



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Analzyation and Prediction of ODI Data

A Report for the Project 2

**BACHELOR OF TECHNOLOGY
IN**

**Computer Science and Engineering
Submitted by**

ANKIT KUMAR NANDAN

**(1613101205)
(16SCSE101164)**

Under the Supervision of

**Dr. SANSAR SINGH CHAUHAN, Ph.D.,
Professor**



**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA-201306
MAY 2020**

DECLARATION

Project Title: CRICKET DATA ANALYTICS AND PREDICTION

Degree for which the project work is submitted: **Bachelor of Technology in
Computer science and Engineering**

I declare that the presented project represents largely my own ideas and work in my own words. Where others ideas or words have been included, I have adequately cited and listed in the reference materials. The report has been prepared without resorting to plagiarism. I have adhered to all principles of academic honesty and integrity. No falsified or fabricated data have been presented in the report. I understand that any violation of the above will cause for disciplinary action by the Institute, including revoking the conferred degree, if conferred, and can also evoke penal action from the sources which have not been properly cited or from whom proper permission has not been taken.

ANKIT KUMAR NANDAN
1613101145
16SCSE101164

DATE: 4/05/2020



SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report “**CRICKET DATA ANALYTICS AND PREDICTION**” is the bonafied work of “**ANKIT KUMAR NANDAN (1613101145)**” who carried out the project work under my Supervision.

Signature of Head

Dr.MUNISH SHABARWAL,

PHD(Management)PHD(CS)

**Professor & Dean,
School of Computing Science &
Engineering**

Signature of Supervisor

Dr SANSAR SINGH CHAHUAN,

M.Tech,PHD,

**Professor,
School of Computing Science &
Engineering**

Abstract

Winning is the goal of any sport. Cricket is the one of most watched sport now a days. Winning in cricket depends on the various factors like home crowd, advantages, performances in past, experience in match, performance at specific venue, Performance against the specific team and the current form of the team and the player. During the past few years lot of work and research papers have been published which measures the players performance and their winning predictions. In this work a model has been proposed that has two methods, first predicts the score of first inning not only on the basis of current run rate but also considers number of wicket fallen, venue of match and batting team. The second method predicts the outcome of the match in the second inning considering the same attributes as of former method along with the target given to the batting team. The methods have been implemented using Linear regression Classifiers or Q- Learning base decision tree approach or Naïve Bayes classifiers for first inning and second inning respectively. In both methods, 5overs interval have been made from 50 overs of the match and at each interval above mentioned attributes have been recorded for all non-curtailed matches played between 2002 and 2014 of every team independently.

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B.Tech. Final Year. We owe special debt of gratitude to

Dr.Sansar singh Chauhan

Department of Computer Science & Engineering, Galgotias University,Gr. Noida for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day. We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Table of Contents

CHAPTER 1 INTRODUCTION TO PROJECT

- 1.1 OVERVIEW
- 1.2 FACTORS
 - 1.2.1 Pitch
 - 1.2.2 Toss
 - 1.2.3 Teamstrength
 - 1.2.4 Home Ground Advantage
- 1.3 PROBLEM STATEMENT
- 1.4 RELATED WORK

CHAPTER 2

- 2.1 METHODOLOGY
 - 2.1.1 Data Set Generation
 - 2.1.2 Data Ser Description
 - 2.1.3 Data Cleaning
 - 2.1.4 Attribute Selection

CHAPTER 3-

- 3.1 CLASS DLAGRAM
- 3.2 DATA FLOW DIAGRAM
 - 3.21 DFD Level
 - 3.2 2 DFD Level 1

CHAPTER 4-

- 4.1 MOTIVATION
 - 4.2 SCOPE
 - 4.3 PROPOSED MODEL
 - 4.4 TECHNOLOGIES USED
 - 4.4.1 JAVA
 - 4.4.2 MySQL
 - 4.5 ALGORITHM
- NAIVE BAYES
- 4.6 HARDWARE AND SOFTWARE REQUIREMENT
 - 4.61 Hardware Requirement

4.6.2 Software Requirement

CHAPTER 5.

5.1 MODULE SPLIT-UP

5.1.1 Adimi module

5.1.2 Users Module

FIGURE 12: RESULT 4.2 ADMIN PAGE

4.2 ADMX PAGE

4.3 REGISTRATION MODULE

FIGURE 15: REGISTERED USERCONCLUSION..

CONCLUSION

FUTURE WORK

REFERENCES

LIST OF FIGURES

1. METHODOLOGY DIAGRAM
2. CSV FILE
3. CLASS DIAGRAM
4. DFD LEVEL 0
5. DFD LEVEL 1
6. PROPOSED MODEL
7. ALGORITHM COMPARISON
8. ADMIN LOGIN
9. ADMIN ACCOUNT MODULE
10. USER LOGIN
11. PHP HOME PAGE
12. REGISTRATION WINDOW.
13. REGISTERED USER

CHAPTER 1

INTRODUCTION TO PROJECT

1.1 Overview

Cricket is being played in many countries all around the world. There are lots of domestic and international tournaments being held in many countries which play cricket. Cricket is a game played between two team comprising of 11 players in each team. The result is either win, loss or tie. However, some time due to bad weather conditions the game is also washed out as cricket is a game which cannot be played in a rain. Moreover, the game is also extremely unpredictable because at every stage of the game the momentum shifts to one of the team between the two. A lot of times the results gets decided on the last ball of the match where the game gets really close. Considering all these unpredictable scenarios of this unpredictable game, there is a huge interest among the spectators to do some predictions either at the start of the game or during the game. Many spectators also play betting games to win money. So, keeping in mind all the responsibilities, this report aim sat studying the problem of predicting the game results before the game has started based on the statistics and data available from the dataset. These are the different way to do prediction. The prediction can be done taking into consideration the players performance as well as team performance. These are many unpredictable things that happen in a cricket game like matches being washed out due to rain, a key plyer getting injured before the game, players changing their teams, etc. sometimes a key player also gets injured during the game and is not able to take further part in the game. All these factors do affect the prediction to some extent. The report discusses a methodology that I followed for the game result prediction. The methodology consists of first the attribute selection algorithms which trim down the list of attributes to only important ones and the data mining algorithms which can be applied on those attributes. The game prediction problem that I studying does not take into consideration the players performance but it does not take into considerations the team past performance at high level extent along with the other factors like toss winner, toss decision, home support,etc.

1.2 Factors

Cricket winning can be predicted like all other games. We need to find the best attributes or factors that influences the match outcome. The result of a cricket match depends upon more of in-game and more of pre-game attributes. Pre-game attributes like pitch, Team strength, weather, venue etc. and in-game attributes like run rate, total run, strike rate, wickets in hand etc. influences a match result predominantly. Below are the attributes that decides out come of the cricket match.

1.2.1 Pitch

Unlike other sports, cricket stadiums shape and size is not fixed except the dimensions of the inner circle and pitch which are 30 yards and 22 yards respectively. Out field variations and pitch can have a substantiate effect on bowling and batting. The spin of the ball, seam movement and the bounce depend upon the nature of the pitch. It depends on the how the wet is the pitch. The more wet the pitch, the slower it will play. On the off chance that it is drying out, those balls will change significantly, yet all it will get less difficult those drier itgets.

Green pitches tend to get easier to bat on. Wickets can get significantly more dry or wet (on the off chance that it downpours) They might start to break up if they are soft (which would help the bowler).

1.2.2 Toss

According to cricket analysts, there is sure measure of advantage for a team if it wins the toss. This might not be deciding factor in a match but it would give the team the opportunity ofchoosing “what they want”.

1.2.3 Team Strength

The team strength should be balanced for winning a match. Captainship in a team is also a deciding factor. Past records: The past team performances can be consider to predict the out comes of match. History of games at that venue show did the team performs, performances at that specific venue, performance against the specific opposition and experience at the specificvenue.

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

In this to design a system that can provide the Score and Winning Prediction in Cricket match, the system can analyze multiple parameters like winning toss, batting side, DL approach, Home ground advantages, player wise performance etc. While declaring a time for the particular championship it is very important to select the best team so that the chances of the team winning the championship become easy. This problem had to be solved to generate the best players from both the teams for the best battle. To solve this problem we have collected the historical data of all some team like (India, Pakistan, Australia, New Zealand etc.), and using prediction algorithm like Naïve Bayesian algorithm we are predicting the best starting players for both the teams that can be used in fantasy league for winning the maximum points.

1.4 Related work

There has been a lot of related study to this problem in various different sports. The papers I have used as references are all related work that had been done on this problem. The paper by Trawinski described the prediction of results using a fuzzy classification system. This paper was predicting the results for basketball games. I had used the attribute selection techniques mentioned in this paper for my project. The attribute selection technique proposed in this paper was done using WEKA so it was a good reference point for me too. The wrapper method algorithms and the ranker method algorithms implemented in this paper were also used in my project. But the prediction part was done using the fuzzy classification system and I did not use that system for my prediction part.

The paper by Haghghat also described the prediction of results in sports using the data mining techniques. But this paper was not specific to any particular sport, rather it was for in general all sports. The attribute selection algorithm that it used was more of an elimination approach where the attributes were eliminated one by one and classification accuracy is computed. Once a good

subset of attributes is achieved, then the eliminated attributes are again added one by one to see if the accuracy improves. But, in my approach I did not use this elimination approach for the attribute selection part, then the paper used various data mining algorithms to perform the classification. I used the Naive Bayes and the decision tree algorithms from the paper in my project and compared my accuracy with that of the paper.

The paper by Zdravevski takes into consideration some of the external factors like the number of players injured before a particular game, the winning streak of a team before a particular game, the fatigue factor of the players, etc. The approach that I used in the project does not take into account the external features like player injury, fatigue or the winning streak. My data set contains more data about the matches and the events happening in the match like toss and player of match rather than the data about external factors.

CHAPTER 2

Methodology

The following methodology are consider in our project. Here 5 phases have been categorized i.e. Generation of information sets, Cleaning of data, Attribute determination data mining and evaluation of result. Each phase would be discussed in the detail:

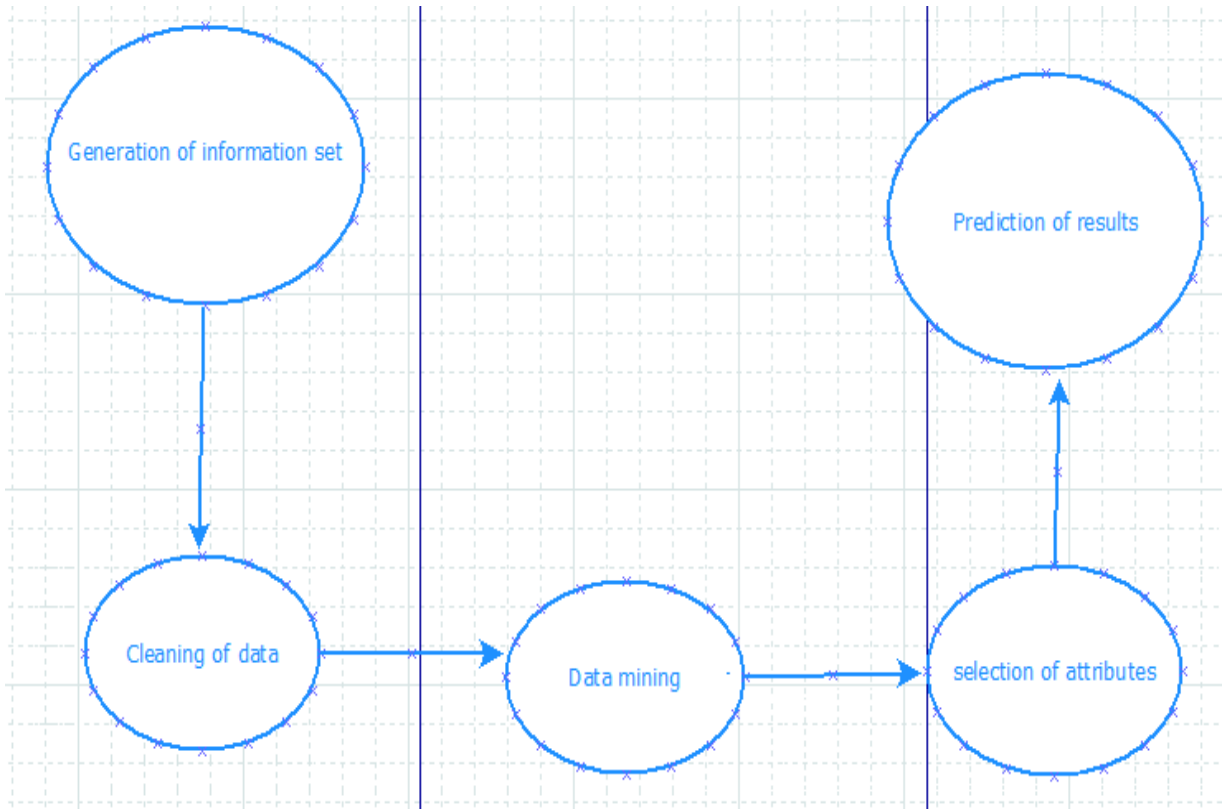


Figure 1: Methodology Diagram

2.1.1 Data Set Generation

The data was collected from the <http://cricsheet.org> website. The website has data about all matches (from 2008 to 2018) of international tournaments held. It is a ODI format of tournaments. It means that each team bats or bowls for maximum 50 overs each. The data set downloaded from this website was in the .CSV format. I used the java classes file and the file writer to read the .CSV file and write the content to new file. The .CSV file had data about 4917 different entries and the java code basically copied the important data from all the 4917 files had then combined multiple files into a single file. The data from the .CSV file contains more irrelevant information like gender, date, umpire information etc. which were all discarded and the java code copied only data about the team, venue, toss, toss _winner, player _of_ match and thee ball by ball data. The .CSV file had data shown in the figure.

In Computing, Commas separated values (CSV) file is a delimited text file that uses a comma to separate values. A CSV files stores tabular data (numbers and text) in a plaintext. Each line of the file is a data record. Each record consists of one or more fields separated by commas. The use of commas as a field separator is the source of the name for this file format

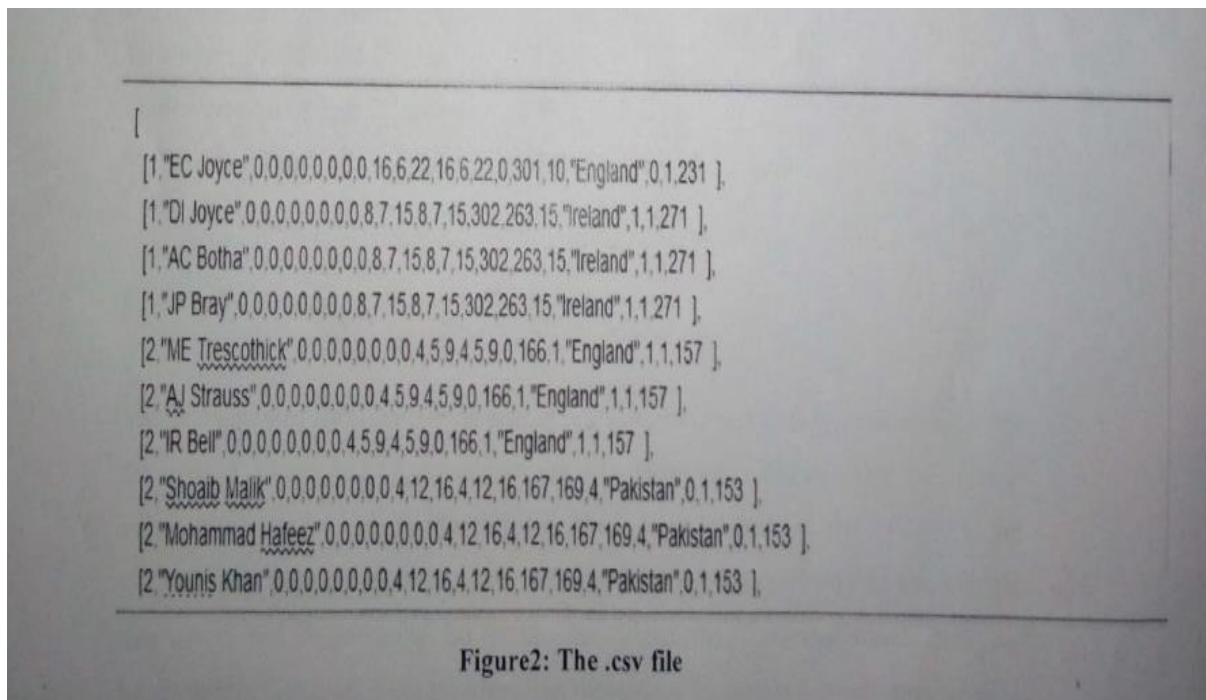


Figure 2: The .csv file

2.1.2 Data Set Description

The new combined data set that I generated had the data from the .csv files. The data set consists of 5 different attributes and 4917 different instances. The data set spans across all the seasons of the One Day International from 2008 to 2018 season. The winning Team is the classifier in the data set and this project is to predict the winning team in the match.

The attributes in the data set are as follows:

- 1). Venue: Playing venue.
- 2). Team1: The Playing team 1.
- 3). Team2: The Playing team 2.
- 4). Player: Add Player 1 team.
- 5). Player: Add Player 2nd team.
- 6). Winning Team: This is the class attribute i.e. the winning team.

2.1.3 Data cleaning

The data obtained from the <http://cricsheet.org> website was already cleaned. So, I did not have to do any sort of cleaning on the data. However, I had to tackle the missing values data and the data for the matches which were washed out due to rain. Those matches data were present in the .json file. So after combining the two data into a single data set, I had to manually fill in the data about the 7 missing teams from the <http://espncricinfo.com> website. Moreover there were 10 matches which were washed out so those instance were filled with null values and they were discarded from the data set. So, the final dataset had 4917 instances with 21 attributes

2.1.4 Attribute Selection

The data set obtained after handling the missing values had to be filtered with the help of the attribute selection algorithm. Since, there were 21 attributes it was necessary to identify all important attributes which would be useful for the data mining tasks. The paper by Haghighit [3] explained the attribute elimination process. Here, first a data set with some number of attributes is selected and then each attribute is eliminated one by one from the set of attributes. The elimination is based upon the result of running the classification algorithm on the set of attributes. An attribute is completely eliminated if the accuracy improves after its removal or else the attribute is kept in the data set. So, at the end of this elimination process we get a set of attributes using which we get the highest accuracy of prediction for the classification algorithms. The paper by Trawinski[2] describes the two types of selection algorithms i.e. the wrapper method and the ranker method that were used by the authors in their study. I used this paper as my reference for the attribute selection phase. I wanted to evaluate the worth of each attribute and rank each attribute to identify their importance before using them for making the prediction during the data mining phase. So I decided to choose these two types of attributes selection algorithms from the papers by Trawinski [2]. These two types of attribute selection methods are available to use in WEKA. After the attribute selection phase, The list of attributes gets minimized to 10 i.e. Team 1, Team 2, Venue, Home _ Team, Toss _ winner, Toss_ decision, Player _ of _ match, Team _ batting _first, Team _batting _ second, and winning _ team. Here the winning _ team is the classifier.

CHAPTER 3

3.1 Class diagram

The class diagram for the proposed system describes the system in terms of classes, attributes, operations, and their associations. In UML, classes and objects are shown by boxes composed of three compartments. Top compartment displays the name of the class or object. The Centre compartment displays its attributes; the bottom com displays its operations

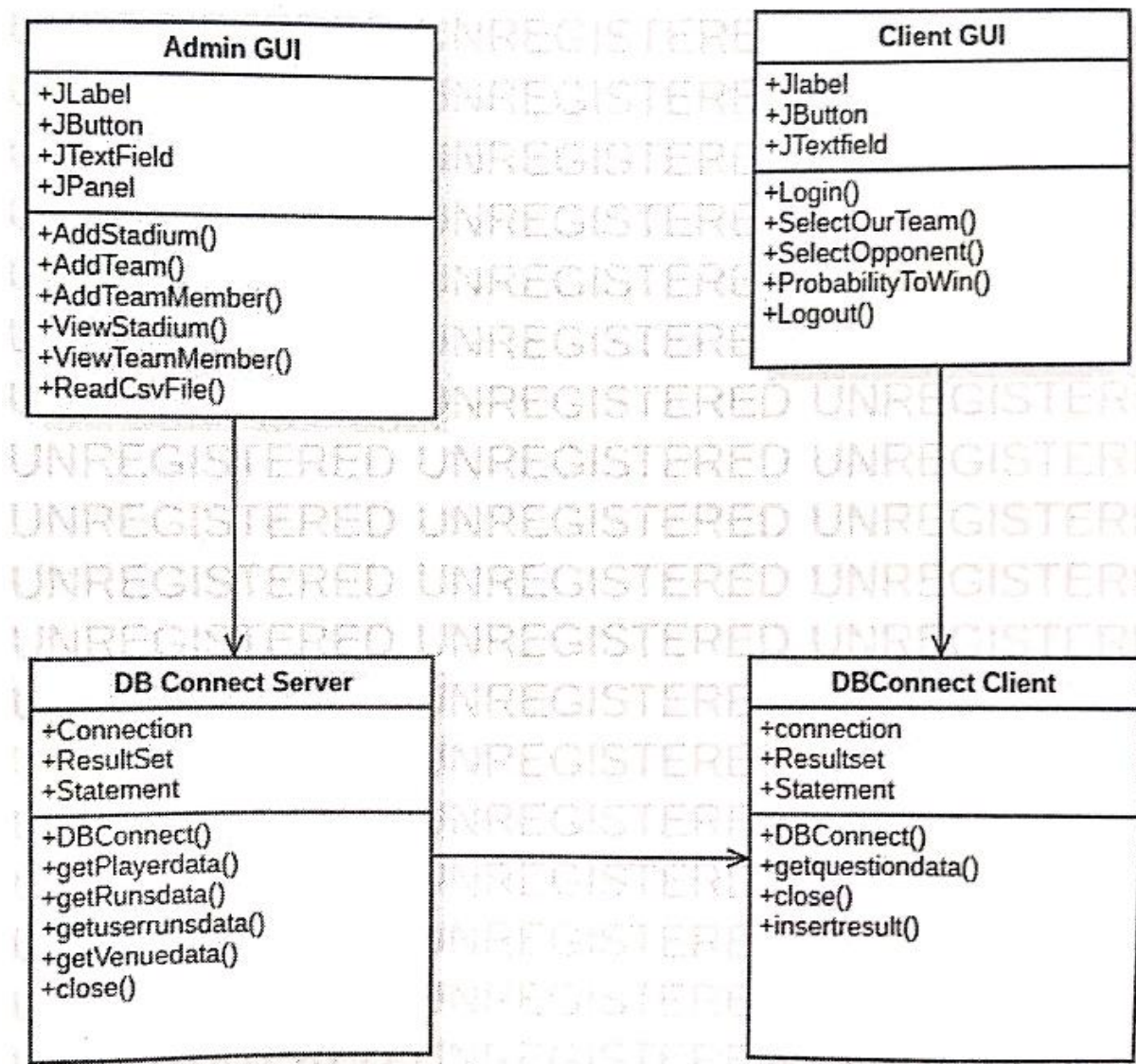


Figure 3: Class Diagram

3.2 Data Flow Diagram

A data flow diagram is a graphical representation of the flow of data of CSPS. It helps us to understand the flow of system.

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modelling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated. Data flow diagrams are one of the three essential perspectives of the structured-systems analysis and design method SSADM.

3.2.1 DFD Level 0

DFD Level 0 is also called a Context Diagram. It's a basic overview of the whole system or process being analyzed or modeled. It's designed to be an at-a-glance view, showing the system as a single high-level process, with its relationship to external entities. It should be easily understood by a wide audience, including stakeholders, business analysts, data analysts and developers.

From the above Data Flow Diagrams we get a detailed understanding of how the flow of each function will take place. DFD level 0 gives the overview of CSPS.

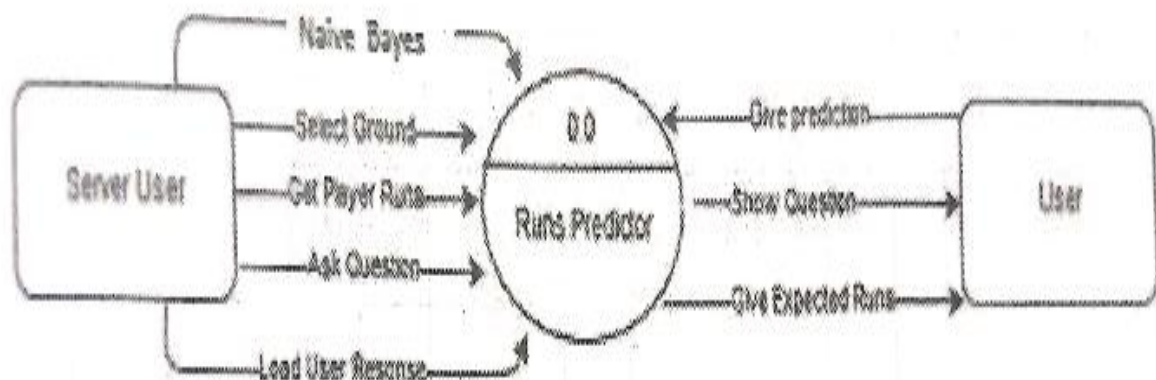


Figure 4: DFD Level 0

3.2.2 DFD Level 1

The Level 0 DFD is broken down into more specific, Level 1 DFD. Level 1 DFD depicts basic modules in the system and flow of data among various modules mentions basic processes and sources of information.

- It provides a more detailed view of the Context Level Diagram.
- Here, the main functions carried out by the system are highlighted into its sub-processes.

Whereas DFD Level 1 of the different functions give a more in depth knowledge of the methods and database usage. The DFD give us a good visualization of the steps involved in functions of the user and the database that is connected to the various methods.

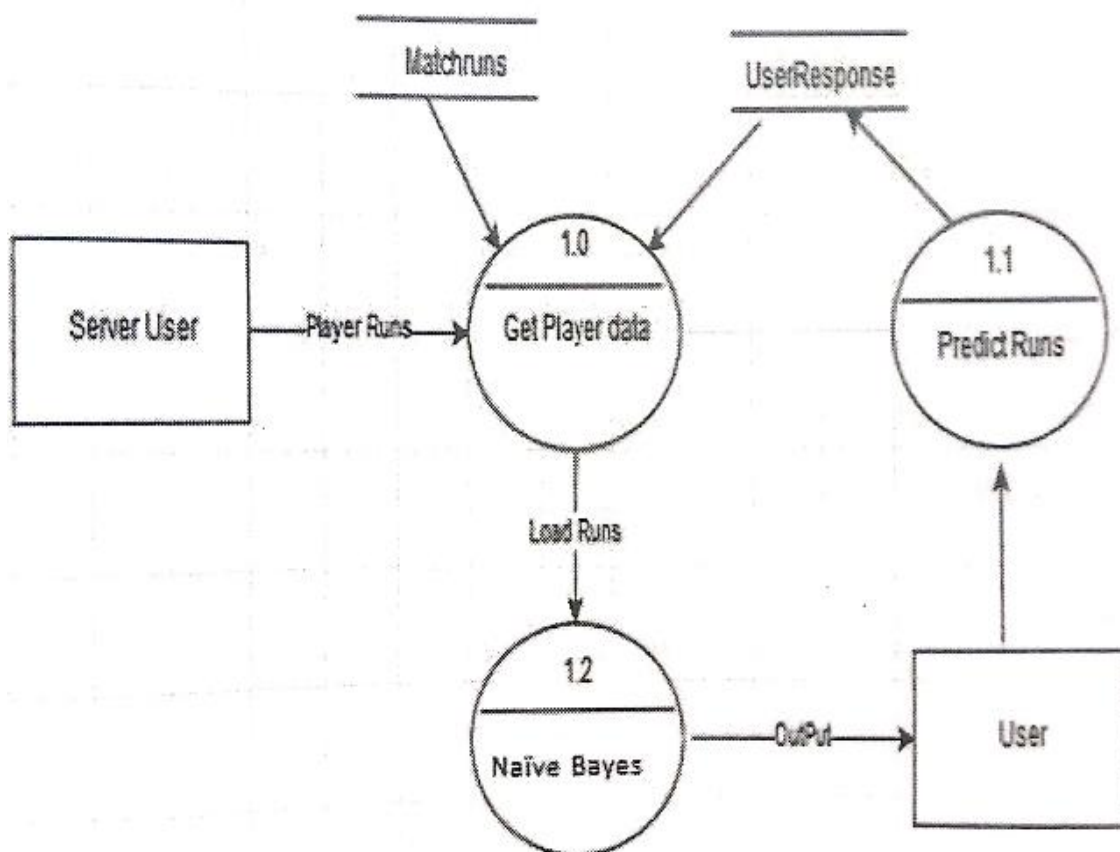


Figure 5: DFD Level 1

CHAPTER 4

4.1 Motivation

There are several areas in our day-to-day life when we need such analysis of data and some of them are listed below that arises the urgent need for such software.

- For instance there are users who bet on the application DREAM11 and there this software can be useful as it will give all information about every player so that the individual can choose his/her team efficiently and it will be productive.
- The commentators or analysts can also use it to demonstrate the stats and the records of the players on a specific field. It can also be useful for some youtubers to analyze the play of some players in their videos.
- It can also be used on a small scale like the inter school tournaments, district level tournaments to analyze the game of various teams and so to improve the game.

4.2 SCOPE:

- It is basically used for the analysis that at which particular ground the winning percentage is higher of a particular team.
- It can be used in predicting the teams on the application like DREAM 11 where people choose their teams and compete with each other to get prizes. It will help the person to form the team of such players that are compatible with the playing conditions.
- It can be used to predict the overall winning chances of a team in a particular match on particular ground.
- It can be used by team management to analyze and make such team that can give higher percentage of winning.

4.3 Proposed Model

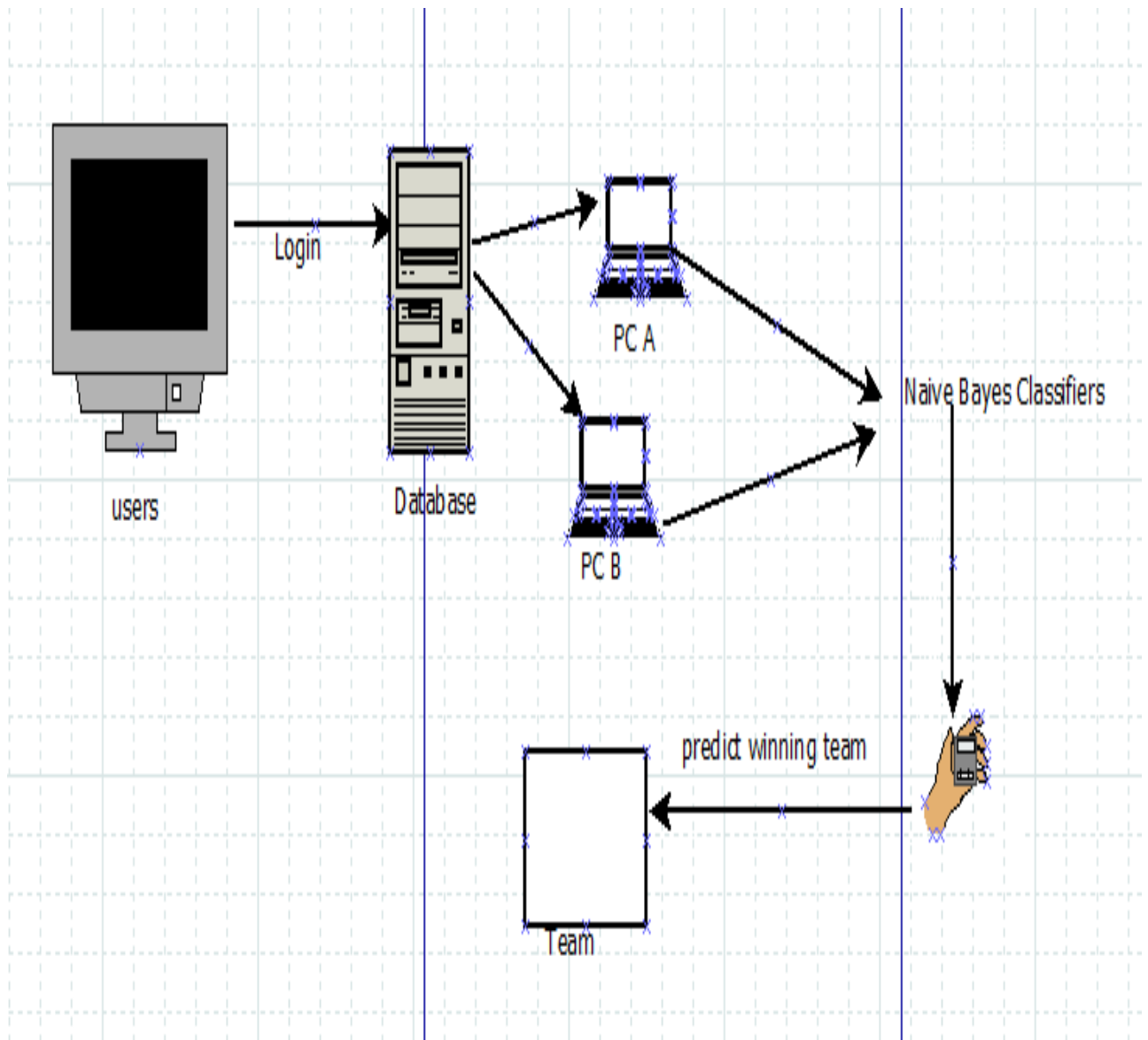


Figure 6: Proposed Model

Technologies Used

4.4.1 JAVA

Java programming language is concise which makes it easy to use and learn. Java virtual Machine (JVM) enables java to be executed in any environment and platform making it portable. Web applications and applets can be accessed in a secure way using java. It is an object oriented and supports multithreading. Java supports cross platform optimized code called bytecode which are faster to execute. Hence it gives high performance.

4.4.2 My SQL

MySQL is an open source database management system suitable for relational data. MySQL is a popular choice of database for use in web applications. It has high availability and can also run on cloud computing platforms. Administration of My SQL is handled with the use of phpMyAdmin which is a free and open source tool for web browser.

4.4.3 Algorithm

Naïve Bayes

The Naive Bayes algorithm is a probabilistic classification algorithm which is based on the Bayes Theorem of conditional probability. The Naive Bayes algorithm assumes that all the attributes in the data set are independent of each other and then computes the probabilities of each attribute. The Naïve Bayes algorithm with a Percentage Split method divides the data set into 70% Training data and 30% Test data and gives an accuracy rate of 57.06%. Here, this accuracy improves only slightly when we use the 10- Fold Cross Validation method.

Input: User input file data record which contains (CScore,Overs, WF), segment of average score from train database of k to n over's.

Output:Project score S_c

Step 1: Read $R\{ \text{Runs, Overs WF} \}$ from current parameters.

Step 2: Map with train features with each sample.

Step 3: calculate average score of train DB with same evidences

$$AvgTScore = \sum_{k=0}^n (S_c)$$

Step 4: PreScore = (Score +Avg TScore)

Step 5: Return PreScore

The accuracy rate of the model on the test data set was used as the evaluation criteria. The table below gives the accuracy rate of each of the algorithm along with how it performed using both the Percentage Split method as well as the K-Fold Cross Validation method.

Algorithm	Percentage Split	K-Fold Cross Validation
Decision Tree (J48)	31.07%	63.24%
Random Forest	48.02%	71.08%
Naïve Bayes	57.06%	60.97%
KNN	52.54%	51.39%

Figure 7:Algorithms Comparison

The model gives an accuracy rate of about 60% to 70% when using the K-Fold Cross Validation whereas the accuracy rate in the case of the Percentage Split is not very good. This is mainly because of the fact that in Percentage Split method, WEKA splits the data manually on the basis of the 70% training and 30% test split that I mentioned. The K-Cross Validation method overcomes this by dividing the data set into K different subsets and thereby computing the results for K different subsets. We can notice that the accuracy rate almost doubles in some algorithms as shown above. The Random Forest algorithm gives the best accuracy rate for the model with 71.08% accuracy rate.

4.5 Hardware And Software Requirement

4.5.1 Hardware Requirement

Processor-DualCore

Speed-1.1GHz

RAM-512 MB (min)

Hard - 20 GB

KeyBoard - Standard Windows Keyboard

Mouse - Two or Three Button Mouse

4.5.2 Software Requirement

Operating System: Windows XP,7,8,10

Front End : Java 7

Technology: Mysql connector java, XAMPP, JDK1.8.0

IDE: Netbeans

CHAPTER 5

5.1 Module Split-up:

Whole project is divided into two module,

- Admin Module
- Users Module

5.1.1 Admin module

Admin module has several sub module.

- First sub module is admin Login Panel. In which admin enter their user name and password to control different function of a system.



Figure 8:Admin Login

- After entering correct username and password, admin get access Admin Account Module. In this module admin get various option like Add Stadium, Add Team, Add Team Members, View Stadiums, View Team Members, Read CSV File. All this option helps admin to run system properly for their users.

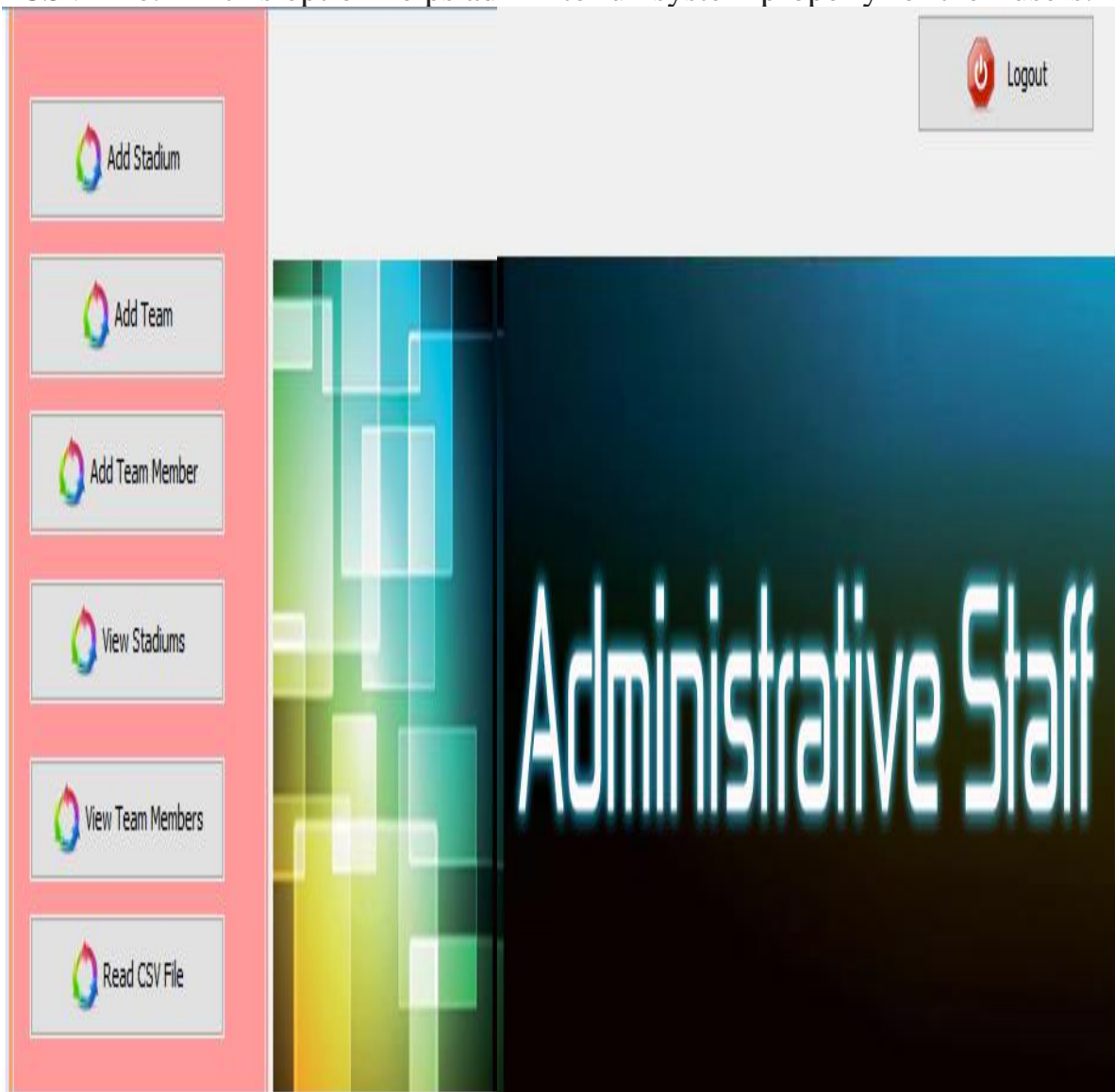


Figure 9: Admin Account Module

5.1.2 User Module

Users module is also divided into several sub module.

- First sub module is user login panel. In which user enter their registered their username and password to get the benefits of our system.



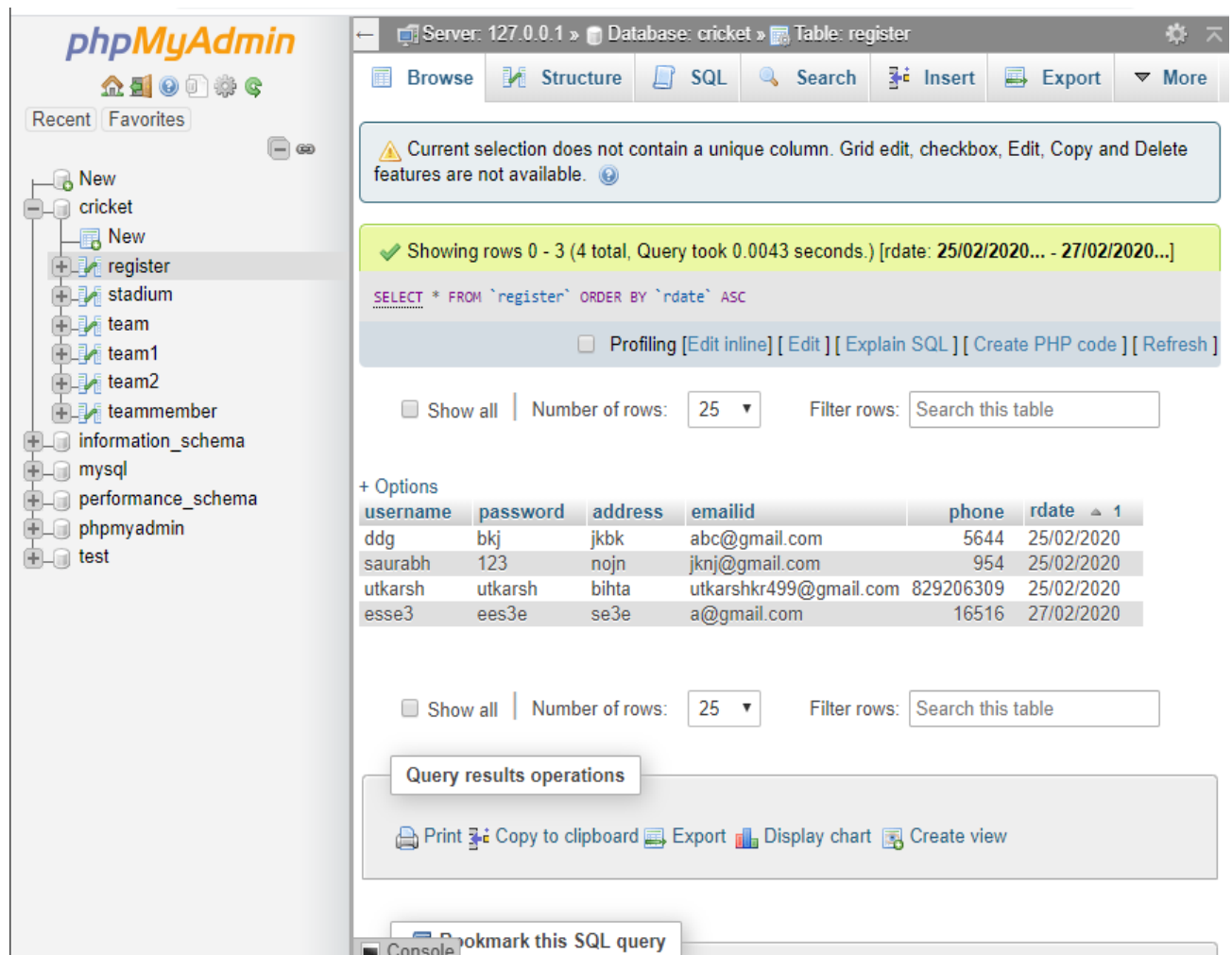
Figure 10: User Login

- After entering correct username and password, user get access to the user account module. In this Module User get various option like Select Team, Select Opponents, Probability to win.

- After selecting two team user click on Probability to win option to see the upcoming match Probability i.e which is going to win.

5.2 Admin Page

PHPMyAdmin is open source free software designed to handle the administration and management of MySQL databases through a graphic user interface. Written in PHP, PHPMyAdmin has become one of the most popular web based MySQL management tools. PHPMyAdmin comes with detailed documentation and is being supported by a large multi-language community. PHPMyAdmin's ever growing list of features supports all commonly used operations such as browsing, dropping, creating, altering MySQL databases, tables, fields and indexes. Also, PHPMYAdmin enables you to manage MySQL users and user privileges. Another commonly used PHPMY Admin feature is its import function. With PHPMyAdmin. MySQL database import from backup is made easy and you can import an SQL or CSV dump with a few mouse clicks. Also, you can easily export your database in CSV, SOL XML Excel and other popular formats.



The screenshot displays the PHPMyAdmin interface. On the left is a navigation tree with a 'cricket' database selected, containing tables like 'register', 'stadium', 'team', etc. The main area shows the 'register' table with a warning message: 'Current selection does not contain a unique column. Grid edit, checkbox, Edit, Copy and Delete features are not available.' Below this, a green status bar indicates 'Showing rows 0 - 3 (4 total, Query took 0.0043 seconds.)'. The SQL query is 'SELECT * FROM `register` ORDER BY `rdate` ASC'. The table data is as follows:

username	password	address	emailid	phone	rdate
ddg	bkj	jkbk	abc@gmail.com	5644	25/02/2020
saurabh	123	nojn	jknj@gmail.com	954	25/02/2020
utkarsh	utkarsh	bihta	utkarshkr499@gmail.com	829206309	25/02/2020
esse3	ees3e	se3e	a@gmail.com	16516	27/02/2020

At the bottom, there are 'Query results operations' including Print, Copy to clipboard, Export, Display chart, and Create view. A 'Console' tab is visible at the very bottom.

Figure 11 Admin Page

5.3 Registration Module

In this module new user can do registration to get access for user client page.

Cricket Team Prediction

Registration Panel

REGISTRATION PANEL

USER NAME

PASSWORD

ADDRESS

EMAIL jLabel11

MOBILE jLabel11



 Register  Cancel

Figure 12: Registration window

Admin can also see all registered users from phpMyAdmin page. Below figure shows all registered users.

The screenshot displays the phpMyAdmin interface for a database named 'cricket'. The left sidebar shows a tree view of databases and tables, with 'cricket' expanded to show tables like 'register', 'stadium', 'team', 'team1', 'team2', and 'teammember'. The main panel shows the 'register' table structure and a query result. A warning message states: 'Current selection does not contain a unique column. Grid edit, checkbox, Edit, Copy and Delete features are not available.' Below this, a green bar indicates 'Showing rows 0 - 3 (4 total, Query took 0.0043 seconds.) [rdate: 25/02/2020... - 27/02/2020...]'. The SQL query is 'SELECT * FROM `register` ORDER BY `rdate` ASC'. The query results are displayed in a table with columns: username, password, address, emailid, phone, and rdate. The results show four rows of user data. Below the table, there are options to 'Show all', 'Number of rows: 25', and 'Filter rows: Search this table'. At the bottom, there are 'Query results operations' including 'Print', 'Copy to clipboard', 'Export', 'Display chart', and 'Create view'. A 'Console' tab is visible at the bottom left.

username	password	address	emailid	phone	rdate
ddg	bkj	jkbk	abc@gmail.com	5644	25/02/2020
saurabh	123	nojn	jknj@gmail.com	954	25/02/2020
utkarsh	utkarsh	bihta	utkarshkr499@gmail.com	829206309	25/02/2020
esse3	ees3e	se3e	a@gmail.com	16516	27/02/2020

Figure 13: Registered User

CONCLUSION

In our proposed Model which is used to do forecasting of ODI data. we have the total attributes available were 21 which is then now reduced to 10. we have used the concept of naïve Bayes theorem to forecast more better results. we have required large number of instances to make a results better. Our future work is to take large number of insatnces to get accurate to predicted the results.

FUTURE WORK

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

- We can also improve our accuracy by regular cheaking of injuries and fatigue and continuous run of matches etc.
- We can also improve by players form in that tournament and opponent team, away condition and no of man of the match etc.
- There is no web/mobile application or UI that my project contains. So, a web/mobile application can be made which would take in the entire data set as input and display the prediction result for each of the instances to a pdf or text file.

REFERENCES

- Dataset: <http://cricsheet.org> [online Accessed].
- <http://espncricinfo.com/> [online accessed].
- Pramila M. Chawan. “Prediction of live cricket score and winning”, (2018) International journal of trend in research and development volume 5(4), ISSN:2394-9333, www.ijtrd.com.
- <https://aajtak11.in/> [online Accessed].
- Haghghat, Maral, Hamid Rastegari, and Nasim Nourafza. “A review of data mining techniques for result prediction in sports.” Advances in computer science : an International Journal2.5.