

Abstract

Cricket prediction can be viewed as one of the objectives of sports analytics, which aims at helping decision makers to gain competitive advantage. Data Analysis is becoming more common specially in Sports. Using data analysis results has become familiar in sports organization such as International Cricket Council (ICC), International Federation of Association Football (FIFA), GrandSlam of the International Tennis Federation. The difficulty of this task depends on many factors, like the availability of data for the past events, the ability to gather data for future events, the knowledge needed to interpret gathered data, and others. In sports like Cricket and Football predicting the results become more difficult due to the changing nature of the Game. People focus on sports results, which is an important aspect, and the result of the game has become the focus and concentration of sports game. Various techniques for modelling a cricket match exist that yield different result prediction algorithms. The modelling can be put under the four generic categories which are empirical models, dynamic systems, statistical techniques, artificial intelligence (including expert systems). In the artificial intelligence category, there are several approaches that focus on Bayesian network modelling. The Matrix factorization technique became very popular in the field of multimedia content recommender systems where it showed good scalability and predictive accuracy. The idea behind using the latent features in our case is to be able to build a successful model. The outcome of this project will be a system which will predict the results of the upcoming matches.

Contents

1	Synopsis	1
1.1	Project Title	2
1.2	Project Option	2
1.3	Internal Guide	2
1.4	Sponsorship and External Guide	2
1.5	Technical Keywords	2
1.6	Problem Statement	2
1.7	Abstract	3
1.8	Goals and Objectives	3
1.9	Relevant mathematics associated with the Project	4
1.10	Names of Conferences / Journals	5
1.11	Plan of Project Execution	6
2	Technical Keywords	7
2.1	Area of Project	8
2.2	Technical Keywords	8
3	Introduction	9
3.1	Project Idea	10
3.2	Motivation of the Project	10
3.3	Literature Survey	10
4	Problem Definition and scope	14
4.1	Problem Statement	15
4.1.1	Goals and objectives	15
4.1.2	Statement of scope	15
4.2	Major Constraints	15
4.3	Methodologies of Problem solving and efficiency issues	15
4.4	Outcome	16
4.5	Applications	16
4.6	Hardware Resources Required	16
4.7	Software Resources Required	17

5 Project Plan	18
5.1 Project Estimates	19
5.1.1 Reconciled Estimates	20
5.1.2 Project Resources	20
5.2 Risk Management w.r.t. NP Hard analysis	21
5.2.1 Risk Analysis	22
5.2.2 Overview of Risk Mitigation, Monitoring, Management	23
5.3 Project Schedule	24
5.3.1 Project task set	24
5.3.2 Task network	24
5.3.3 Timeline Chart	25
5.4 Team Organization	26
5.4.1 Team structure	26
5.4.2 Management reporting and communication	26
6 Software Requirement Specification	27
6.1 Introduction	28
6.1.1 Purpose and Scope of Document	28
6.1.2 Overview of responsibilities of Developer	28
6.2 Usage Scenario	29
6.2.1 User profiles	29
6.2.2 Use-cases	29
6.2.3 Use Case View	30
6.3 Data Model and Description	30
6.3.1 Data Description	30
6.3.2 Data objects and Relationships	31
6.4 Functional Model and Description	31
6.5 Non-functional Requirement	31
6.5.1 Performance Requirement :	31
6.5.2 Safety and Security Requirement:	31
6.6 UML Diagrams:	31
6.6.1 Data Flow Diagram	31
6.6.2 Activity Diagram:	35
6.6.3 Design Constraints	36
6.6.4 Software Interface Description	36
7 Detailed Design Document	37
7.1 Introduction	38
7.2 Architectural Design	38
7.3 Data design	39
7.3.1 Internal software data structure	39

7.3.2	Global data structure	39
7.3.3	Temporary data structure	40
7.3.4	Database description	40
8	Project Implementation	41
8.1	Introduction	42
8.2	Tools and Technologies Used	42
8.3	Hardware Resources Required	43
8.4	Software Resources Required	43
8.5	Methodologies/Algorithm Details	43
8.5.1	Algorithm 1/Pseudo Code	44
8.6	Verification and Validation for Acceptance	46
9	Software Testing	47
9.1	Type of Testing Used	48
9.2	Test Cases and Test Results	49
10	Results	50
10.1	Output Screenshots	51
11	Deployment and Maintenance	55
11.1	Installation and un-installation	56
12	Conclusion and Future Scope	57
References		58
13	References	59
Annexure A	Laboratory assignments on Project Analysis of Algorithmic Design	62
Annexure B	Laboratory assignments on Project Quality and Reliability Testing of Project Design	65
Annexure C	Project Planner	67
C.0.1	Timeline Chart	69
Annexure D	Reviewers Comments of Paper Submitted	70
Annexure E	Plagiarism Report	72
Annexure F	Term-II Project Laboratory Assignments	74

Annexure G Information of Project Group Members **76**

List of Figures

1.1	Project Planner	6
5.1	Waterfall Model	19
5.2	NP Complete	21
5.3	Task network	25
5.4	Timeline chart	25
6.1	Use Cases	29
6.2	Use case diagram	30
6.3	DFD0	32
6.4	DFD1	33
6.5	DFD2	34
6.6	Activity diagram	35
7.1	Architecture diagram	38
9.1	Test Cases	49
10.1	Home Page	51
10.2	Selection of Stadium	51
10.3	Selection of First Team Players	52
10.4	Confirmation Message	52
10.5	Selection of Second Team Players	53
10.6	Toss Selection	53
10.7	Predicted Score	54
10.8	Statistics	54
A.1	IDEA Diagram	63
C.1	Project Planner	68
C.2	Task network	69

List of Tables

5.1	Risk Table	22
5.2	Risk Probability definitions	22
5.3	Risk Impact definitions	22
5.4	Table 1	23
5.5	Table 2	23
5.6	Table 3	24
5.7	Team Structure	26

CHAPTER 1

SYNOPSIS

1.1 Project Title

Winning Prediction in One Day International Cricket using Machine Learning Techniques.

1.2 Project Option

Internal Project

1.3 Internal Guide

Prof. S. V. Limkar

1.4 Sponsorship and External Guide

SAAPTHAHIK DAUND NAVANIRMAN NEWSPAPER

Haresh Hasmukh Rambhiya

1.5 Technical Keywords

- Machine Learning
- Neural Network
- Decision Tree
- Clustering
- Linear Regression
- Client Server Architecture

1.6 Problem Statement

To develop a system which will predict the win or lose percentage of a cricket match played between two teams.

1.7 Abstract

Cricket prediction can be viewed as one of the objectives of sports analytics, which aims at helping decision makers to gain competitive advantage. Data Analysis is becoming more common specially in Sports. Using data analysis results has become familiar in sports organization such as International Cricket Council (ICC), International Federation of Association Football (FIFA), GrandSlam of the International Tennis Federation. The difficulty of this task depends on many factors, like the availability of data for the past events, the ability to gather data for future events, the knowledge needed to interpret gathered data, and others. In sports like Cricket and Football predicting the results become more difficult due to the changing nature of the Game. People focus on sports results, which is an important aspect, and the result of the game has become the focus and concentration of sports game. Various techniques for modelling a cricket match exist that yield different result prediction algorithms. The modelling can be put under the four generic categories which are empirical models, dynamic systems, statistical techniques, artificial intelligence (including expert systems). In the artificial intelligence category, there are several approaches that focus on Bayesian network modelling. The Matrix factorization technique became very popular in the field of multimedia content recommender systems where it showed good scalability and predictive accuracy. The idea behind using the latent features in our case is to be able to build a successful model. The outcome of this project will be a system which will predict the results of the upcoming matches.

1.8 Goals and Objectives

- Better understanding of Machine Learning.
- To develop a system which will predict the win or lose percentage of a cricket match played between two teams.
- To achieve accurate results.
- To develop and implement a successful prediction system.

1.9 Relevant mathematics associated with the Project

System Description:

$$S = [I, Fn, O, Sc, Fc]$$

- Input: $[I_1, I_2, I_3]$

Where,

I_1 =Select the two teams

I_2 =Select the ground and the country

I_3 =Select the 11 team players

- Function: The set of records are partitioned in two subsets that leads to the left branch of the tree and the right one.

Predicates or Quantifiers:

$$Ti = \{t \in T : t(A) \leq x\}$$

$$Tr = \{t \in T : t(A) > x\}$$

Regression (Final addition):

$$\Delta = I(E) - \sum_{i=1}^k \left(\frac{|E_i|}{|E|} \right) * I(E_i)$$

- Output: $[O_1, O_2]$

Where,

O_1 =Main prediction

O_2 =Result of fluctuating prediction

- Success Conditions: When the result of the main prediction will match with the result of the fluctuating prediction.
- Failure Conditions: When the result of the fluctuating prediction will differ from the main prediction in the aspect of the winning team.

1.10 Names of Conferences / Journals

- Journal Of Emerging Technologies and Innovative Research(JETIR)
- International Journal of Engineering Research and Technology (IJERT)

1.11 Plan of Project Execution

NO	TASK	DURATION (Days)	START DATE	END DATE
1	Group Formation	1	16 Jun	17 Jun
2	Decide Project Idea Of Interest	3	20 Jun	23 Jun
3	Requirement Gathering	7	30 Jun	6 Jul
4	Literature Survey	6	23 Jul	29 Jul
5	Mathematical Modeling	9	18 Aug	26 Aug
6	Feasibility Test	10	1 Sept	10 Sept
7	Synopsis Submission	2	14 Sept	15 Sept
8	UML Diagram	4	25 Sept	29 Sept
9	GUI Design	12	03 Jan	14 Jan
10	Functionality Implementation	25	22 Jan	14 Feb
11	Testing	8	12 Mar	20 Mar
12	Documentation	4	24 Mar	28 Mar

Figure 1.1: Project Planner

CHAPTER 2

TECHNICAL KEYWORDS

2.1 Area of Project

Machine Learning

2.2 Technical Keywords

- Machine Learning
- Neural Network
- Clustering
- Decision Tree
- Linear Regression
- Client Server Architecture

CHAPTER 3

INTRODUCTION

3.1 Project Idea

We were trying to implement machine learning techniques on the subject which we are familiar with so keeping this thing in mind we thought of implementing it on a subject like Cricket. We found that cricket is the appropriate topic as there are no such softwares or systems available which would predict the accurate percentile of winning or loosing of a particular team before the start of the match. We also found that there have been a number of sports like football, basketball, volleyball, tennis etc on which the prediction software or systems are available. The systems available for predictions are restricted only for a particular tournament or leagues so we decided to develope a system that will overcome these issues.

3.2 Motivation of the Project

- Earlier the predictions were made after the rst innings because of which there were least chances of getting the correct predictions.The main aspect of the prediction was dependent on the fan following as well as the runs made in the rst half of the match which gave the inaccurate result. In this project we intend to predict the accurate results of the matches depending on the historical data stored in the database.
- In this project the system will give the accurate percentile of the matches based on the previous results.The Batting Averages will be the main aspect for calculating the prediction.

3.3 Literature Survey

1. Tianxiang Cui, Jinpeng Li, et al[1] proposed that they have used GP to the problem of predicting the outcomes of English Premier League games with the result being either win, lose or draw. They have selected 25 features from each game as the input to GP system. The advantage of GP system is that it can generate as many high-quality functions as per required. It uses Bayesian network together with some other machine learning techniques including a decision tree and KNN to predict the results. The overall average accuracy for the Bayesian network is (52.21) Best accuracy Is achieved by using an ANN (68.8). The best overall accuracy achieved by GP system is (76). Limitation of this system is that only 25 features are selected. Hence to overcome this filter can applied to these features in order to detect the importance

2. S. Dobravee , et al[3] proposed to build a goal scores prediction model that uses latent features obtained from matrix factorization process. Nave Bayes classifier is also added to be able to predict outcomes of the match. In this the Matrix Factorization technique is used to build the successful model even when the expect knowledge is not available. Limitation of this system is that the size of database is very small, a short-termed dataset is used. Hence to improve this prediction success is to append regression model that would be used to improve latent features models-based predictions.
3. J. Pan , et al[4] proposed DS-evidence theory to calculate the uncertainty of the data and the unknown a prior probability. For the tennis prediction results this paper is processed a method with evidence theory to compute the uncertainty of competition results. Hybrid hadam evaluation method is proposed in this paper although there is some dejiciary in this. This paper finds that evidence theory can still calculate data uncertainty even though data sources of initialization input is absent. Method accuracy of predicting completion is 70 percent in this paper.
4. T. L. W. Walls, E. J. Bass , et al[5] proposed a regression-based prediction model. This was developed to allow better prediction of attendance for the student general admission seats. At university of Virginia, a regression-based approach was used in this system. The advantage of having this is that the use of regression is a promising method, especially for longer term planning.
5. B. G. Aslan, M. M. Inceoglu , et al[10] proposed two different input vector parameters have been tested via learning vector quantization network in order to emphasize the importance of input parameter selection. Neural networks are used for building the forecasting system about soccer matches. This system has been improved to calculate the probability of outcome of soccer match. The goal ratio compare model had been proposed for predicting the soccer model results. The two input methods have been used i.e. LVQA LVQB. The accuracy provided by the LVQA is 51.29 percent while the accuracy provided by the LVQB is about 53.25 percent. The accuracy of hybrid network is 52.29 percent. Elo has a successful prediction rate of 47.71 percent while goal ratio has corrected prediction rate of 49.02 percent. Limitations of this system is selection of input data to be used in forecasting systems is a critical issue. Hence different leagues, different input parameters and of

course different network structures should be tested in order to achieve a well-balanced generic forecasting system.

6. K. Wickramaratna, Min Chen, et al[11] proposed a Neural Network based framework for semantic events detection in soccer videos. A Hidden Markov model is used to detect the play and break events from soccer videos. There are many issues related to HMM and SVM hence to tackle the issues a novel learning-based event detection framework is processed. In this paper, which incorporates both strength of multi model analysis and ability of neural network ensembles to enhance the generalization capabilities. In this paper, an advanced framework for goal event detection in soccer videos are proposed using multi model processing and the classification power of neural network ensembles the future work is to extend the framework for multiple event detection at different domains.
7. D. Comanier, V. Ramesh, et al[12] proposed a prediction after based prediction filter to track the players distinctively in interaction events of the volleyball crash pattern model and other pattern models are use to detect each tracking objects after intersection. Algorithms such as Mean-shift, Cam-shift, SIFT, Kalman Filter, Extended Kalman Filter, Particle Filter had their own merit. This paper proposes a prediction after intersection-based Particle filter the success rate of our proposal is around 80 percent while the conventional one is about 30 percent.
8. Z. Yijie, x. Sun , et al[13] proposed offers finalized description of outcome prediction for sport completions. It proposes a novel based prediction model and team model-based study on existing technologies EM algorithm is used in this paper. Different models like defect analysis of traditional model, Game model, Team model is used. The result in the offensive and defensive rounds and the total score to predict the exact rate of is between 65 percent and 70 percent which all approximate to traditional methods.
9. B. Zhao, L. Chen, et al[14] proposed a prediction model of sports result based on knowledge discovery in database. The method combines the multiple light weighted models with a variety technical to improve the accuracy of prediction model of sport result. This paper proposed KDD modelling method to analysis and predict the non-against games and predict the result of the game and analysis and determine individual of the game.

10. Q. Wang, Z. Sun, et al[18] proposed the problem of spin classification firstly. An adopted and improved the extreme Learning Machine model is presented. Trajectory prediction methods can be classified into Two categories: Experience model and Parameter model. Experience model uses local weighted regression, Parameter model uses the stress analysis study. Extreme Learning Machine is newly developed neural network algorithm. The experiment result should be a major improvement fojr improved ELM in both classification precision and efficiency.

CHAPTER 4

PROBLEM DEFINITION AND SCOPE

4.1 Problem Statement

To develop a system which will predict the win or lose percentage of a cricket match played between two teams.

4.1.1 Goals and objectives

- Better understanding of Machine Learning.
- To develop a system which will predict the win or lose percentage of a cricket match played between two teams.
- To achieve accurate results.
- To develop and implement a successful prediction system.

4.1.2 Statement of scope

- To develop and implement a successful prediction system.
- The system is completely user friendly and easy to use.
- To achieve accurate results.
- Better understanding of Machine Learning.

4.2 Major Constraints

The Batsmans Performance and the difference between uctuated prediction and nal prediction can impact the manner of the project.

4.3 Methodologies of Problem solving and efficiency issues

- At the Client side the end user must select the two teams. Depending on the selection of both the teams, the country and the ground where the match will be held will display automatically.
- After the toss, the playing 11 players of both the teams will be short listed.

- Depending on the batting average of the player as well as the experience of that player on that ground against rival country. By keeping these averages in consideration the server will perform the main prediction.
- After the match starts the prediction can differ depending on the following factors like Batsmens Performance, Number of wickets, Difference between projected score and the current score
- The fluctuated prediction will be compared with the main prediction.(Fluctuated prediction means again the graph will be formed on real time values and will be compared with the graphs that have been made by main predictions.)
- When the result of the fluctuated prediction matches with the result of the main prediction then we have the accurate output.
- When the result of the fluctuated prediction differs from the result of the main prediction then the system will give a reason of failure.

4.4 Outcome

Successful Prediction.

4.5 Applications

- Media: The system will work as an expert analyzer.
- Newspaper: It will give the descriptive knowledge about the playing eleven players of the particular match.
- Online Applications: Example Dream11: It will help the users to know the best playing 11 for a particular match and use this knowledge to play on different Cricket applications.

4.6 Hardware Resources Required

- CPU Speed: 2Ghz. If it is less execution time may be more.
- RAM: 3GB. Lower RAM may lead to system lag.
- Hard Disk Drive: 5 GB. High volume generated data.

4.7 Software Resources Required

1. Eclipse
2. XAMPP Server

CHAPTER 5

PROJECT PLAN

5.1 Project Estimates

We have used Waterfall Model for our Project as we had a period of 9 months to implement our entire Project.

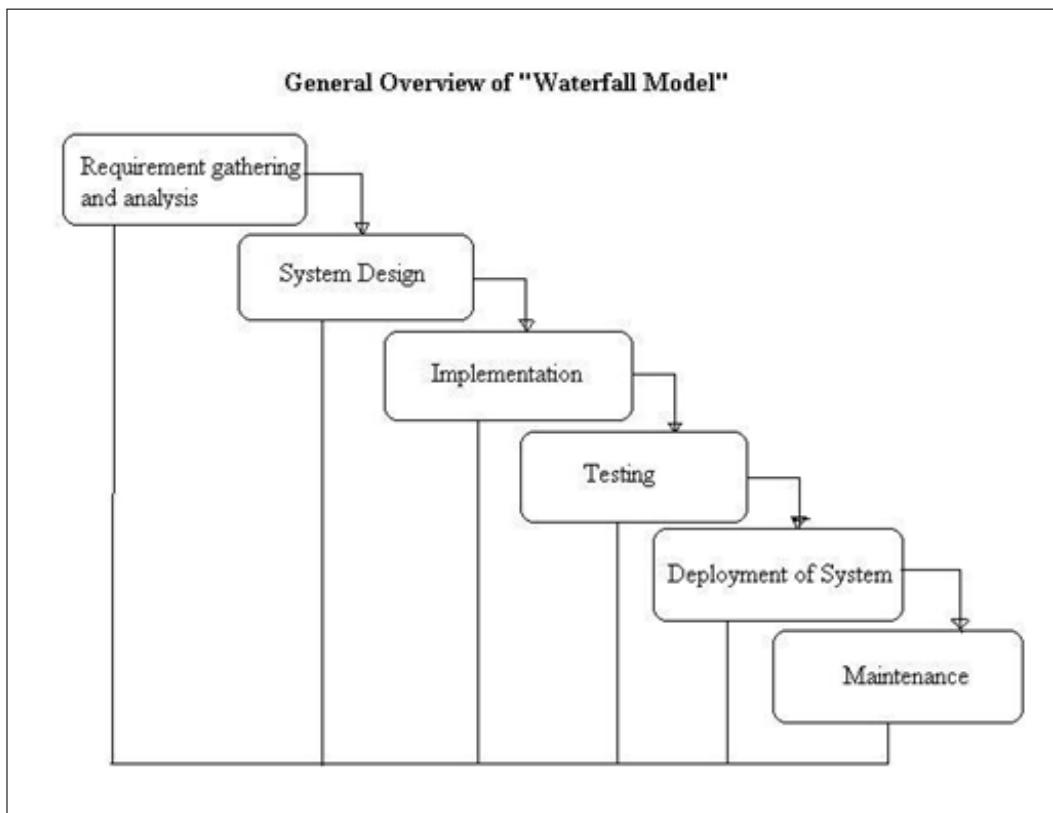


Figure 5.1: Waterfall Model

1. Requirement Gathering and Analysis:

In this step of Waterfall Model, we have identified the various requirements needed for our Project such as software and hardware, database and interfaces.

2. System Design:

In the system design phase, we have designed the system which is easily understandable for the end - user, i. e., in a user - friendly manner. We have designed UML diagrams and data flow diagrams to understand the system module and sequence of execution.

3. Implementation:

In implementation phase of our Project, we have implemented various modules required to successfully get expected outcome at different module levels.

4. Testing:

Different test cases were generated to test whether the Project modules are giving expected outcome or not in given time limit. All units developed were integrated into a system after testing of each unit. Post integration, the entire system was tested for any faults and failures.

5. Deployment of System:

After the functional and non - functional testing was done, the product was shown to the Sponsor.

6. Maintenance:

There are certain issues which arise in the client environment. In maintenance phase, those issues need to be fixed and better versions are released at regular intervals.

In Waterfall Model, the progress is seen flowing downwards like a waterfall. The next phase initiates only after the defined set of goals are achieved for previous phase, so the name 'Waterfall Model'.

5.1.1 Reconciled Estimates

5.1.1.1 Cost Estimate

Not Applicable.

5.1.1.2 Time Estimates

Approximately 9 months.

5.1.2 Project Resources

- LAN connection
- Eclipse
- XAMPP Server

5.2 Risk Management w.r.t. NP Hard analysis

Np Complete: NP-Complete(NPC): NP-Complete problems belong to NP class, these are a set of all decision problems whose solutions can be verified in polynomial time; NP may be equivalently defined as the set of decision problems that can be solved in polynomial time. In our proposed system the complexity of solving a linear regression model changes according to the numerical method chosen to calculate it. The theoretical computation complexity for regression model is $O(NC^2)$. The run time of the model varies linearly with the number of instances for a given number of attributes. .

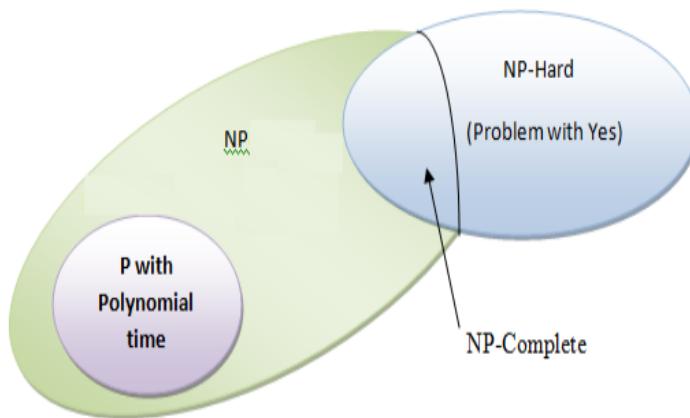


Figure 5.2: NP Complete

5.2.1 Risk Analysis

The risks for the Project can be analyzed within the constraints of time and quality

ID	Risk Description	Probability	Impact		
			Schedule	Quality	Overall
1	Accuracy	Low	Low	High	Low
2	Availability	High	Low	High	High
3	Compatibility	High	Low	High	High

Table 5.1: Risk Table

Probability	Value	Description
High	Probability of occurrence is	> 75%
Medium	Probability of occurrence is	26 – 75%
Low	Probability of occurrence is	< 25%

Table 5.2: Risk Probability definitions

Impact	Value	Description
Very high	> 10%	Schedule impact or Unacceptable quality
High	5 – 10%	Schedule impact or Some parts of the project have low quality
Medium	< 5%	Schedule impact or Barely noticeable degradation in quality Low Impact on schedule or Quality can be incorporated

Table 5.3: Risk Impact definitions

5.2.2 Overview of Risk Mitigation, Monitoring, Management

Following are the details for each risk.

Risk ID	1
Risk Description	No Internet Connection
Category	Networking
Source	Server
Probability	High
Impact	High
Response	Mitigate
Strategy	Continues Internet Connection
Risk Status	Identified

Table 5.4: Table 1

Risk ID	2
Risk Description	Inappropriate Data
Category	Software
Source	Database
Probability	Low
Impact	High
Response	Mitigate
Strategy	Accurate Data
Risk Status	Identified

Table 5.5: Table 2

Risk ID	3
Risk Description	System Crash
Category	Web Browser
Source	System
Probability	Medium
Impact	Moderate
Response	Mitigate
Strategy	Better testing will resolve this issue
Risk Status	Identified

Table 5.6: Table 3

5.3 Project Schedule

5.3.1 Project task set

Major Tasks in the Project stages are:

- Task 1: Correctness.
- Task 2: Availability.
- Task 3: Integrity.

5.3.2 Task network

Project tasks and their dependencies are noted in this diagrammatic form.

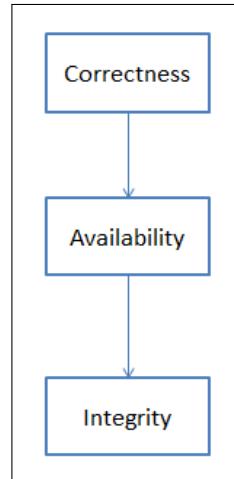


Figure 5.3: Task network

5.3.3 Timeline Chart

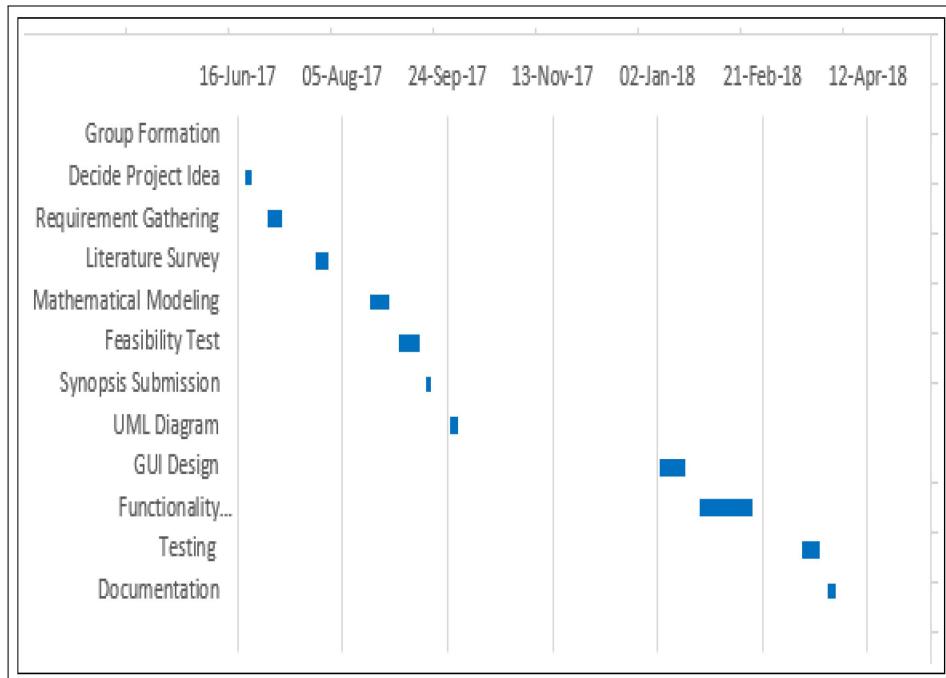


Figure 5.4: Timeline chart

5.4 Team Organization

5.4.1 Team structure

The team structure for the project is identified. Roles are defined.

Sr. No.	Member	Responsibilities
1	Shivanee Pawar	Developer and Database Management
2	Sahil Mulani	Testing and Developer
3	Minakshee Naik	Developer and Requirement Gathering
4	Abhishek Naik	Project Analysis,Developer and Design

Table 5.7: Team Structure

5.4.2 Management reporting and communication

Well planning mechanisms are used for progress reporting and inter/intra team communication are identified as per requirements of the project

CHAPTER 6

SOFTWARE REQUIREMENT SPECIFICATION

6.1 Introduction

6.1.1 Purpose and Scope of Document

Sports play a very vital role in the media coverage. Cricket, which comes in the top five games, played in the world, covers almost half of the population (followers) in Asia. It is also a very popular sport in Australia(National Sport) and England. India is considered as a dominant team in Cricket and as Data Analysis is becoming more common especially in Sports. Using data analysis results has become familiar in sports organization such as International Cricket Council (ICC), International Federation of Association Football (FIFA), and Grand Slam of the International Tennis Federation. In sports like Cricket and Football predicting the results become more difficult due to the changing nature of the Game. This project utilizes Machine learning techniques using client-server architecture to predict which team will win between the two sides playing based on past data available. Different algorithms namely Logistic Regression, Neural Network, KMeans clustering are used. In this model, the outcome is predicted based on the players' performance on that particular ground. The model takes into account the players' bowling and batting average, career average of the player as well as the experience of that particular player against that particular rival. There is also a fluctuating prediction which comes into picture when the live match starts. This fluctuating prediction differs from the main prediction when the player performs less or the player performs very well than the past performance. Hence, this model will give the accurate prediction because nowadays, people focus on sports results, which is an important aspect, and the result of the game has become the focus and concentration of sports game.

6.1.2 Overview of responsibilities of Developer

1. To have understanding of the problem statement.
2. To know what are the hardware and software requirements of proposed system.
3. To have understanding of proposed system.
4. To do planning various activities with the help of planner.
5. Designing, programming, testing etc

6.2 Usage Scenario

This section provides various usage scenarios for the system to be developed.

6.2.1 User profiles

There is an user and actor which is as follows,

1. User: The user will choose Random or Live match and accordingly select the teams and the ground respectively.
2. System: System consists of a database as well as server to perform multiple tasks and identifying the best players average.

6.2.2 Use-cases

All use-cases for the software are presented. Description of all main Use cases using use case template is to be provided.

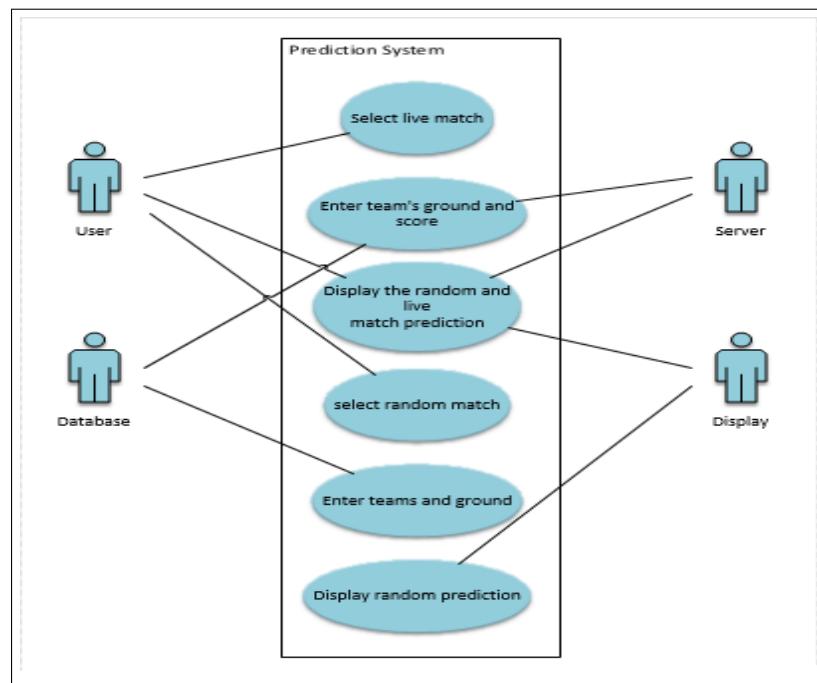


Figure 6.1: Use Cases

6.2.3 Use Case View

Use Case Diagram. Example is given below

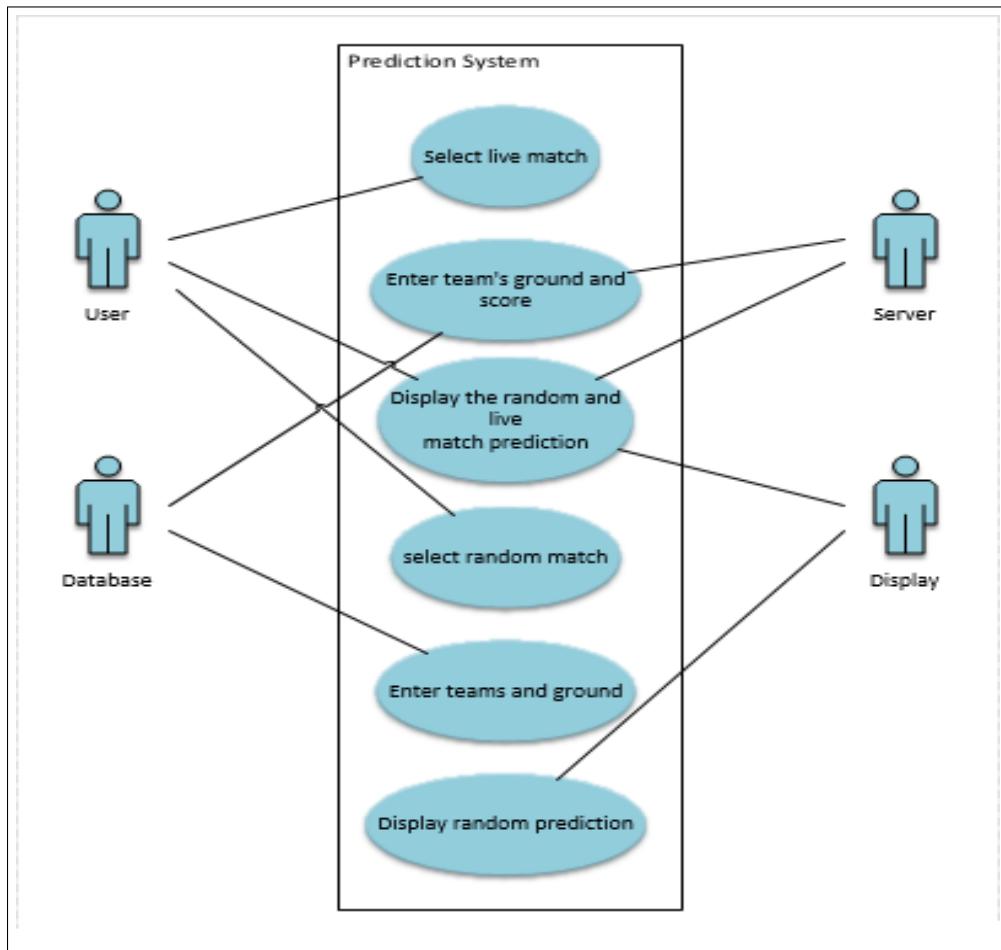


Figure 6.2: Use case diagram

6.3 Data Model and Description

6.3.1 Data Description

The data streams will be processed and analyzed by framework. This data needs to be converted into a suitable form so that it can be compared with the database and proceed efficiently to get the desired result for end user.

6.3.2 Data objects and Relationships

The data streams are given to framework which will classify that data and then this data is given to server for processing and analysis purpose, so this is how the two modules are connected.

6.4 Functional Model and Description

Data streams will be processed and analyzed by framework and server. The server will use the past data as well as players statistical data in order to show the runs that the particular player will score.

6.5 Non-functional Requirement

6.5.1 Performance Requirement :

- Every module of the system should work efficiently.
- The system should perform fast in all weather conditions.

6.5.2 Safety and Security Requirement:

The java application will not affect other applications on user's system.

6.6 UML Diagrams:

6.6.1 Data Flow Diagram

6.6.1.1 Level 0 Data Flow Diagram

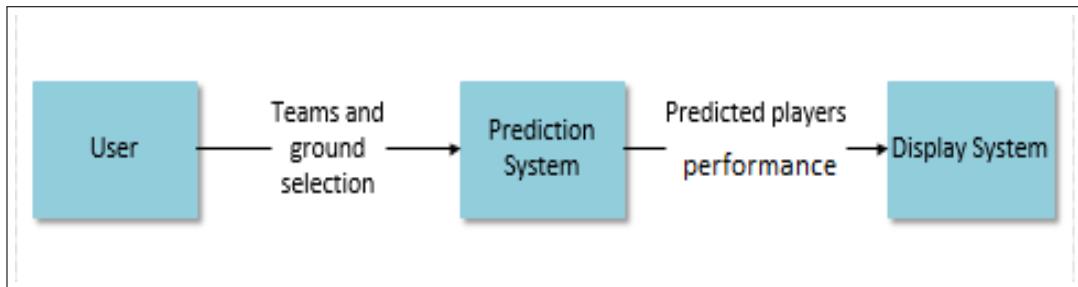


Figure 6.3: DFD0

6.6.1.2 Level 1 Data Flow Diagram

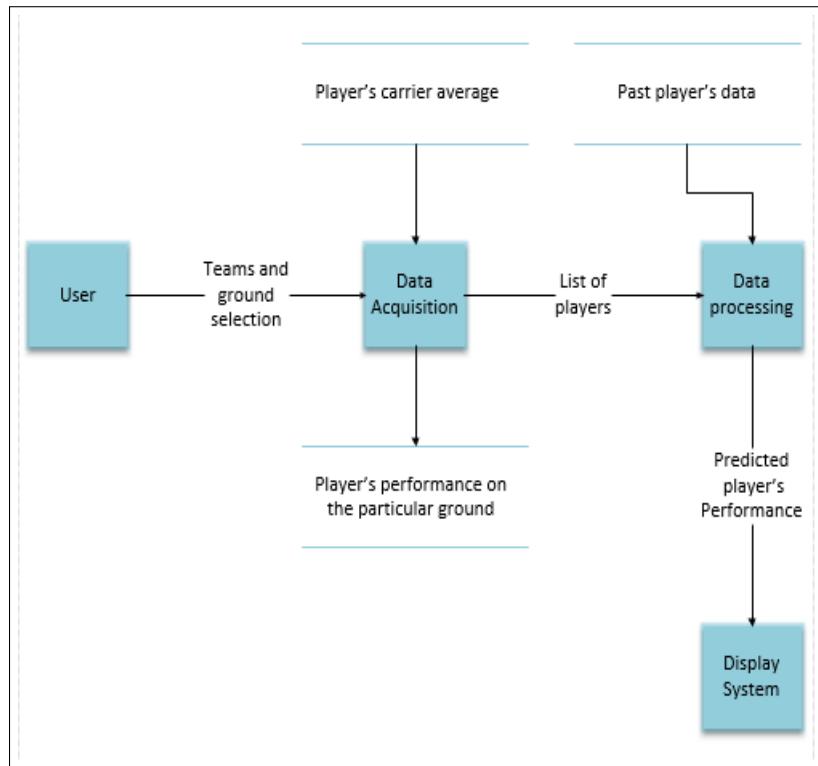


Figure 6.4: DFD1

6.6.1.3 Level 2 Data Flow Diagram

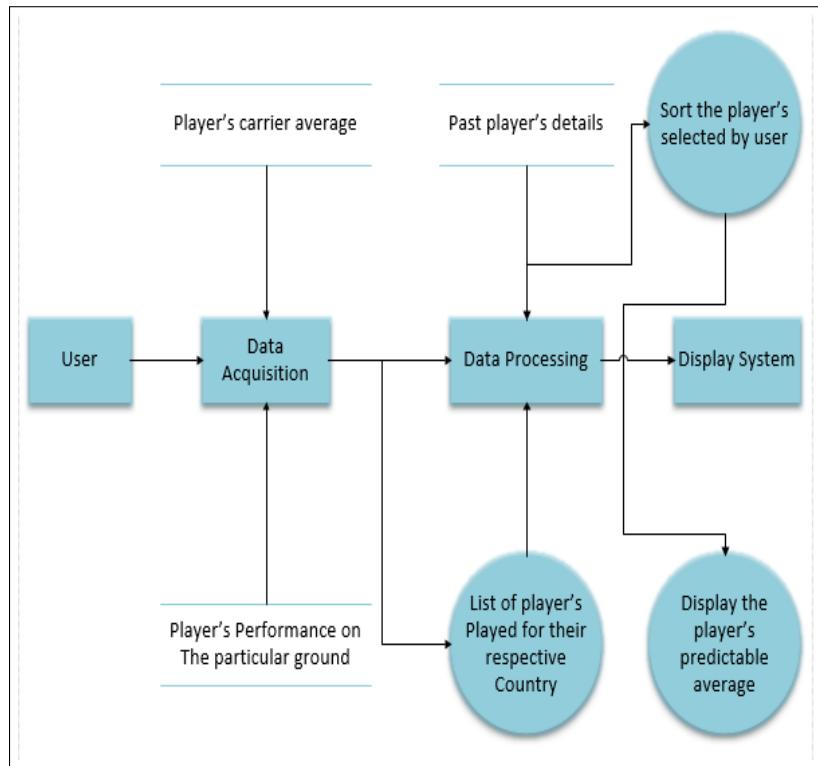


Figure 6.5: DFD2

6.6.2 Activity Diagram:

The Activity diagram represents the steps taken.

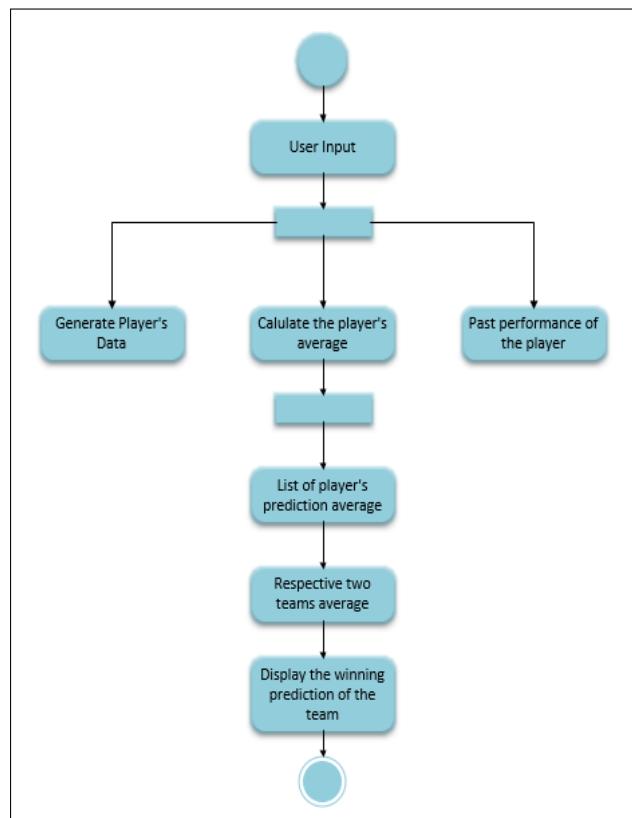


Figure 6.6: Activity diagram

6.6.3 Design Constraints

Xampp Web Server and Internet connectivity.

6.6.4 Software Interface Description

The software interface to the outside world is very good and user friendly for analyzed Data The requirements for interfaces are Windows 7/8/10, Linux(Fedora/Ubuntu),MacOS.

CHAPTER 7

DETAILED DESIGN DOCUMENT

7.1 Introduction

This document specifies the overall architectural design of the project under implementation.

7.2 Architectural Design

A description of the program architecture is presented. Subsystem design or Block diagram, Package Diagram, Deployment diagram with description is to be presented.

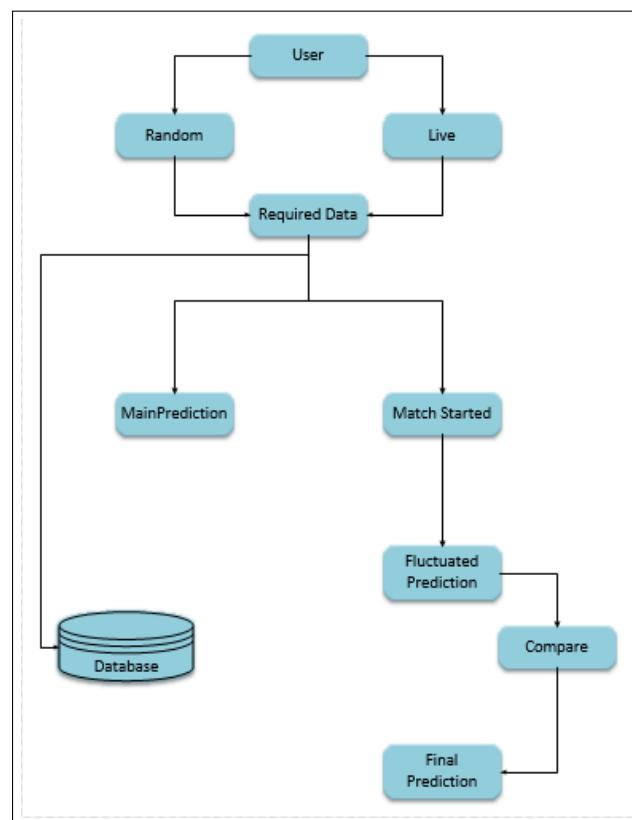


Figure 7.1: Architecture diagram

- At the Client side the end user must select the two teams. Depending on the selection of both the teams, the country and the ground where the match will be held will display automatically.

- After the toss, the playing 11 players of both the teams will be short listed by the server and displayed to the client.
- Depending on the batting average of the player as well as the experience of that player on that ground against rival country. By keeping these averages in consideration the server will perform the main prediction.
- After the match starts the prediction can differ depending on the following factors like Batsmens Performance, Number of wickets, Difference between projected score and the current score.
- The fluctuated prediction will be compared with the main prediction.(Fluctuated prediction means again the graph will be formed on real time values and will be compared with the graphs that have been made by main predictions.)
- When the result of the fluctuated prediction matches with the result of the main prediction then we have the accurate output.
- When the result of the fluctuated prediction differs from the result of the main prediction then the system will give a reason of failure.

7.3 Data design

A description of all data structures including internal, global, and temporary data structures, database design (tables), file formats. .

7.3.1 Internal software data structure

Data related to players is stored in the database. This data is continuously read and processed to get the accurate information of the players and is made available in the form of text to the end user.

7.3.2 Global data structure

Global data structure would be the data stored in the database or live data. Here data is stored in the database. Database is used to store a large number of variables or data. Data structures that are available to major portions of the architecture are describe.

7.3.3 Temporary data structure

- String: Strings are among the most popular types. We can create them simply by enclosing characters in quotes.
- int: They are often called just integers or int, are positive or negative whole numbers with no decimal point.
- Float: Also called floats, they represent real numbers and are written with a decimal point dividing the integer and fractional parts. Floats may also be in scientific notation, with E or e indicating the power of 10 ($2.5\text{e}2 = 2.5 \times 10^2 = 250$).

7.3.4 Database description

Database includes attributes like:

- 1.Player
- 2.Stadium
- 3.Opponent
- 4.Career Average
- 5.Chasing
- 6.Average of ground
- 7.Average of country
- 8.Average of country w.r.t team
- 9.Average of ground w.r.t team
- 10.Win/Loss

CHAPTER 8

PROJECT IMPLEMENTATION

8.1 Introduction

Different tools and technologies are used in implementation of this project. Description of each is given below.

8.2 Tools and Technologies Used

1. Xampp Web Server

- XAMPP is a free and open source cross-platform web server solution stack package developed by Apache Friends, consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages.
- XAMPP stands for Cross-Platform (X), Apache (A), MariaDB (M), PHP (P) and Perl (P). It is a simple, lightweight Apache distribution that makes it extremely easy for developers to create a local web server for testing and deployment purposes. Everything needed to set up a web server application (Apache), database (MariaDB), and scripting language (PHP) is included in an ex-tractable file
- XAMPP is also cross-platform, which means it works equally well on Linux, Mac and Windows. Since most actual web server deployments use the same components as XAMPP, it makes transitioning from a local test server to a live server extremely easy as well.

2. Eclipse

- Eclipse is an IDE (integrated development environment) used in computer programming, and is the most commonly used IDE for Java development . It contains a base workspace and an extensible plug-in system for plotting environment. It was one of the first Integrated Development Environment to run under GNU Class path and it runs with no problems under IcedTea.
- Eclipse is a free of cost, Java-based development platform known for its plugins that allow developers to develop and test code written in other programming languages. Eclipse is released under the terms of the Eclipse Public License.

- Eclipse is written in C and Java language which makes it more user friendly and platform independent. Eclipse tool basically used to create the project of different technology.
3. JavaScript
- JavaScript is a dynamic client-side scripting language, which means the source code is processed by clients web browser rather than web server. Thus, JavaScript can run after a webpage is loaded without communicating with the server. JavaScript is an interpreted, light-weight, network-centric application-driven programming language. It is complimentary to and integrated with Java. It supports event-driven, functional, and imperative programming styles

8.3 Hardware Resources Required

Not required

8.4 Software Resources Required

1. Eclipse
2. Xampp Server

8.5 Methodologies/Algorithm Details

- At the Client side the end user must select the two teams. Depending on the selection of both the teams, the country and the ground where the match will be held will display automatically.
- After the toss, the playing 11 players of both the teams will be short listed.
- Depending on the batting average of the player as well as the experience of that player on that ground against rival country. By keeping these averages in consideration the server will perform the main prediction.
- After the match starts the prediction can differ depending on the following factors like Batsmens Performance, Number of wickets, Difference between projected score and the current score

- The fluctuated prediction will be compared with the main prediction.(Fluctuated prediction means again the graph will be formed on real time values and will be compared with the graphs that have been made by main predictions.)
- When the result of the fluctuated prediction matches with the result of the main prediction then we have the accurate output.
- When the result of the fluctuated prediction differs from the result of the main prediction then the system will give a reason of failure.

8.5.1 Algorithm 1/Pseudo Code

1. Decision Tree Algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, decision tree algorithm can be used for solving regression and classification problems too. The general motive of using Decision Tree is to create a training model which can use to predict class or value of target variables by learning decision rules inferred from prior data(training data). The understanding level of Decision Trees algorithm is so easy compared with other classification algorithms. The decision tree algorithm tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label.
 - 1.Place the best attribute of the dataset at the root of the tree.
 - 2.Split the training set into subsets. Subsets should be made in such a way that each subset contains data with the same value for an attribute.
 - 3.Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.
2. Input
 - Data partition, D, which is a set of training tuples and their associated class labels.
 - Attribute List: The set of candidate attributes
 - Attribute Selection Method, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a splitting attribute and, possibly, either a split-point or splitting subset.

3. Decision Tree Construction

- Tree is constructed in a top-down, recursive, divide-and-conquer manner.
- At start, all the training examples are at the root.
- Attributes are categorical (Note that if continuous-valued, they are discretized in advance)
- Test attributes are selected on the basis of a heuristic or statistical measure.

4. Conditions for stopping partitioning

- All samples for a given node belong to the same class.
- There are no remaining attributes for further partitioning - majority voting is employed for classifying the leaf.
- There are no sample left.

5. Attribute Selection

- If dataset consists of n attributes then deciding which attribute to place at the root or at different levels of the tree as internal nodes is a complicated step. By just randomly selecting any node to be the root can't solve the issue. If we follow a random approach, it may give us bad results with low accuracy.
- For solving this attribute selection problem, researchers worked and devised some solutions. They suggested using some criterion like information gain, gini index, etc. These criterions will calculate values for every attribute. The values are sorted, and attributes are placed in the tree by following the order i.e, the attribute with a high value(in case of information gain) is placed at the root. While using information Gain as a criterion, we assume attributes to be categorical, and for gini index, attributes are assumed to be continuous.

• Pseudo-Code

```
if (the sub-table passed to the algorithm is empty)
return NULL;
if (the sub-table passed to the algorithm is homogeneous (if all the rows have
the same value for the last column)
```

```

{mark this node as a leaf node;
label this node with the value of the last column
return a pointer to this node}

else

{decide a column to split the table on based on information gain
set the node's split on value to this column's name
for all the values that the splitting column can take:
create a new node
set the new node as the current node's child node
prune the sub-table so that all the rows with this value of the last column
are removed
recursively call the function by passing it the new pruned table and the new
node}

```

8.6 Verification and Validation for Acceptance

- Validation

1. This application performs all verifications of field like if team1 is selected as India then it is not allowed to select same team as team2.
2. Also all the fields are validated for null values.

- Verification

3. When user selects teams and players then our application shows proper o/p and allows him/her to change chasing value. And it is verified that the system gives proper output values and shows accurate graph results.

CHAPTER 9

SOFTWARE TESTING

9.1 Type of Testing Used

- 1.Unit Testing
- 2.Integration Testing
- 3.Acceptance Testing

- Unit testing: In this testing we have evaluated all the modules while they are implemented. Also we have tested each module separately like selecting team, players. Also the result module is tested separately.
- Integration testing: In this testing we have tested basic working of all modules after integrating individual modules. After integration of all module our system is working as per the requirement.
- Acceptance testing: Acceptance testing is the last phase of functional testing and is used to assess whether or not the final piece of software is ready for delivery. It involves ensuring that the product is in compliance with all of the original business criteria and that it meets the end users needs. This requires the product be tested both internally and externally, meaning you will need to get it into the hands of your end users for beta testing along with those of your QA team. Beta testing is key to getting real feedback from potential customers and can address any final usability concerns.

9.2 Test Cases and Test Results

Sr No.	Test Id	Test Case	Pass Criteria	Pass/Fail
1.	Test-01	Selection of the Second Team	Same Team should not be repeated	Pass
2.	Test-02	Select 11 players of team 1 and team 2	5 batsman, 1 wicketkeeper, 1 all rounder and 4 bowlers must be selected	Pass
3.	Test-03	Toss Selection	Chasing or Batting select any one	Pass

Figure 9.1: Test Cases

CHAPTER 10

RESULTS

10.1 Output Screenshots

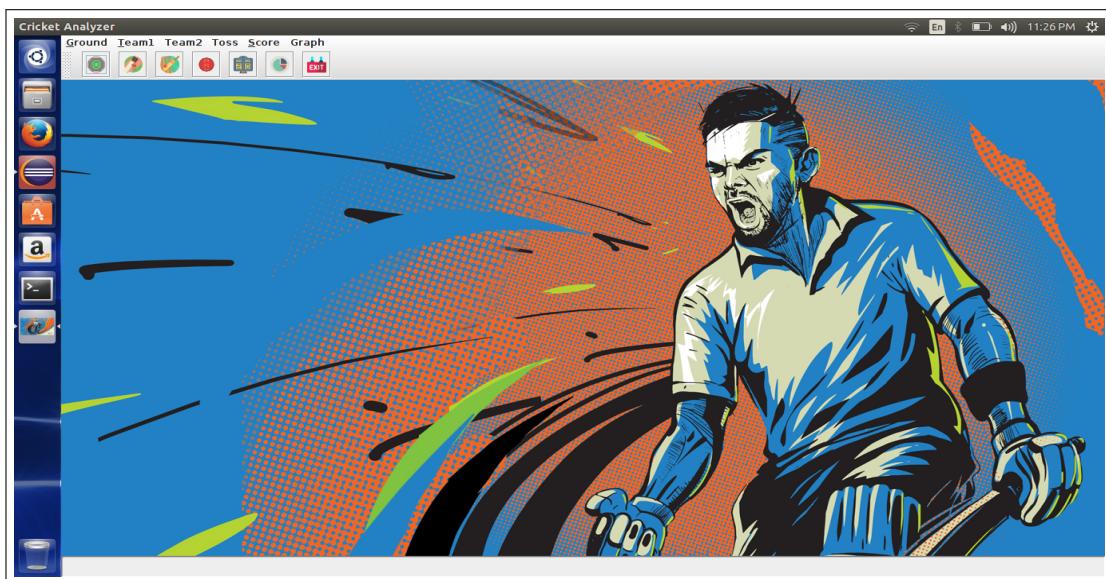


Figure 10.1: Home Page

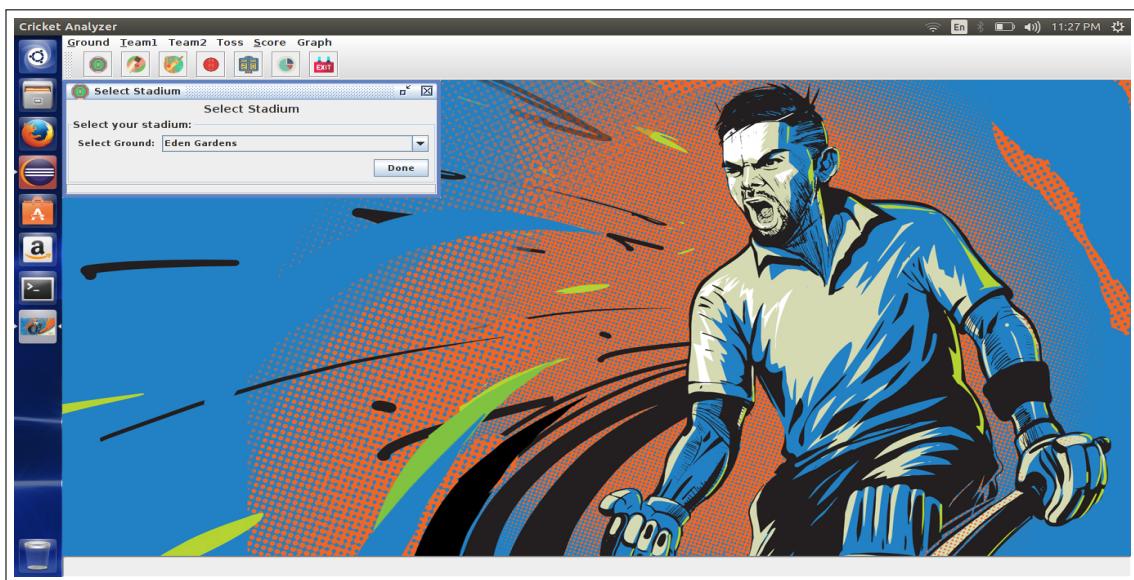


Figure 10.2: Selection of Stadium

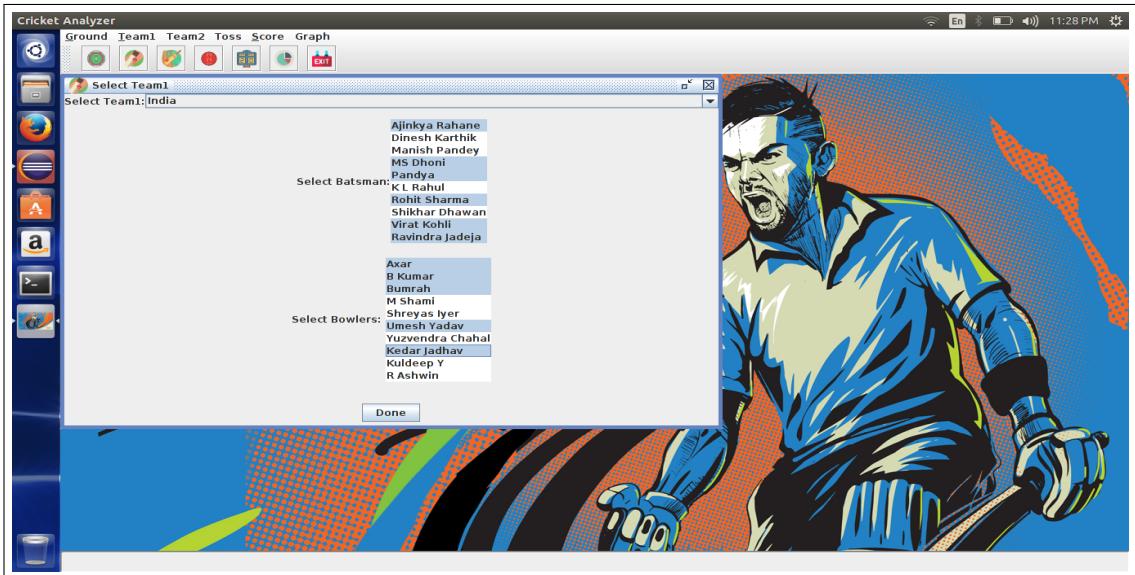


Figure 10.3: Selection of First Team Players

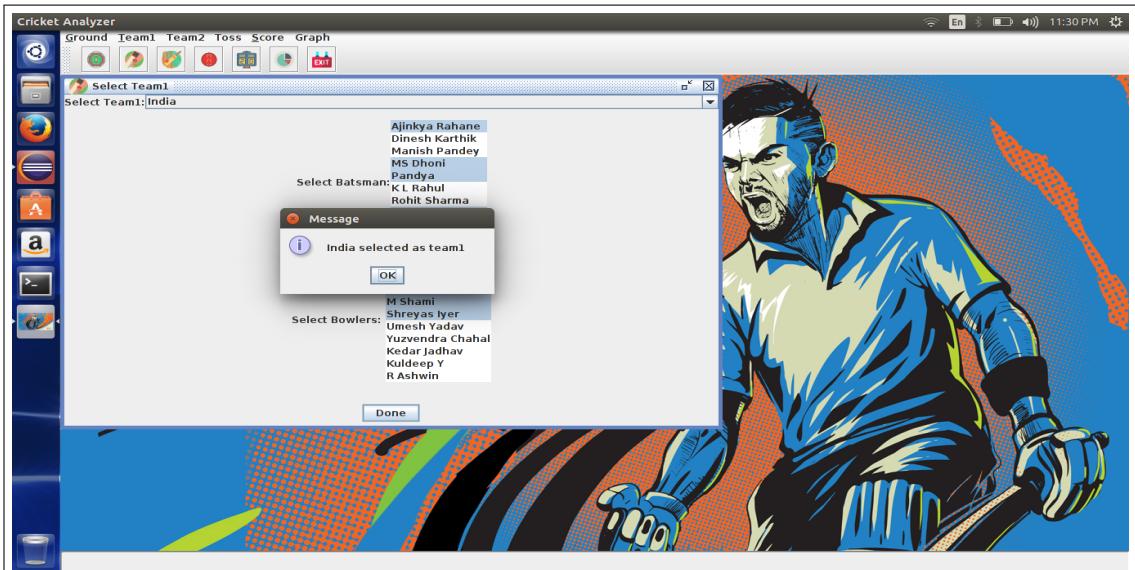


Figure 10.4: Confirmation Message

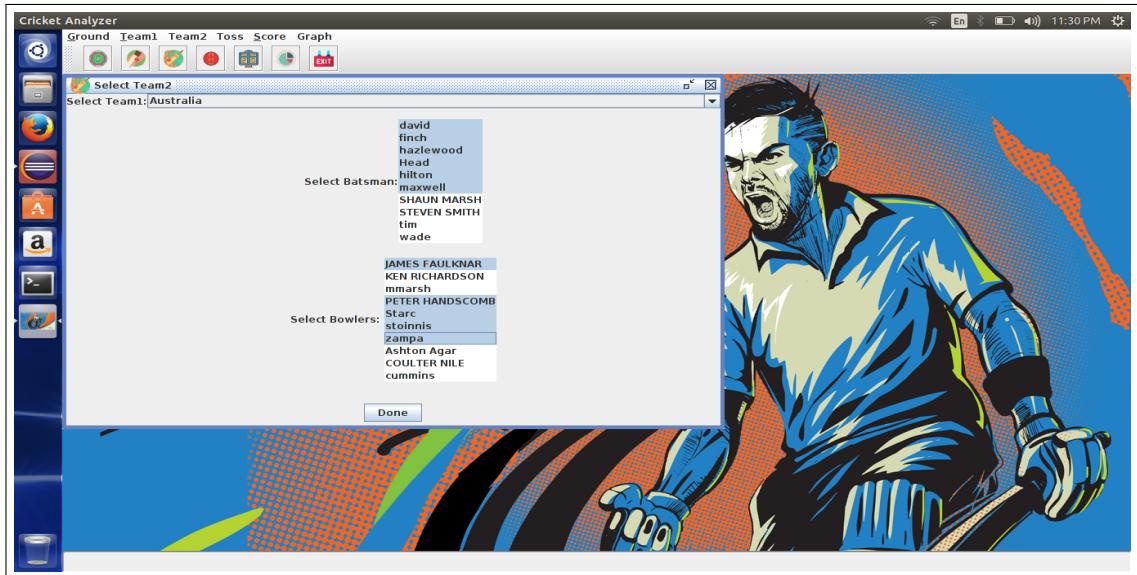


Figure 10.5: Selection of Second Team Players

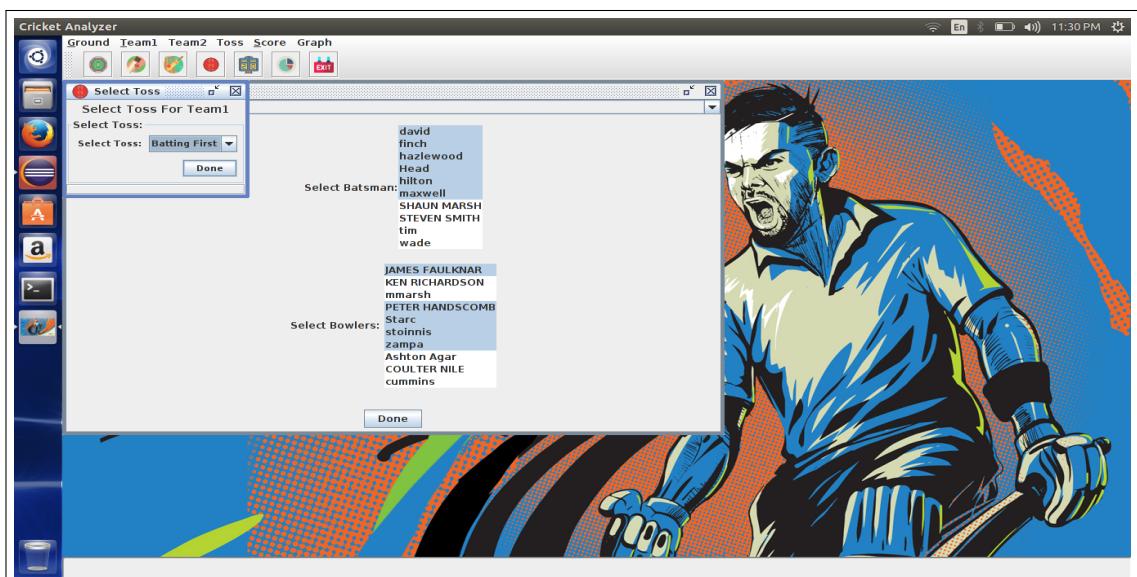


Figure 10.6: Toss Selection

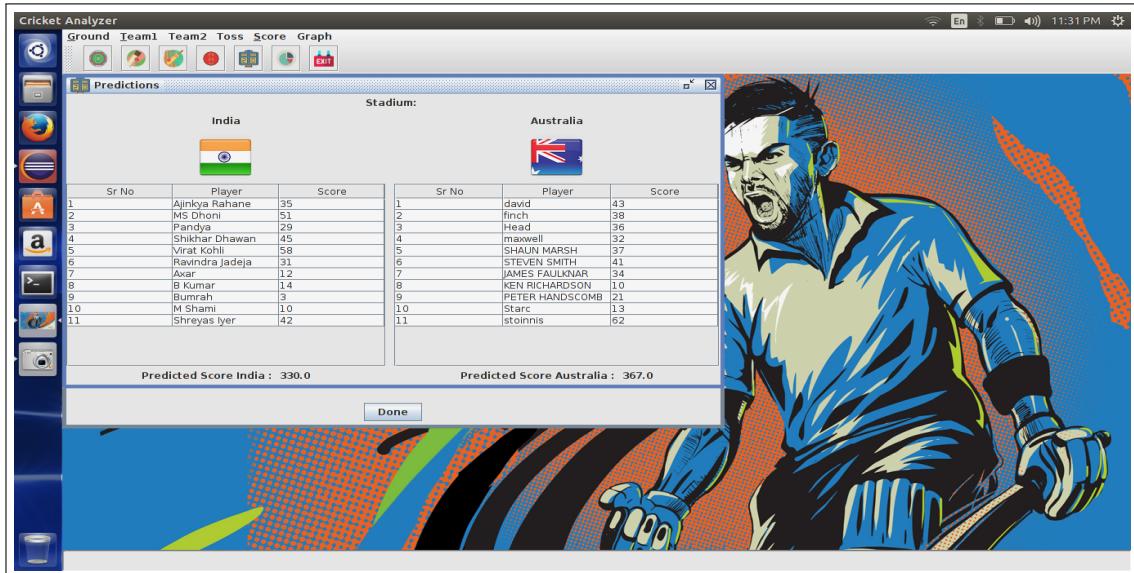


Figure 10.7: Predicted Score



Figure 10.8: Statistics

CHAPTER 11

DEPLOYMENT AND MAINTENANCE

11.1 Installation and un-installation

- Installation

Since our system is developed using PHP(Admin side) and Core java(Client side). In order to install our system we just need to create executable jar for client side and for installing admin panel we need to host our site on cpanel on any of the server and also we need to create database on server side using cpanel.

- Uninstallation

For uninstalling server side we just need to block the access to the admin url. Or we can remove hosted files from the cpanel. For uninstalling client side we dont need any steps as we are using executable jar.

CHAPTER 12

CONCLUSION AND FUTURE SCOPE

Conclusion:

We proposed a client server architecture for Winning Prediction System for a sport like cricket keeping in mind the popularity which cricket has in India. Based on the historical data of the players and considering the major four aspects i.e. batting,bowling, carrier average and the past experience of the player we have calculated the accurate prediction for the respective teams. Unlike the different prediction systems which makes use of ANN strategy our winning prediction system makes use of the logistic regression, neural network and k-means clustering strategies and relatively gives a better result. As we know, people nowadays focus on the sport results, which is an important aspect and the result of the game has become the focus and concentration of sports game, so our system turn out to be the best as it gives the correct percentile of the game.

Future Scope:

In future we can give Predictions in different formats of cricket like Test matches and Twenty-Twenty. We can also give the Prediction of major series like Champions Trophy, Ashes series, World Cups (50-50, 20-20). Predictions can also be made when the match is abandoned due to rain, bad light etc.

CHAPTER 13

REFERENCES

- 1 Tianxiang Cui, Jingpeng Li, J. R. Woodward and A. J. Parkes, An ensemble based Genetic Programming system to predict English football premier league games, 2013 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), Singapore, 2013, pp. 138-143
- 2 J. Hucaljuk and A. Rakipovi, Predicting football scores using machine learning techniques, 2011 Proceedings of the 34th International Convention MIPRO, Opatija, 2011, pp. 1623-1627.
- 3 S. Dobravec, Predicting sports results using latent features: A case study, 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2015, pp. 12671272.
- 4 J. Pan, Tennis Match Prediction Model Based on Improved D-S Evidence Theory, 2016 International Conference on Robots Intelligent System (ICRIS), Zhangjiajie, 2016, pp. 414-417
- 5 T. L. W. Walls and E. J. Bass, A regression-based predictive model of student attendance at UVA mens basketball games, Proceedings of the 2004 IEEE Systems and Information Engineering Design Symposium, 2004., Charlottesville, VA, 2004, pp. 203-208.
- 6 D. Miljkovi, L. Gaji, A. Kovaevi and Z. Konjovi, The use of data mining for basketball matches outcomes prediction, IEEE 8th International Symposium on Intelligent Systems and Informatics, Subotica, 2010, pp. 309-312.
- 7 K. Trawinski, A fuzzy classification system for prediction of the results of the basketball games, International Conference on Fuzzy Systems, Barcelona, 2010, pp. 1-7.
- 8 Y. Saito, M. Kimura and S. Ishizaki, Real-time prediction to support decision making in soccer, 2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), Lisbon, 2015, pp. 218-225.
- 9 B. G. Aslan and M. M. Inceoglu, A Comparative Study on Neural Network Based Soccer Result Prediction, Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007), Rio de Janeiro, 2007, pp. 545-550.
- 10 K. Wickramaratna, Min Chen, Shu-Ching Chen and Mei-Ling Shyu, Neural network based framework for goal event detection in soccer

videos, Seventh IEEE International Symposium on Multimedia (ISM05), 2005, pp. 8 pp.

- 11 D. Comaniciu, V. Ramesh, and P. Meer, Player Tracking using prediction after intersection based practicle filter for Volley Ball match, IEEE Transactions on, Vol. 25, No. 5, pp.564-577, May. 2003.
- 12 Z. Yijie and X. Jun, Competition Results Prediction Model Based on Athlete Ability Data Simulation and Analysis, 2015 Sixth International Conference on Intelligent Systems Design and Engineering Applications (ISDEA), Guiyang, 2015, pp. 222-225.
- 13 B. Zhao and L. Chen, Prediction Model of Sports Results Base on Knowledge Discovery in Data-Base, 2016 International Conference on Smart Grid and Electrical Automation (ICSGEA), Zhangjiajie, 2016, pp. 288-291.
- 14 A. McCabe and J. Trevathan, Artificial Intelligence in Sports Prediction, Fifth International Conference on Information Technology: New Generations (itng 2008), Las Vegas, NV, 2008, pp. 1194-1197.
- 15 R. Vuillemot and C. Perin, Sports Tournament Predictions Using Direct Manipulation, in IEEE Computer Graphics and Applications, vol. 36, no. 5, pp. 62-71, Sept.-Oct. 2016.
- 16 Z. Bo, Q. Chaoling, X. Xiaoli and Z. Fanbo, GM (1,1) Model Gray Prediction for the Gold-Medal Result of Womens Put Shot in the 30th Olympic Games, 2011 International Conference on Future Computer Science and Education, Xian, 2011, pp. 334-337.
- 17 Q.Wang and Z. Sun, Trajectory identification of spinning ball using improved extreme learning machine in table tennis robot system, 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), Shenyang, 2015, pp. 551-554.
- 18 J. Liu, Z. Fang, K. Zhang and M. Tan, A new data processing architecture for table tennis robot, 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, 2015, pp. 1780-1785.

ANNEXURE A

LABORATORY ASSIGNMENTS ON

PROJECT ANALYSIS OF

ALGORITHMIC DESIGN

- To develop the problem under consideration and justify feasibility using concepts of knowledge canvas and IDEA Matrix.

Refer innovationbook for IDEA Matrix and Knowledge canvas model. Case studies are given in this book. IDEA Matrix is represented in the following form. Knowledge canvas represents about identification of opportunity for product. Feasibility is represented w.r.t. business perspective.

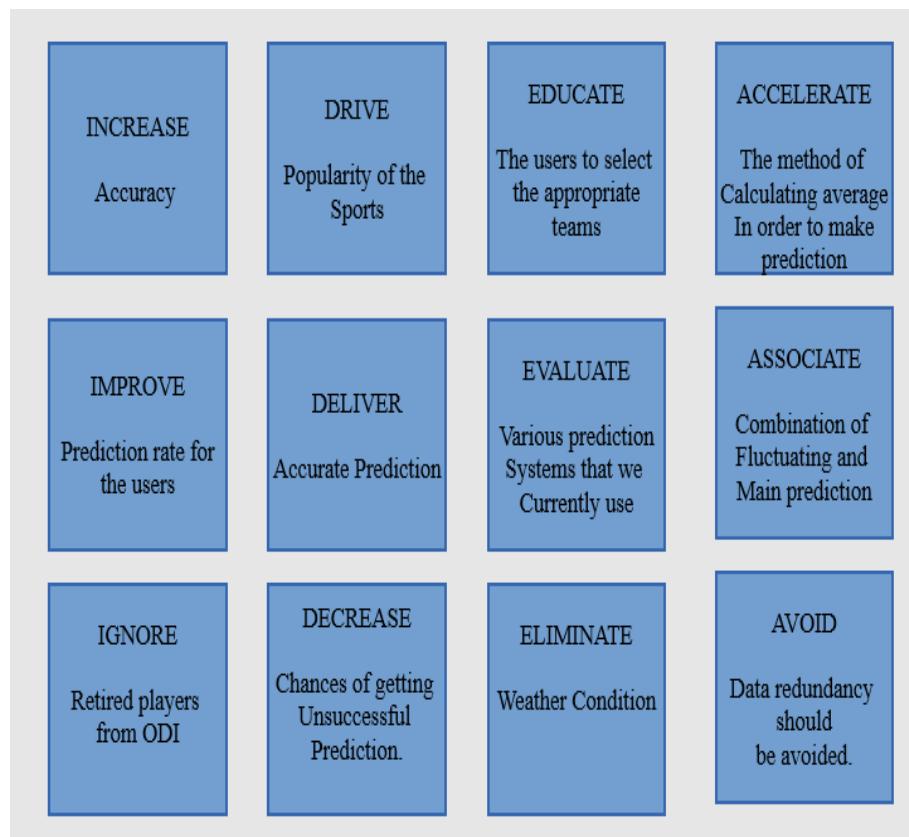


Figure A.1: IDEA Diagram

- Project problem statement feasibility assessment using NP-Hard, NP-Complete or satisfy ability issues using modern algebra and/or relevant mathematical models.

ANNEXURE B

LABORATORY ASSIGNMENTS ON

PROJECT QUALITY AND

RELIABILITY TESTING OF

PROJECT DESIGN

It should include assignments such as

- Use of divide and conquer strategies to exploit distributed/parallel/concurrent processing of the above to identify object, morphisms, overloading in functions (if any), and functional relations and any other dependencies (as per requirements). It can include Venn diagram, state diagram, function relations, i/o relations; use this to derive objects, morphism, overloading
- Use of above to draw functional dependency graphs and relevant Software modeling methods, techniques including UML diagrams or other necessities using appropriate tools.
- Testing of project problem statement using generated test data (using mathematical models, GUI, Function testing principles, if any) selection and appropriate use of testing tools, testing of UML diagram's reliability. Write also test cases [Black box testing] for each identified functions. You can use Mathematica or equivalent open source tool for generating test data.

ANNEXURE C

PROJECT PLANNER

NO	TASK	DURATION (Days)	START DATE	END DATE
1	Group Formation	1	16 Jun	17 Jun
2	Decide Project Idea Of Interest	3	20 Jun	23 Jun
3	Requirement Gathering	7	30 Jun	6 Jul
4	Literature Survey	6	23 Jul	29 Jul
5	Mathematical Modeling	9	18 Aug	26 Aug
6	Feasibility Test	10	1 Sept	10 Sept
7	Synopsis Submission	2	14 Sept	15 Sept
8	UML Diagram	4	25 Sept	29 Sept
9	GUI Design	12	03 Jan	14 Jan
10	Functionality Implementation	25	22 Jan	14 Feb
11	Testing	8	12 Mar	20 Mar
12	Documentation	4	24 Mar	28 Mar

Figure C.1: Project Planner

C.0.1 Timeline Chart

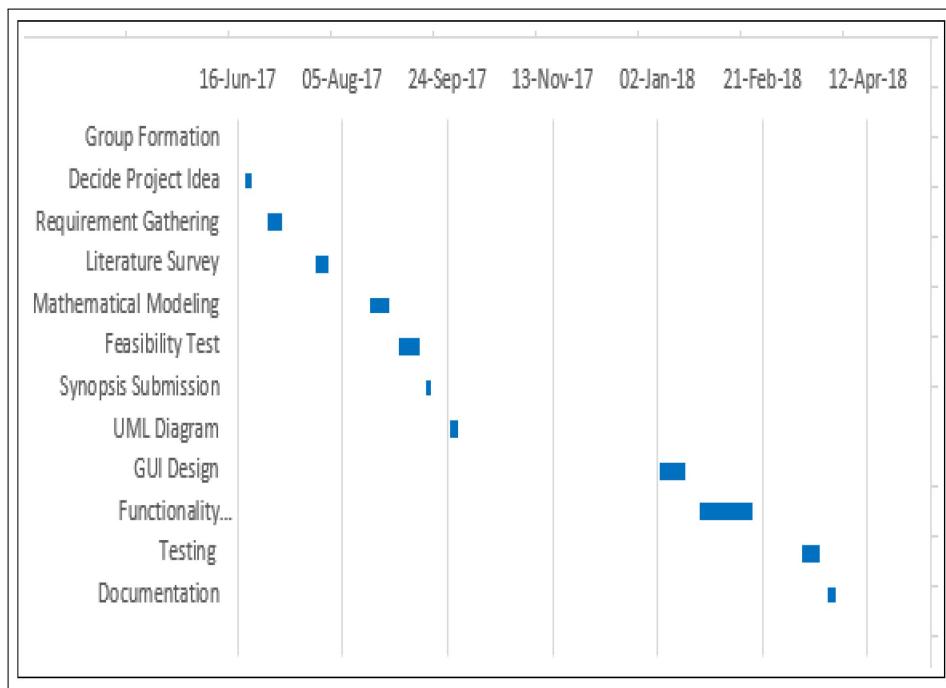


Figure C.2: Task network

ANNEXURE D

**REVIEWERS COMMENTS OF
PAPER SUBMITTED**

Survey Paper

- Paper Title: Winning Prediction Analysis in One Day International Cricket using Machine Learning Techniques.
- Authors : Abhishek Naik, Shivanee Pawar, Minakshee Naik, Sahil Muliani
- Name of the Conference/Journal where paper submitted : International Journal Of Emerging Technology and Computer Science(IJETCS).
- Reviewer Comments :

 - (a) Subject content is good.
 - (b) Technical content is good.
 - (c) Domain of the paper is good.

- Paper accepted/rejected : Accepted
- Overall evaluation : Strongly accepted

Research Paper

- Paper Title: Winning Prediction Analysis in One Day International Cricket using Machine Learning Techniques
- Authors : Abhishek Naik, Shivanee Pawar, Minakshee Naik, Sahil Muliani
- Name of the Conference/Journal where paper submitted : Journal Of Emerging Technologies and Innovative Research(JETIR).
- Paper accepted/rejected : Accepted
- Overall evaluation : Strongly accepted

ANNEXURE E

PLAGIARISM REPORT

Plagiarism report

ANNEXURE F

**TERM-II PROJECT LABORATORY
ASSIGNMENTS**

1. Initially we acquired the necessary sponsorship from, Saapthahik Daund Navanirman Newspaper.
2. After the data acquisition, a model was designed which provided overview of the system which was checked in accordance with the project guide. The next stage involved developing the server.
3. The server design and code was implemented after taking valuable input from the project guide. The server on completion of implementation was verified by the project guide and expected changes were implemented.
4. The data sets involved different parameters related to players which varied linearly. Thus, this project also implements Linear Regression algorithm along with the decision tree for accurate prediction of players performance.
5. The client side required development of the website. Different website design were studied and the various features required by our website were discussed and implemented.
6. The concluding part involved if any further features were required which was reviewed by the review panel and the project guide.
7. In report generation, an appropriate report of various players performance predicted that were selected by the user were generated. This provided an overview of the team win prediction carried out by the system.
8. In order to make the project error-free, various testing methodologies were carried out.