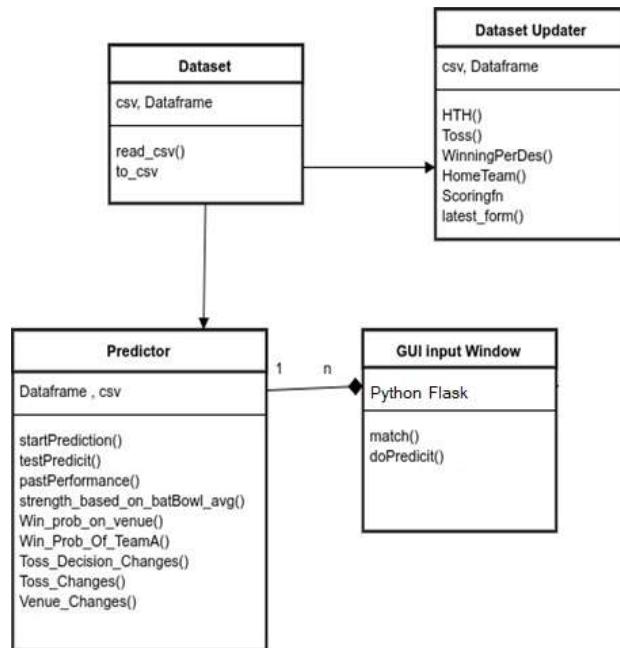


Fig 4.2 Class Diagram

4.6 FLOWCHART OF SYSTEM

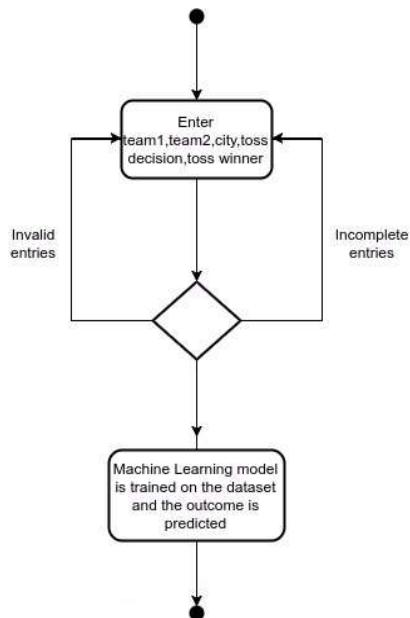


Fig 4.3 Flowchart

4.7 SEQUENCE DIAGRAM

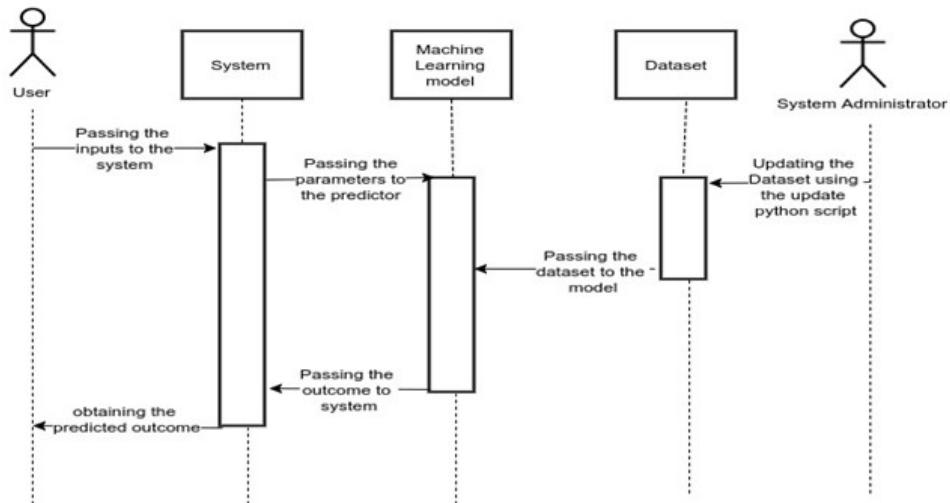


Fig 4.4 Sequence Diagram

CHAPTER 5.0

SYSTEM DESIGN

5.1 DESIGN PHASES

The whole project is divided into 3 phases.

- Data Preprocessing
- Model Generation and its testing
- GUI Development for better user interaction.

5.1.1 Data Pre-processing

This stage involves fetching data from web and then processing it to make it suitable for modelling. The data I obtained is in the form of csv files where each file describes each match details. The details that single csv file constitutes are match id, team names, toss winner, toss decision taken after winning, venue of the match, date of the match, scores of each team after every over. This data also contains few missing entries. With the help of python, I first parsed all the files and stored the data in a single file. While parsing I have fetched only required attributes from each file rather than storing whole contents of file. Now the resultant file each row corresponds to details of each match. After parsing I handled missing values by taking mean of that particular feature for which the value is missing.

5.1.2 Model Generation and its testing

After pre-processing I used the machine learning to generate models which can be used to predict the results. I used different machine learning models and calculated the accuracy from each model and finally selected the one with best accuracy. I used decision tree, random forest, logistic regression and SVM. Out of all these I used logistic regression as it gave us the better accuracy among all other models. While generating the models I used K Fold cross validation technique so as to make use of whole dataset as training as well as testing. For future match prediction I first take the inputs from user via GUI which is created using python flask framework. Then I first convert those inputs in numeric form like 1 for team A and 0 for team B and generate other important features using the modules I have

used in phase 2. Then we convert these whole inputs into panda's data frame and pass it to model for prediction. The model does the processing and gives us the winner of the match.

5.1.3 GUI Development for better user interaction

In this phase I developed a simple user interface so that a user need not to go to command prompt for entering all the required inputs. The interface has been developed using python framework flask. The interface contains a single form where user can enter some match details. The form developed also validates if the provided input is valid or not. If the entered information is invalid it returns an error message. Also, if the user doesn't give all the required inputs then it displays an error message to fill all the required entries. Thus, it is easy and better way of getting all the valid inputs.

5.2 USER INTERFACE SNAPSHOTS



Figure 5.1 Home page



Figure 5.2 Predicted Output Page

CHAPTER 6.0

IMPLEMENTATION PLANNING

6.1 IMPLEMENTATION ENVIRONMENT

Hardware:

- Model: Dell Inspiron 15 3000
- Processor: Intel Core i5 7th Generation
- Installed Memory (RAM)- 8.00 GB
- System type- 64-bit Operating System, x64 based processor

Software:

- PyCharm 2019
- Flask 1.0.2
- Technology- HTML, CSS, JavaScript, Jquery, Bootstrap and Python

6.2 MODULE SPECIFICATION

Module 1: Data preprocessing

This stage involves fetching data from web and then processing it to make it suitable for modelling.

Module 2: Model generation and testing

In this module used the machine learning to generate models which can be used to predict the results. I used different machine learning models and calculated the accuracy from each model and finally selected the one with best accuracy.

Module 3: GUI Design

In this module interface has been developed using python framework flask. The interface contains a single form where user can enter some match details. Thus, it is easy and better way of getting all the valid inputs.

CHAPTER 7.0

SYSTEM TESTING

- Before actually implementing the new system/software into operations, a test run of the system/software is done removing all the bugs, if any. It is an import Phase of a successful system. In the system/software testing stage, we check the overall behavior of the system/software against the functional and the performance requirements. There are two type of testing:
- White – box: Internal part (code) of the project is tested.
- Black -- box: System behavior (input/output) is checked.

Test Suite 1: Registration

Sr. No.	Test case	Test data	Expected result	Actual result	Status (Pass/Fail)
1.	User can submit data	Home Team	Select any one team	Show selected team	Pass
2.	User can submit data	Away team	Select any one team	Show selected team	Pass
3.	User can submit data	Home team & Away team same	Please select different team	Popup message shown	Pass
4.	User can submit data	Venue	Ground selected of home team	Show home teams' ground	Pass
5.	User can submit data	Toss winner	Select 1 toss winner team	1 toss winner is selected	Pass
6.	User can submit data	Toss decision	Any one is selected	One check box is selected	Pass

Table 7.1 Input form test suite

CHAPTER 8.0

CONCLUSION

8.1 SELF ANALYSIS OF PROJECT VIABILITIES

There is always a room for improvements in any machine learning algorithm however good and efficient it may be done but the most important thing should be flexible to accept further modifications. Right now, I am just dealing with few features only but in future this web application may be extended to new advance features. I am using more future to increase the accuracy of the prediction and give accurate output to user.

8.2 SUMMERY OF PROJECT WORK

- The outcome of the project is abiding by the rules and requirements specified before the development of the project.
- The project also contributes to the overall objectives of the organization.
- The system can also be integrated with older systems and newer extensions can also be added in future.

8.3 SOCIAL BENEFIT

- Cricket enthusiasts can know who has the higher chances of winning the match even before the match starts.
- It also assists team captains in deciding whether to choose bat or field at a specific stadium if they win the toss.

CHAPTER 9.0

BIBLIOGRAPHY

- Tejinder Singh, Vishal Singla, Parteek Bhatia (2015).Score and Winning Prediction in Cricket through Data Mining
- Munir, Fahad, Hasan, Md. Kamrul, Ahmed, Sakib, Md. Quraish, Sultan (2015). Predicting a T20 cricket match result while the match is in progress.
- Sasank Viswanadha, Kaustubh Sivalenka, Madan Gopal Jhawar,Vikram Pudi(2006). Dynamic Winner Prediction in Twenty20 Cricket: Based on Relative Team Strengths.
- Monalisha Pattnaik, Anima Bag (2015). Fitting of Logistic Regression Model for Prediction of Likelihood of India Winning or Losing in Cricket Match.
- A Kaluarachchi, SV Aparna (2010).A classification based tool to predict the outcome in cricket.