**MINI PROJECT REPORT**

**ON**

**IPL 2008-20 Python analysis**

**BY**

**1.Omkar Ganore**

**Course: MCA I Sem.: II**

**Roll No.: 21066**

**PRATIBHA INSTITUTE OF BUSINESS MANAGEMENT**

**CHINCHWAD, PUNE-411019**

**2021-2022**

**DECLARATION**

**PRATIBHA INSTITUTE OF BUSINESS MANAGEMENT**

**CHINCHWAD, PUNE-411019**

**2021-2022**

**Master in Computer Application**

This is certify that Mr. Omkar Ganore(66)

have successfully completed Project titled

“IPL 2008-20 Python analysis” for MCA (Computer

Application) Sem-II in academic year 2021-2022.

**Project Guide Head of Department Mr.Shubham Nagure (Master In Comp.Application)**

**External Examiner Internal Examiner**

**ACKNOWLEDGEMENT**

Words fall insufficient to describe the efforts taken by all the staff member and colleagues for their encouragement and willing co-operation for the completion of this project. We would like to place our sincere gratitude their worthy suggestions and timely cooperation “Defeat is not when you fall down; it is when you refuse to get up”. We faced many difficulties during our project tenure, right from the requirements Gathering to implementation. We cannot retain our self from according heartfelt thanks to **Mr.Shubham Nagure** our project guide cum mentor for sailing us through various odds at various stages of this project without his guidance this project wouldn’t have reached this stage. Apart from the efforts of me, the success of any project depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project. Last but not least; we are thankful our friends for their innovation suggestions which helped us in completing this project.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. no.** | **topic name** | **page no.** |
| **1** | **CHAPTER 1: INTRODUCTION**  1.1 Introduction  1.2 Existing System and need for System  1.3 Limitations of existing system |  |
| **2** | **CHAPTER 2: PROPOSED SYSTEM**  2.1 Problem statement/s  2.2 Objectives of proposed system  2.3 Scope of the system  2.4 Module specifications  2.5 Detail Description of Technology Used  2.6 Operating environment |  |
| **3** | **CHAPTER 3: SYSTEM ANALYSIS & DESIGN**  3.1 Use Case Diagrams  3.2 Sequence Diagram  3.3 Activity Diagram  3.4 User Interface Design (Input Screens etc.) |  |
| **4** | **CHAPTER 4 : Implementation**  4.1 : user interface screens  4.2 :output reports with data (if any) |  |
| **5** | **Chapter 5 : Drawbacks and Limitations** |  |
| **6** | **Chapter 6 : Conclusion** |  |
| **7** | **Chapter 7 : Bibliography** |  |

**Chapter 1. INTRODUCTION**

**1.1 Introduction**

Data science is the study of data to extract knowledge and insights from the data and apply knowledge and actionable insights. In this project, we will work on IPL Data Analysis and Visualization Project using Python where we will explore interesting insights from the data of IPL matcheslike most run by a player, most wicket taken by a player, and much more from IPL season 2008-2020.So if you are an IPL cricket fan and love data analysis with Python this project is perfect for you.Every sporting event today generates a lot of data about the game, which is used to analyze the performance of players, teams, and every event of the game. So the use of data science is in every sport today. Currently, IPL is one of the popular sporting events being held in India. So in this project we will use NumPy and Pandas libraries of Python for data analysis and for data visualization Seaborn and Matplotlib libraries. In this project we are perform some analysis process means with python using pandas and numpy using some mathematical operations and better understanding for use matplotlib and seaborn using visualize some graph and learn some new thing.

**1.2 Existing System and Need for System**

The current system is manual. All function is carried out manually. The data are coming in large number in so the management is unable to manage the all things manually. Retrieving records, making recipes, etc. is done manually. Manually and record of all these are maintained in registers, which also required a lot of space and lot of paper. The information about the run ,wicktes matches records is maintained as record in files. Hence the existing system contains lot of paperwork which takes a lot of time and also required lot of documentation and registers of records. Therefore, it is difficult to handle all conceptual data. This affects management satisfaction so there is need of computer-based system to overcome all this issues.

There are many different analysis software are availbel in market like microsoft excel. Those systems do not offer all the functionality that is needed for a analsis.So in this project we will use NumPy and Pandas libraries of Python for data analysis and for data visualization Seaborn and Matplotlib libraries. In this project we are perform some analysis process means with python using pandas and numpy using some mathematical operations and better understanding for use matplotlib and seaborn using visualize some graph.

**1.3 Limitations of existing system**

1. Lack of data not able to manage in a single excel sheet

2. Time consuming.

3. In the IPL we observe that only the current run-rate

is used to predict the final scorewhich is not an efficient way,

since there are many other factors which can affect the

projected score

4. Manual work and consumes large volumes of data

5. Lack of data security

**CHAPTER 2: PROPOSED SYSTEM**

**2.1Problem statement/s**

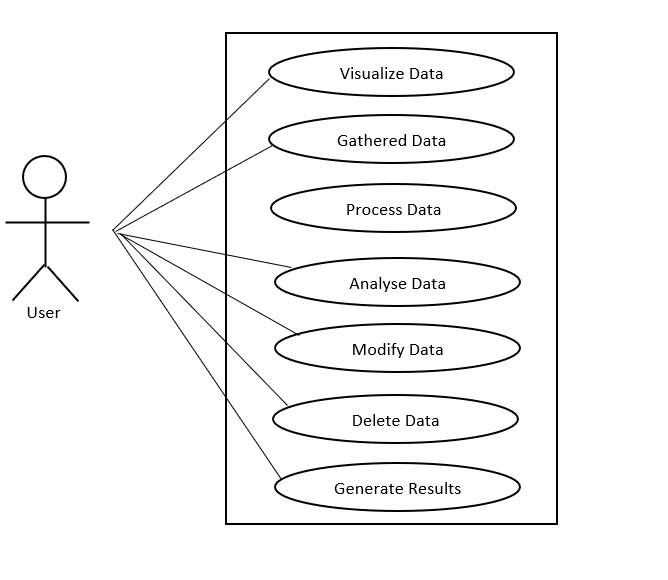
The proposed system uses Random Forest algorithm for Score Prediction of an innings and Logistic Regression for win prediction. The system scrapes data of all IPL matches from 2008 to 2020 from espncricinfo.com. And then we perform Data Cleaning followed by Data Preprocessing. Data Analysis and Visualization is done to make better understanding of the results and exploring various valuable insights. A model is deployed using one of the salesforce platforms, Heroku. From the available data,we can analysedifferentfields of the game by obtaining the statistics from which we can interpret some useful and interesting results. These visualizations can be helpful for the teams and players to understand the areas of improvement and to plan new strategies against opponents.

**2.1Objectives of proposed system**

The main objective of this project is to give the team players information about how each venue makes a difference to the game. And give feedback of how the players can improve their own performance in each game. And also give have a better planning of how the match should be played overall by the whole team regardless of the toss decision. From the available data,we can analyse different fields of the game by obtaining the statistics from which we can interpret some useful and interesting results. These visualizations can be helpful for the teams and players to understand the areas of improvement and to plan new strategies against opponents. We have analysed the performance of teams in death overs by comparing the average number of runs scored per over by teams vs the average number of runs given per over by that team in the death overs. We have used a grouped bar plot to visualize this data. Each teams’ average runs scored per over vs average runs given per over are plotted side by side as a group to show a better visualization of all the teams’ performances. We need to fill the data of the current situation of the match as shown in the above image. To predict the first innings score, we need to provide the current over, current score, current wickets, names of striker and non-striker, current scores of striker and non-striker, runs and wickets in the last 3 overs etc. The page displays the range of projectedscores which the team is expected to score. To predictthe winner, we need to provide the runs remaining along with the data mentioned av. The page displays the winning percentage of the chasing team in that situation.

**CHAPTER 3: SYSTEM ANALYSIS & DESIGN**

**3.1 Use Case Diagrams**

****

**Drawbacks and Limitations**

1. Sometimes difficult in creating model occurred due to lack correct data, the result shown above are not correctly justified due to this problem.

2. lack of data size.

3.Time Consuming

4. Need proper record of data

5. Difficulty in tracking customers past history

6. Manual work and consumes large volumes of data

7. Lack of data security

**Conclusion**

This report shows how different data mining techniques can be used to predict the win outcome of the cricket game, this data is fetched from the cricket, due lack of data size the data is artificial generated to produce a certain size. The challenges were deciding the correct model to be used. Python are used to created models, also removing the noise from the was one of the toughest one to. The difficult in creating model occurred due to lack correct data, the result shown above are not correctly justified due to this problem. The Linear Regression, Multilinear Regression and The factors such team matches played and win outcome does affect rating and points that is gained by team. The study does not correctly predict win outcome but, we can presume from the study if this is conducted on a cleaned dataset the models does have the ability produce the better, where. The finding of this are mainly dependent on the three model which performed better result, which reveals that cricket win outcome is affected many factors such as team players, number matches it has played and the number matches won. If the team performance is consistent it is likely that is happy to the rating and points increasing whereas the team who does not perform well tend have low rating and points which affect the players. Correlation between team playing matches and rating and points increasing.

**Bibliography**

We are taken references from these following website to build our project and ideas only for study material purpose.

https://youtube.com

https://wikipedia.com