PROJECT - II

Report on

CRICKSTER

Submitted in partial fulfillment of the requirements

of the degree of

Bachelor of Engineering
(Electronics and Telecommunication Engineering)

by

RAO VARUN PRASANNA (19ET2005)
RAI AMOG GANGADHAR (19ET2016)
SAINI ARPIT SANJEEV(19ET2007)
LABADE SIDDESH KAMALAKAR(19ET1018)

Supervisor

Dr. AYUSH SAXENA

Department of Electronics and Telecommunication Engineering
Ramrao Adik Institute of Technology,
Sector 7, Nerul , Navi Mumbai
(Affiliated to University of Mumbai)
April 2023

Ramrao Adik Education Society's

Ramrao Adik Institute of Technology

(Affiliated to the University of Mumbai)

Dr. D. Y. Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai 400 706.

Certificate of Approval

This is to certify that, the Project - II report entitled

"CRICKSTER "

is a bonafide work done by

RAO VARUN PRASANNA(19ET2005)
RAI AMOG GANGADHAR(19ET2016)
SAINI ARPIT SANJEEV (19ET2007)
LABADE SIDDESH KAMALAKAR(19ET1018)

and is submitted in the partial fulfillment of the requirement for the degree of

Bachelor of Engineering
(Electronics and Telecommunication Engineering)
to the
University of Mumbai.

Examiner	Supervisor

Project Coordinator	Head of Department	Principal

Ramrao Adik Education Society's
Ramrao Adik Institute of Technology
(Affiliated to the University of Mumbai)
Dr. D. Y. Patil Vidyanagar,Sector 7, Nerul, Navi Mumbai 400 706.

Declaration

We wish to state that work embodied in this dissertation entitled "**CRICKSTER**" has been carried out under the guidance of "Dr. AYUSH SAXENA" at Department of Electronics and Telecommunication Engineering, Ramrao Adik Institute of Technology during 2022-2023.

We declare that the work being presented forms our own contribution and has not been submitted for any other Degree or Diploma of any University/Institute. Wherever references have been made to previous works of others, it has been clearly indicated. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

VARUN RAO (19ET2005):

AMOG RAI (19ET2016):

ARPIT SAINI (19ET2007):

SIDDESH LABADE (19ET1018):

Acknowledgments

This project is spanned over a period of one year and the number of people who have helped us are outstanding. To begin with; we would like to thank our project guide Dr. Ayush Saxena. In spite of their busy schedules, it was very kind of her to spare some of her precious time and give us valu- able suggestions which ultimately formed the heart of this report. Also I would like to thank the Head of Electronics and Tele Communications Engineering Department Dr. Chandrakant Gaikwad for his valuable guid- ance, ability to motive us and willingly help us to solve difficult equations that arosed during the making of this project. He was a great source of in- spiration and allowed us to work in the lab for extended hours. My sincere thanks to Principal Dr. Mukesh Patil for providing us the necessary facility to carry out the work. We also extend our thanks to all professors of our departments for their valuable guidance. We would also like to thank our fellow students for their support and good wishes. Finally, we thank entire Electronics and Telecommunication Engineering Department for excellent facilities and encouragement given during the course of this project.

Abstract

Cricket, especially the Twenty20 format, has maximum uncertainty, where a single over can completely change the momentum of the game. With millions of people following Cricket, developing a model for predicting the outcome of its matches is a real-world problem. A cricket match depends upon various factors, and in this work, the factors which significantly influence the outcome of a Twenty20 cricket match are identified. Each players performance in the field is considered to find out the overall weight (relative strength) of the team. A multivariate regression based solution is proposed to calculate points of each player in the league and the overall weight of a team is computed based on the past performance of the players who have appeared most for the team. In this paper, the past seven years data of cricket containing the players details, match venue details, teams, ball to ball details, is taken and analyzed to draw various conclusions which help in the improvement of a players performance. Various other features like how the venue or toss decision has influenced the winning of the match in last seven years are also predicted. Machine learning and data extraction models are considered for prediction is random Forest Regressor. Machine learning models were trained and used for predicting the outcome of each 2018 match. We performed three types of prediction toss-winner prediction, 2 inning score prediction and the actual winner prediction. The prediction results are impressive.

Contents

Abstract **iv**

List of Figures **vi**

1 PROBLEM STATEMENT **1**
1.1 Solution:- 1

2 LITERATURE SURVEY: **2**

3 ML FLOWCHART **4**

4 SOFTWARES: **5**
4.1 EXCEL: 5
4.2 VS CODE: 6

5 WORK FLOW: **7**
5.1 Methdology: 7
5.2 Working: 8
5.2.1 Random Forest: 8
5.2.2 Naive Bayes: 8
5.2.3 Support Vector Machine: 9

6 DETAILS OF HARDWARE AND SOFTWARE: **10**
6.1 Hardware Requirements: 10
6.2 Software Requirements: 10
6.2.1 Python: 10
6.2.2 SDLC: 13

7 RESULTS: **19**
7.1 ML OUTPUT and WEBSITE 19

8 SUMMARY: **26**
8.1 CONCLUSION: 26

9 REFERENCES: **28**

List of Figures

3.1 ML flowchart 4

4.1 Excel Workbook 5
4.2 VS code 6

5.1 Flow of the Project 7
5.2 Random Froest Flowchart 8

6.1 DJango flowchart 13
6.2 SDLC flowchart 14
6.3 Iterative Model 18

7.1 1st Inning Score 19
7.2 Winning Team 20
7.3 Website 21
7.4 Website 22
7.5 Before Toss 23
7.6 Winner Prediction 24
7.7 Target Prediction 25

Chapter 1

PROBLEM STATEMENT

In order to create a system that can predict the outcome of a cricket match and offer the score, the system can examine a variety of factors, including player-specific performance and statistics like batting average and bowling average. It is crucial to choose the greatest squad when setting a deadline for a certain championship so that the team has a better chance of winning.

1.1 Solution:-

To produce the best players for a team and the best outcome, this issue needed to be solved. The best players for the team that can be used in international tournaments to win the most matches are predicted using prediction algorithms like the Random Forest method. This problem has been solved by gathering historical data on a few players.

Chapter 2

LITERATURE SURVEY:

We discuss the analysis role of machine learning in the improvement of performances of players and the team in different sports and how it helps the players to know their performance levels and further improvements. We aim to make a model which initially collects raw data and these data are stored and clustered since the data is very large .

1. Kalpdrum Passi and Niravkumar Pandey discussed about the prediction accuracy in terms of runs scored by batsman and the no. of wickets taken by the bowler in each team.
2. I. P. Wickramasinghe proposed a methodology to predict the performance of batsman for the previous five years using hierarchial linear model .
3. R.P.Schumaker et. al, discussed about different statistical simulations used in predictive modeling for different sports.
4. John McCullagh implemented neural networks and datamining techniques to identify the talent and also for the selection of players based on the talent in Australian Football League.
5. Bunker et. al, proposed a novel sport prediction framework to solve specific challenges and predict sports results .
6. Ramon Diaz-Uriarte et. al, investigated the use of random forest for classification of microarray data and proposed a new method of gene selection in classification problem based on random forest .
7. Rabindra Lamsal and Ayesha Choudhary, proposed a solution to calculate the weightage of a team based on the player's past performance of IPL using linear regression .
8. Akhil Nimmagadda et. Al, proposed a model using Multiple Variable Linear Regression and Logistic regression to predict the score in different innings and also the winner of the match using Random Forest algorithm .
9. Ujwal U J et. Al, predicted the outcome of the given cricket match by analyzing previous cricket matches using Google Prediction API.
10. Rameshwari Lokhande and P.M.Chawan came up with live cricket score predicton using linear regression and Naïve Bayes classifier .
11. Abhishek Naik et. Al, proposed a new model used matrix factorization technique to analyze and predict the winner in ODI cricket match .
12. Esha Goel and Er. Abhilasha discussed the improvements in Random Forest Algorithmand described the usage in various fields like agriculture, astronomy, medicine, etc. .

13. Amit Dhurandhar and Alin Dobra proposed a new methodology for analysing the error of classifiers and model selection measures to analyse the decision tree algorithm .
14. H. Yusuff et. Al, performed logistic regression using mammograms to find the accuracy with valid samples

Chapter 3

ML FLOWCHART

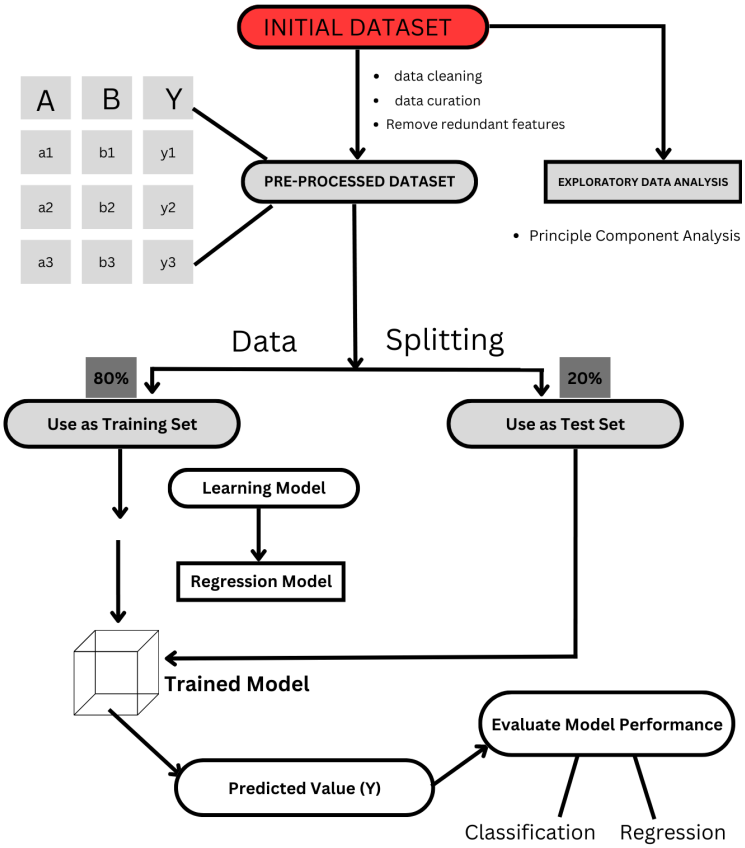

Figure 3.1: ML flowchart

The above Flowchart shows a basic working of an ML model which describes how a Model works and what all parameters it uses. It shows us that how a dataset is initially pre processed and then converted into a training and testing model after which it is further processed to make the required model.

Chapter 4

SOFTWARES:

4.1 EXCEL:

Figure 4.1: Excel Workbook

Microsoft produced Microsoft Excel, a spreadsheet, for Windows, macOS, Android, and iOS. It has calculating or computing capabilities, graphing tools, pivot tables, and the Visual Basic for Applications (VBA) macro programming language. The Microsoft Office programme suite includes Excel. In order to organise data manipulations like arithmetic operations, spreadsheets like Microsoft Excel use a grid of cells organised in numbered rows and letter-named columns. It has a variety of built-in functionalities to address financial, engineering, and statistical requirements. Additionally, it has a very limited three-dimensional graphical display and can present data as line graphs, histograms, and charts. Using pivot tables and the scenario manager, it enables segmenting of data to view its reliance on multiple parameters for diverse viewpoints. A data analysis tool is a pivot table. This is accomplished by using PivotTable fields to condense big data sets. It features a programming component called Visual Basic for Applications that enables users to apply a wide range of numerical techniques, such as those for solving differential equations in mathematical physics, and then send the results back to the spreadsheet. The spreadsheet presents itself as a so-called application, or decision support system (DSS), via a specially designed user interface, such as a stock analyzer, or more generally, as a design tool that asks the user questions and provides answers and reports. It also has a variety of interactive features that

enable user interfaces that can completely hide the spreadsheet from the user.In a more complex implementation, an Excel programme can automatically query outside databases and measuring devices in accordance with a timetable for updates, analyse the data, create a Word report or PowerPoint slide show, and send out these presentations on a regular basis to a list of participants. Excel wasn't intended to be a database, though.

4.2 VS CODE:

Figure 4.2: VS code

Microsoft created the source-code editor Visual Studio Code, generally known as VS Code, for Windows, Linux, and macOS using the Electron Framework. Debugging support, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git are among the features. Users can modify the theme, keyboard shortcuts, preferences, and add functionality by installing extensions.C, C++, Fortran, Go, Java, JavaScript, Node.js, Python, and Rust are just a few of the programming languages that may be utilised with Visual Studio Code, a source-code editor. The Electron framework, which is used to create Node.js web apps that leverage the Blink layout engine, serves as its foundation. The editor component (codenamed "Monaco") used in Azure DevOps, formerly known as Visual Studio Online and Visual Studio Team Services, is also utilised in Visual Studio Code. The majority of popular programming languages have minimal support out of the box in Visual Studio Code. This fundamental support consists of configurable snippets, code folding, bracket matching, and syntax highlighting. Along with debugging support for Node.js, Visual Studio Code also comes with IntelliSense for JavaScript, TypeScript, JSON, CSS, and HTML. The VS Code Marketplace's freely downloadable extensions can add support for more languages.

Chapter 5

WORK FLOW:

5.1 Methdology:

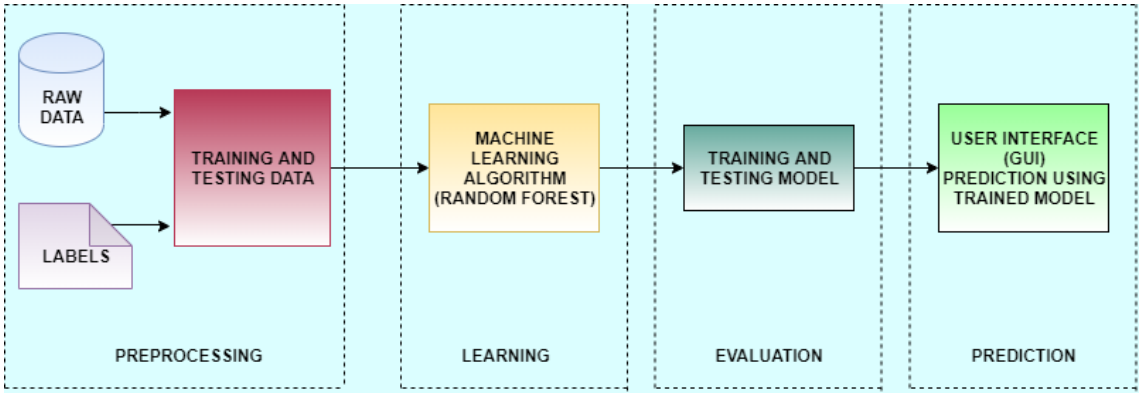

Figure 5.1: Flow of the Project

The three key parts of the methodology are data preprocessing, data preparation, and data encoding. The seven IPL seasons' real-time dataset is initially collected in CSV format. The data are unreliable, noisy, and inconsistent during the preparation stage. To get the best outcomes, the data preparation is important. To assess several predictor variables, an outcome measure must be selected. Label each term with a brief name and encode it as a numerical value for predictive modelling during the data encoding phase.

5.2 Working:

5.2.1 Random Forest:

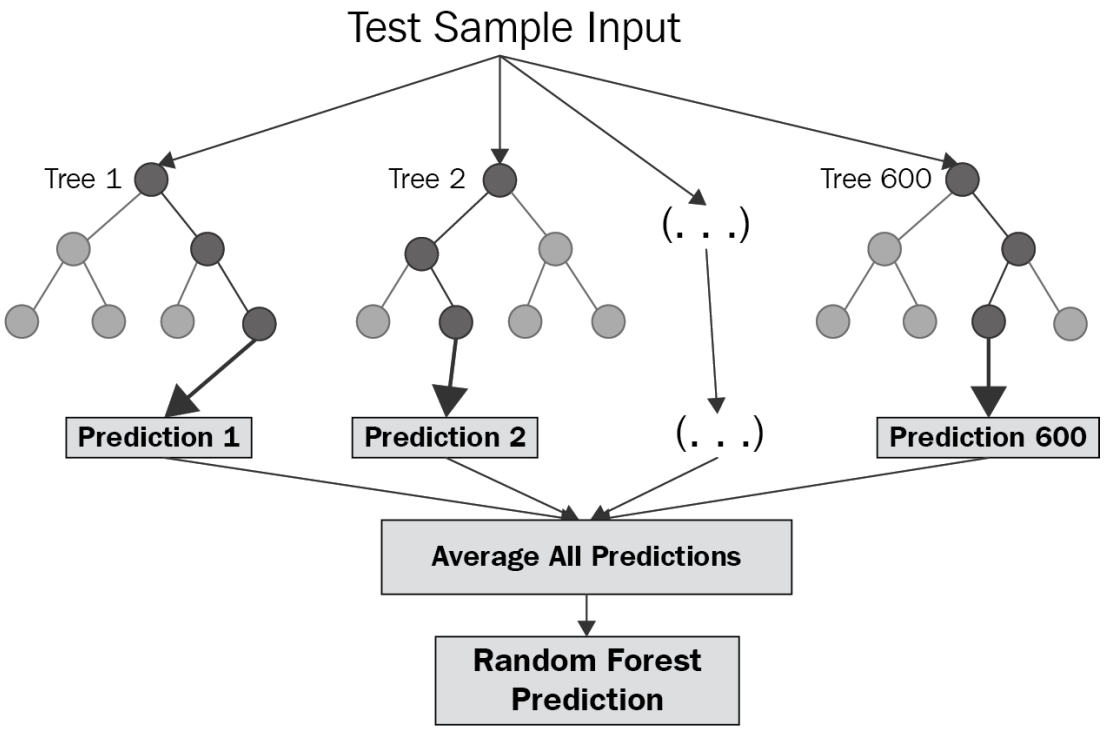

Figure 5.2: Random Froest Flowchart

A supervised learning algorithm called random forest uses the ensemble learning approach for classification and regression. In contrast to boosting techniques, random forest is a bagging approach. Random forests operate their trees in parallel. While the trees are being built, there is no contact between them. It works by building a large number of decision trees during training period and producing the class that is the mean of the predictions (regression) or the mode of the classes (classification) of the individual trees. A random forest is a meta-estimator that aggregates many decision trees with certain useful alterations, i.e., it mixes the outcome of multiple forecasts. 1. A certain percentage of the entire number of features (referred to as the hyperparameter) can be split on at each node. This makes that the ensemble model uses all of the potentially predictive features fairly and does not rely unduly on any one feature. 2. To add another element of randomization and avoid overfitting, each tree selects a random sample from the original data set while creating its divides.

5.2.2 Naive Bayes:

A group of classification algorithms built on the Bayes' Theorem are known as naive Bayes classifiers. It is a family of algorithms rather than a single method, and they are all based on the idea that every pair of features

being classified is independent of the other. Naive Bayes Classifier Types: Multinomial Simple Bayes: This is mostly used for document classification issues, such as determining whether a document falls under the sports, politics, technology, etc. category. The frequency of the terms included in the document is one of the features/predictors that the classifier uses. Bernoulli Naive Bayes: Similar to the multinomial naive bayes, the Bernoulli naive bayes uses boolean variables as predictors. The only options for the factors we use to predict the class variable are yes or no, as in whether a word is in the text or not. Gaussian Naive Bayes: We assume that the values of the predictors are samples from a gaussian distribution when they take up a continuous value and are not discrete.

5.2.3 Support Vector Machine:

Finding a hyperplane in an N-dimensional space (N is the number of features) that categorises the data points clearly is the goal of the support vector machine algorithm. There are a variety of different hyperplanes that might be used to split the two classes of data points. Finding a plane with the greatest margin—that is, the greatest separation between data points from both classes—is our goal. Maximising the margin distance adds some support, increasing the confidence with which future data points can be categorised. Support vectors are data points that are closer to the hyperplane and have an impact on the hyperplane's position and orientation. By utilising these support vectors, we increase the classifier's margin. The hyperplane's location will vary if the support vectors are deleted. These are the ideas that aid in the development of our SVM.

Chapter 6

DETAILS OF HARDWARE AND SOFTWARE:

6.1 Hardware Requirements:

- 4GB RAM.
- 10 GB HDD.
- Intel 1.66 GHz Processor Pentium 4

6.2 Software Requirements:

- Windows 7
- Python 3.6.3
- VS code

6.2.1 Python:

High-level programming skills are available in Python, a general-purpose object-oriented programming language. Its obvious and simple syntax, portability, and ease of learning have helped it become well-known. Python is a computer language with elements of both Java and C. For object-oriented programming, it offers classes and objects similar to Java, and it offers the approach of producing elegant code similar to C.Python is an interpreted, object-oriented, high-level, dynamically semantic programming language. It is particularly desirable for Rapid Application Development as well as for usage as a scripting or glue language to tie existing components together due to its high-level built-in data structures, dynamic typing, and dynamic binding. Python's straightforward syntax emphasises readability and makes it simple to learn, which lowers the cost of programme maintenance. Python's support for modules and packages promotes the modularity and reuse of code in programmes. For all popular platforms, the Python interpreter and the comprehensive standard library are freely distributable and available in source or binary form. At the

National Research Institute for Mathematics and Computer Science in the Netherlands, Guido van Rossum created Python in the late 1980s and early 1990s. ABC, Modula-3, C, C++, Algol-68, SmallTalk, the Unix shell, and other scripting languages are just a few that Python is derived from. Copyright applies to Python. Python source code is currently accessible under the GNU General Public Licence (GPL), just like Perl. Although a core development team at the institution is now responsible for maintaining Python, Guido van Rossum continues to play a key role in guiding its development.

Libraries:

PANDAS

Pandas is a data analysis and manipulation software package created for the Python programming language. It includes specific data structures and procedures for working with time series and mathematical tables. It is free software distributed under the BSD license's three clauses. The word is derived from "panel data," a phrase used in econometrics to refer to data sets that contain observations for the same persons throughout a range of time periods. Its name is a pun on "Python data analysis" in and of itself. From 2007 to 2010, Wes McKinney worked as a researcher at AQR Capital, where he began creating the pandas that are known today.

NUMPY

It is a library for the Python programming language that adds support for sizable, multi-dimensional arrays and matrices as well as a substantial number of sophisticated mathematical operations that may be performed on these arrays. Jim Hugunin originally developed Numeric, the predecessor to NumPy, with assistance from a number of other programmers. Travis Oliphant developed NumPy in 2005 by heavily altering Numeric to incorporate capabilities of the rival Numarray. Numerous people have contributed to the open-source programme NumPy. A project sponsored financially by NumFOCUS is NumPy. NumPy targets the non-optimizing CPython bytecode interpreter, which is the Python reference implementation. Due to the lack of compiler optimisation, mathematical algorithms created for this version of Python frequently execute considerably slower than their compiled counterparts. Multidimensional arrays, efficient array-based functions, and operators are some of the ways that NumPy addresses the slowness issue. Using them necessitates rewriting some code, primarily inner loops, in NumPy. Since both Python and MATLAB are interpreted languages, NumPy can be used to create programmes quickly as long as

the majority of operations are performed on arrays or matrices rather than scalars. In contrast, MATLAB has a lot of extra toolboxes, including Simulink, but NumPy is inextricably linked to Python, a more advanced and comprehensive programming language. There are also complementing Python programmes available, like Matplotlib, a plotting package, and SciPy, a library that adds more MATLAB-like functionality. Internally, BLAS and LAPACK are used by both MATLAB and NumPy to perform effective linear algebra computations.

SCI-KIT LEARN

A free freeware machine learning library for the Python programming language is scikit-learn, sometimes referred to as sklearn. Support-vector machines, random forests, gradient boosting, k-means, and DBSCAN are just a few of the classification, regression, and clustering algorithms it offers. It is also built to work with Python's NumPy and SciPy scientific and numerical libraries. A NumFOCUS fiscally sponsored project is scikit-learn. Scikit-learn is mostly designed in Python and heavily utilises NumPy for high-performance array and linear algebra operations. To enhance performance, some fundamental algorithms are also written in Cython. Support vector machines, logistic regression, and linear support vector machines are implemented using wrappers built in Cython for LIBSVM and LIBLINEAR, respectively. It might not be viable to expand these methods with Python in such circumstances Scikit-Learn works nicely with many other Python libraries, including SciPy, Pandas dataframes, NumPy for array vectorization, Matplotlib and plotly for graphing, and many more. David Cournapeau, a French data scientist, founded the scikit-learn project as scikits.learn during the Google Summer of Code. The idea that the project is a "SciKit" (SciPy Toolkit), an independently created and distributed third-party extension to SciPy, is the source of the project's name.

Framework:

DJANGO:

Django is a model-template-views (MTV)-based web framework that is free and open-source that uses Python.The 501(c)(3) non-profit Django Software Foundation (DSF), a separate entity founded in the US, is responsible for maintaining it. Django's main objective is to make it simpler

to create intricate, database-driven websites. The framework places an emphasis on the principles of don't repeat yourself, minimum coupling, rapid development, and the "pluggability" of components. Everywhere, even in settings, files, and data models, Python is employed. Additionally, Django offers a customizable administrative creation, read, update, and delete interface that is dynamically built by introspection. Django is used by some well-known websites, including Nextdoor, Clubhouse, Mozilla, Instagram, Disqus, and Disqus.

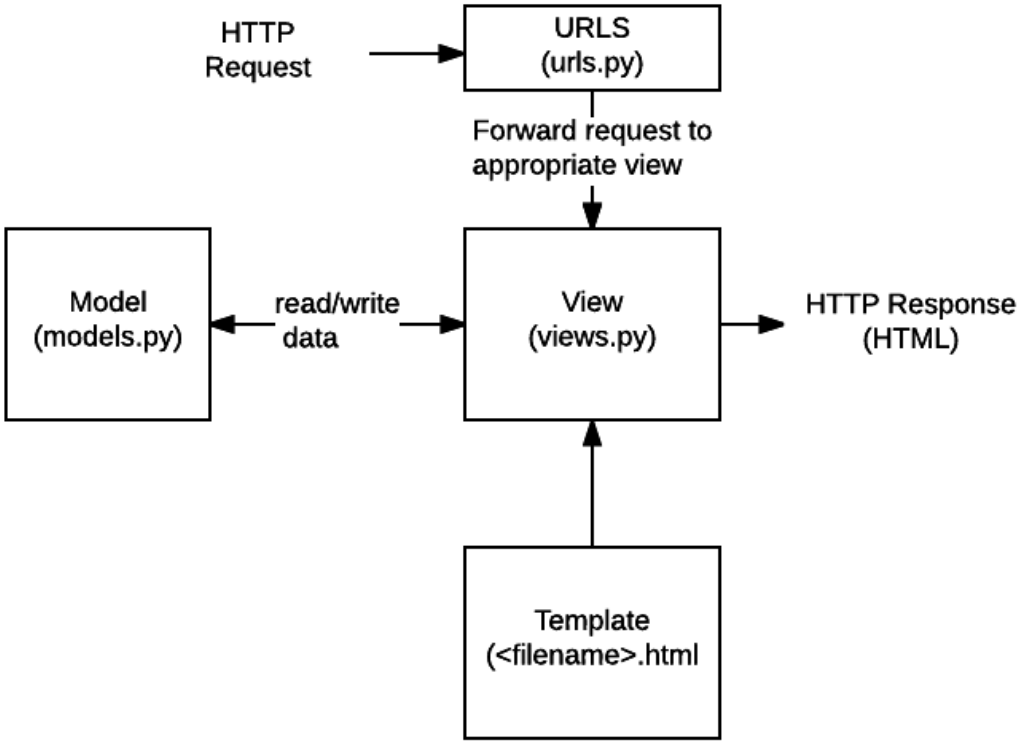

Figure 6.1: DJango flowchart

6.2.2 SDLC:

The process of creating information systems through research, analysis, design, implementation, and maintenance is known as the system develop-

ment life cycle. Information systems development and application development are other names for the System Development Life Cycle (SDLC).

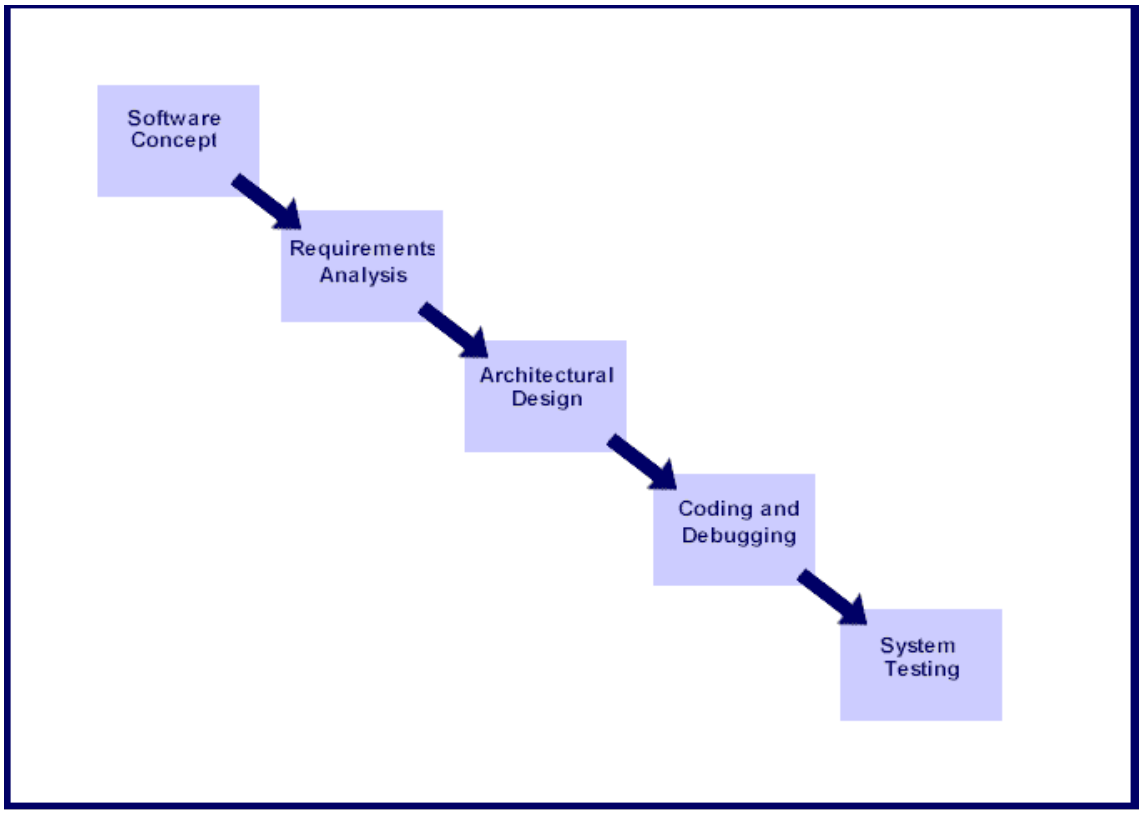

Figure 6.2: SDLC flowchart

Steps involved in SDLC:

Below are the steps involved in the System Development Life Cycle. Each phase within the overall cycle may be made up of several steps.
Step 1: Software Concept Finding a need for the new system is the first step. This will involve identifying potential business opportunities or problems, completing a feasibility assessment to see if the suggested solution is practical, and creating a project plan. End users that have a suggestion for how to make their work better may be involved in this process. To make sure that IT is being used to support the organisation in achieving its strategic goals, the process should ideally take place concurrently with a review of the organization's strategic plan. Before any funds are budgeted for its development, management may need to accept concept ideas.

Step 2:Requirements Analysis Requirements analysis is the process of examining end users' information needs, the organisational environment, and

any existing systems in use in order to produce the functional specifications of a system that can satisfy users' needs. Additionally, the requirements ought to be written down in a letter, email, storyboard for the user interface, executable prototype, or any other format. To make sure the developing project is in line with user demands and requirements, the requirements documentation should be consulted throughout the remaining stages of the system development process.To guarantee that the new system will perform properly and meet end users' demands and expectations, professionals must involve end users in this process.

Step 3:Architectural Design The specifications for the hardware, software, human, and data resources, as well as the information products that will satisfy the functional requirements of the proposed system, can be established once the needs have been established. The design will act as the system's blueprint and aid in identifying issues before they are included into the finished product. Professionals design the system, but users must examine their work to make sure it satisfies their demands.

Step 4:Coding and Debugging The process of developing the finished system involves coding and debugging. The software developer completes this phase.

Step 5:System Testing To compare the system's actual functionality to its intended or expected functionality, it must be tested. Other challenges to think about at this time include transferring old data into the new system and training staff members to use it. End users will be crucial in assessing if the produced system satisfies the required specifications and how much it is actually used.

Step 6:Maintenance The system will inevitably require maintenance. When software is provided to the customer, changes will undoubtedly be made. The change has a number of causes. Unexpected input values entering the system could lead to change. Additionally, modifications to the system may have an immediate impact on how the software functions. The programme must to be created to support adjustments that can be made after implementation. There are many models of software processes, including:-

- Prototyping Model
- RAD Model
- The Spiral Model

• The Spiral Model
• The Waterfall Model
• The Iterative Model
For the development of our project, we chose the iterative model (the linear sequential model) out of all of these process models.

Iterative Model:

The cascading influence from one phase to the next, as shown in Figure 1.1, is how the waterfall model gets its name. Each phase in this paradigm has a clear beginning and end, as well as recognisable delivery to the following phase. This paradigm is also known as the software life cycle or the linear sequential model. The model consists of six distinct stages, namely:
1. In the requirements analysis phase
(a) The targeted service objectives (goals) are provided along with the problem.
(b) The constraints are identified
2. The detailed definitions of (a) and (b) above are used to generate the system specification during the specification process. The product function should be explicitly stated in this text.
3. The system specifications are converted into a software representation throughout the system and software design phase. At this point, the software engineer is focused on:
• Data structure
• Software architecture
• Algorithmic detail
• Interface representations
At this point, the hardware requirements are also established, along with an overview of the system architecture. The software engineer ought to be able to recognise the connections between the hardware, software, and related interfaces by the end of this phase. It is ideal to avoid passing any specification errors "downstream."
4. The designs are transformed into the software domain during the implementation and testing phase.
• The amount of coding required can be considerably decreased by thorough documenting from the design phase.
• At this step, testing focuses on ensuring that any problems are found and that the programme complies with the specified requirements.
5. All programme modules are integrated and tested as part of the system

testing step to make sure the entire system complies with the software requirements. The software is then given to the client for acceptance testing [Deliverable - The software product is given to the client for acceptance testing].
6. The software's maintenance phase is typically the longest. The software is updated during this phase to:
• Meet the changing customer needs
• Adapted to accommodate changes in the external environment
• Correct errors and oversights previously undetected in the testing phases
• Enhancing the efficiency of the software
Observe how feedback loops enable the model to incorporate corrections. For instance, an issue or change in the design phase necessitates a'revisit' of the specs phase. The pertinent documentation should be updated to reflect any changes made at any phase.

Advantages of the Iterative Model:
• Every stage of the iterative model includes testing by default.
• It is an enforced disciplined approach
• It is driven by documentation, therefore documentation is created at each level

Disadvantages of the Iterative Model:
The oldest and most popular paradigm is the waterfall model. However, many projects hardly ever adhere to its logical order. This is as a result of its strict format's inherent issues. Namely:
• Changes may be quite confusing as the project develops because iteration is only included indirectly.
• This IM has trouble addressing the initial project's inherent unpredictability because the client typically just has a general understanding of what is expected from the software solution.
• After the product has been coded, only a working version is presented to the customer. If any problems go unnoticed and progress to this point, it might be disastrous.

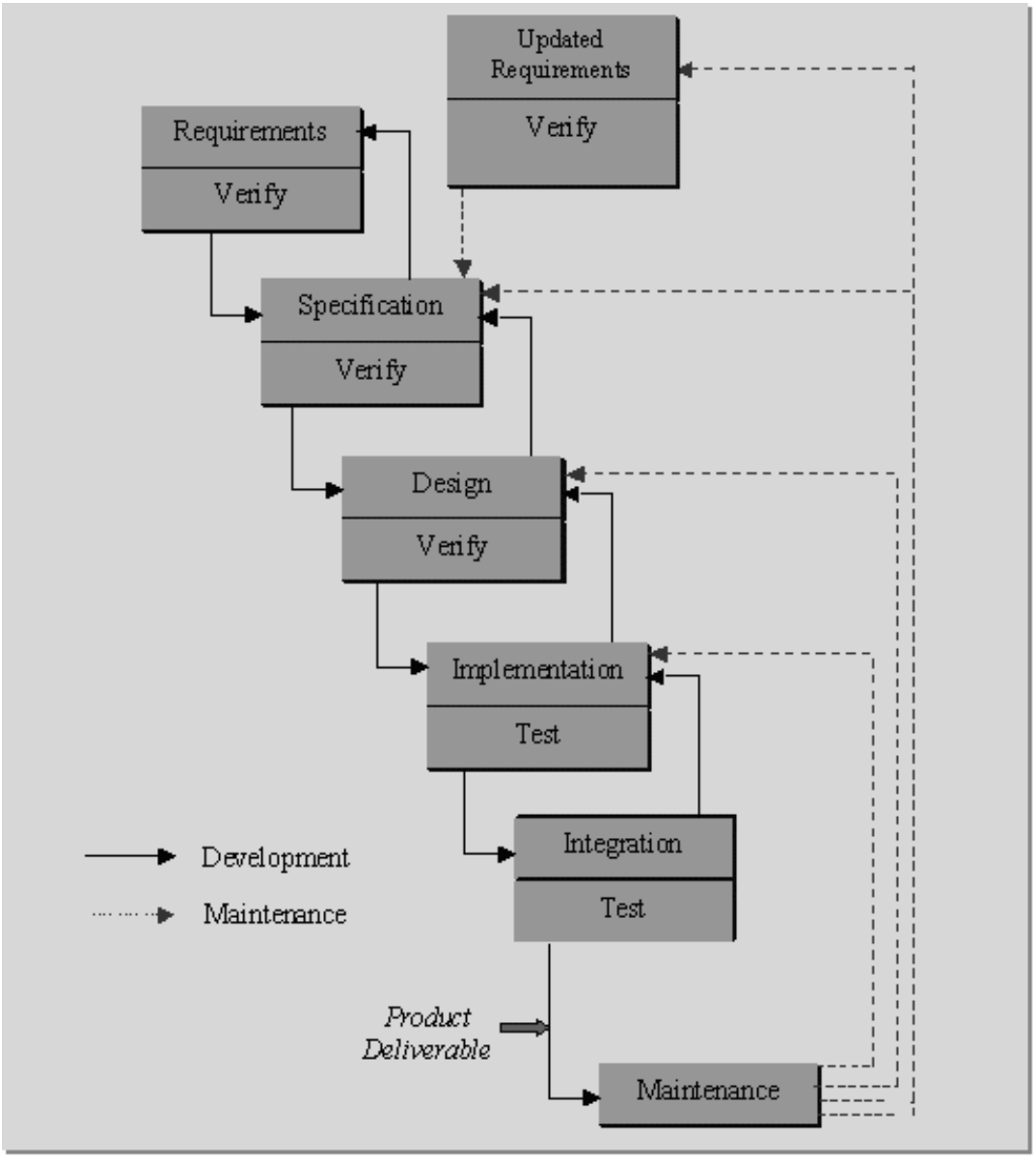

Figure 6.3: Iterative Model

Chapter 7

RESULTS:

7.1 ML OUTPUT and WEBSITE

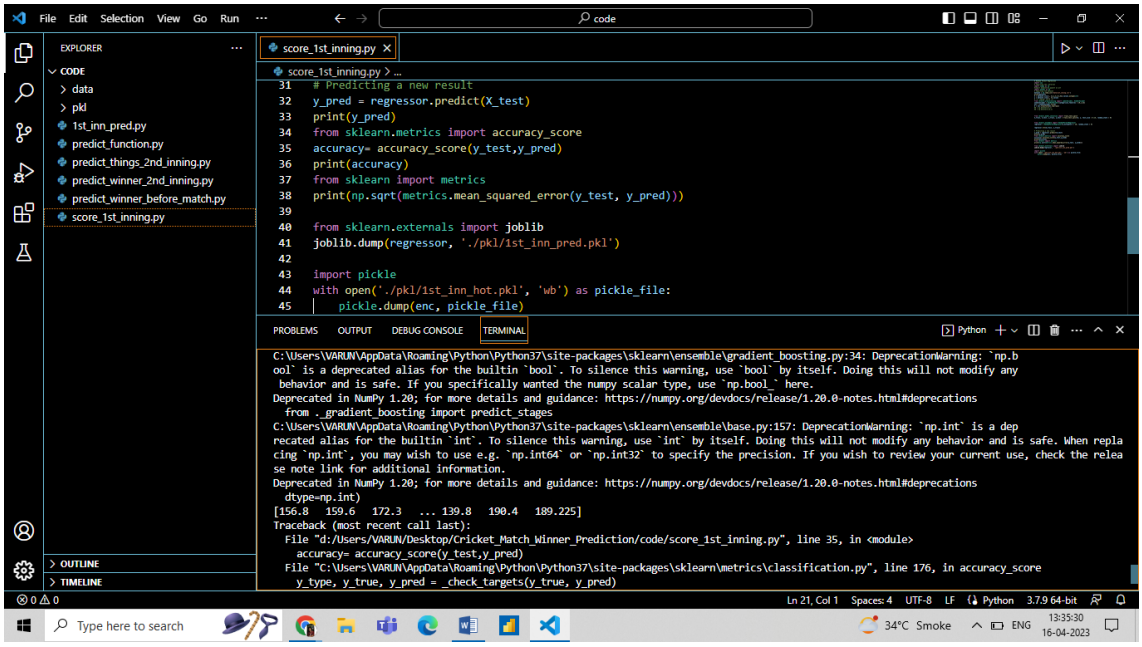

Figure 7.1: 1st Inning Score

In above Image we have displayed the Machine Learning (Random Forest Regressor Model) code in which we have displayed the output as well as the result which shows the score prediction.

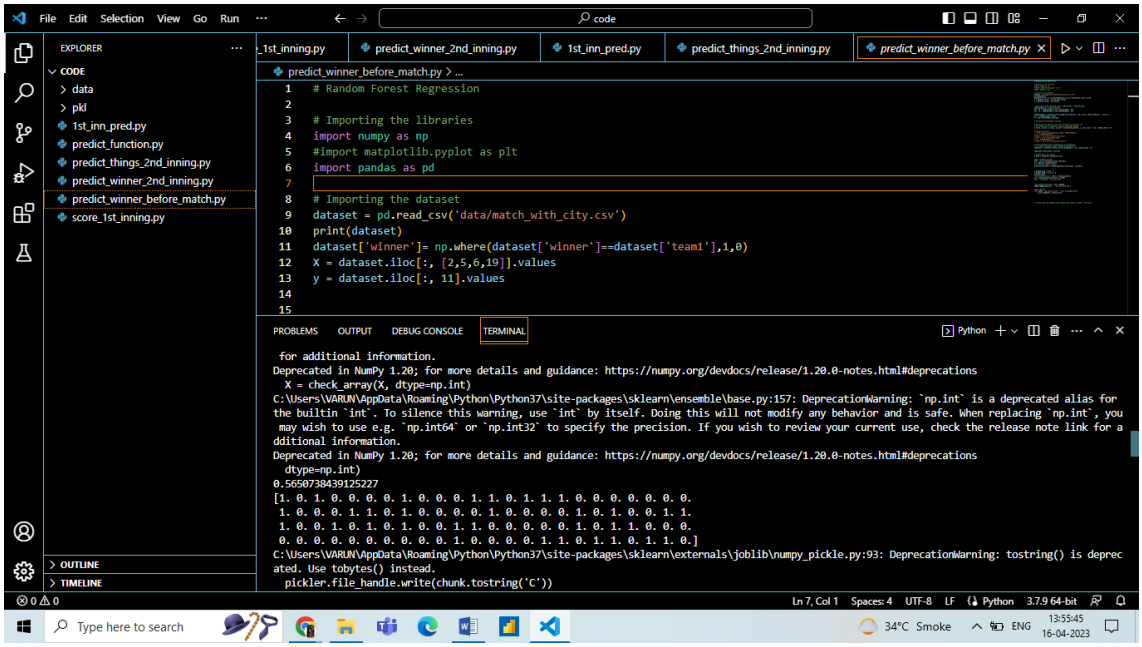

Figure 7.2: Winning Team

In above Image we have displayed the Machine Learning (Random Forest Regressor Model) code in which we have displayed the Winning chances of a team given by 1's and 0's using basic function like One Hot Encoding.

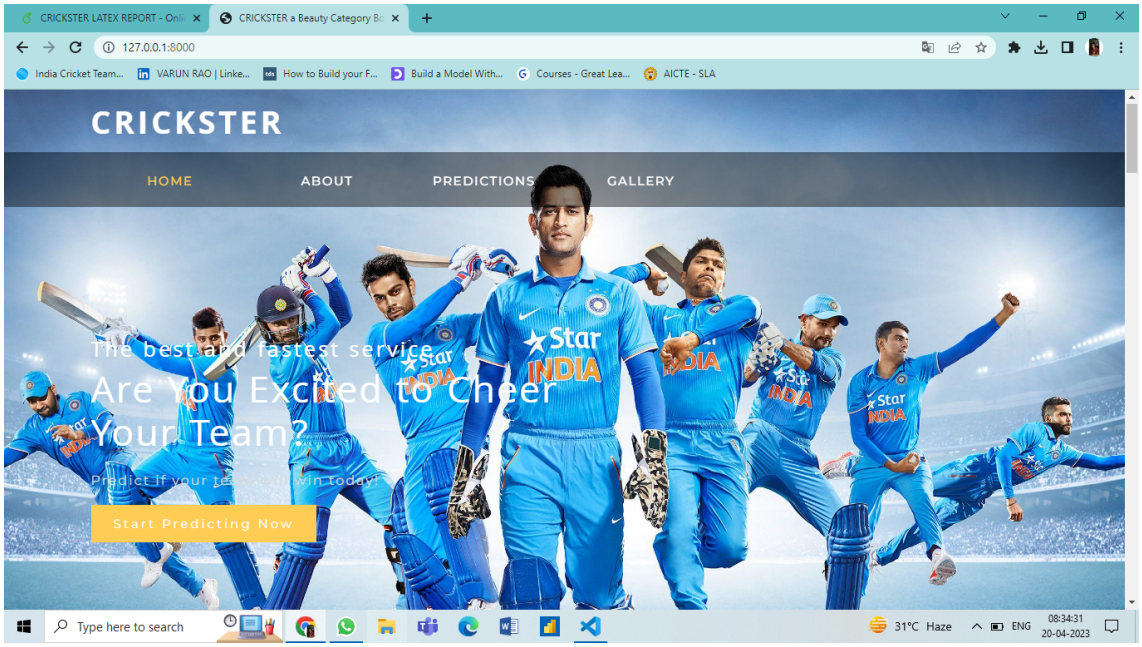

Figure 7.3: Website

In above Image we have displayed the Basic User Interface Website of our Project. The website displays basic Features like HOME, ABOUT, PREDICTIONS, CONTACT etc. Apart from this our portal also redirects to the official IPL website for latest updates.

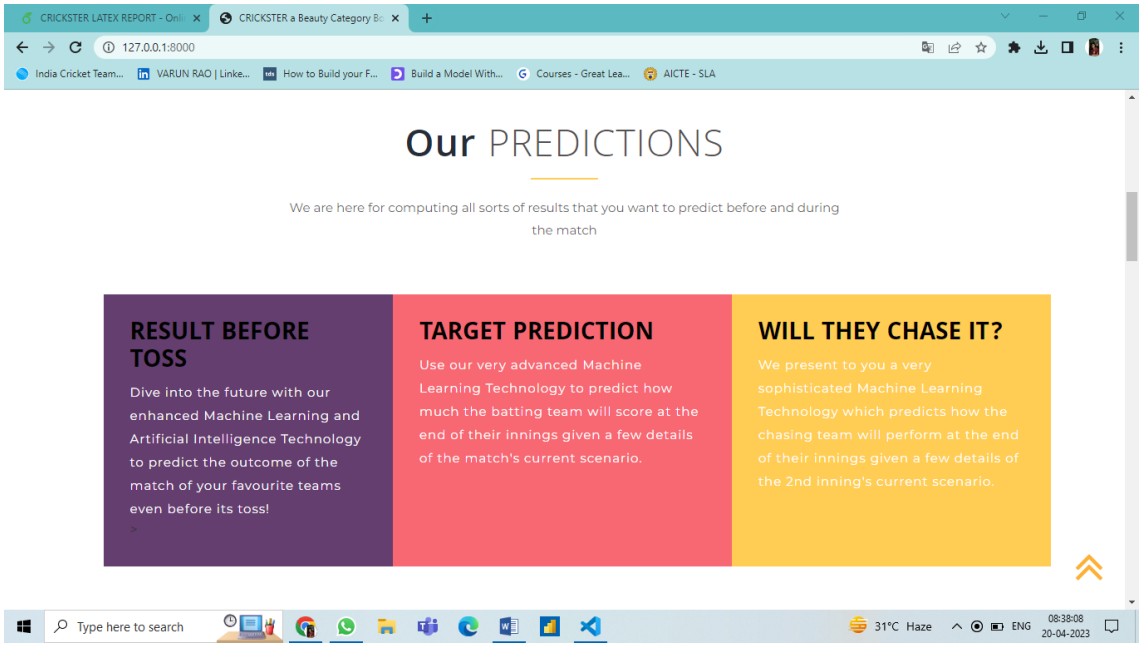

Figure 7.4: Website

In the given Image we have displayed the website page in which all our Prediction parameters are shown, namely result before the Toss, Target predictions, will they chase it each having it's own specific working and function.

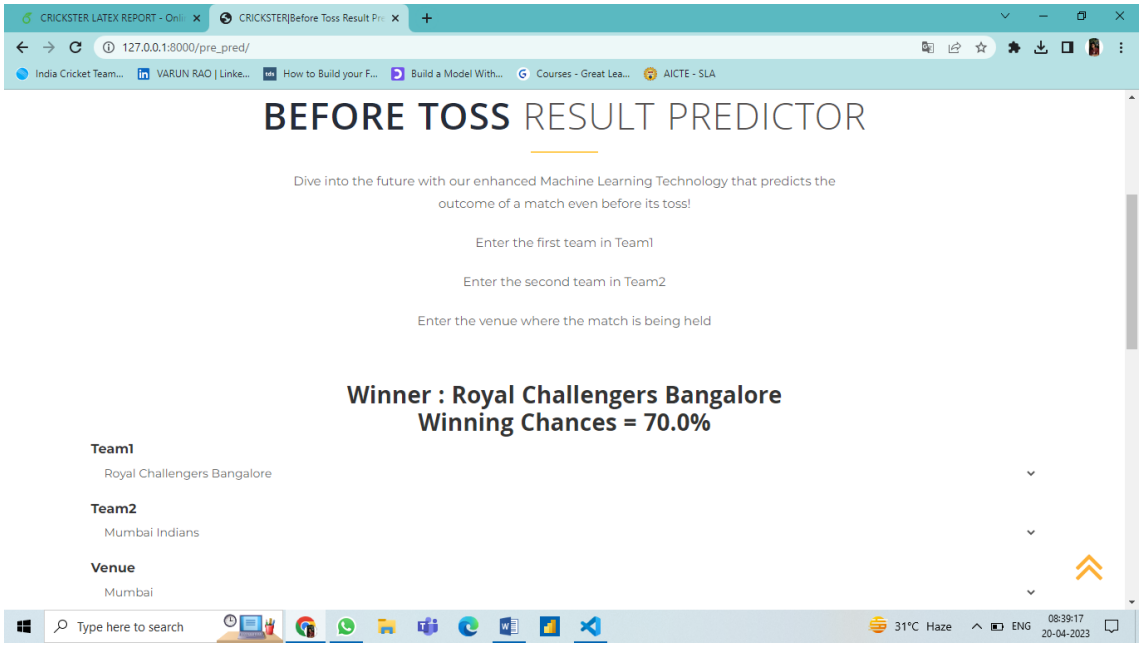

Figure 7.5: Before Toss

The following above image displays the result of Two teams and it shows the prediction of which team between the 2 will have higher chances of winning that match.

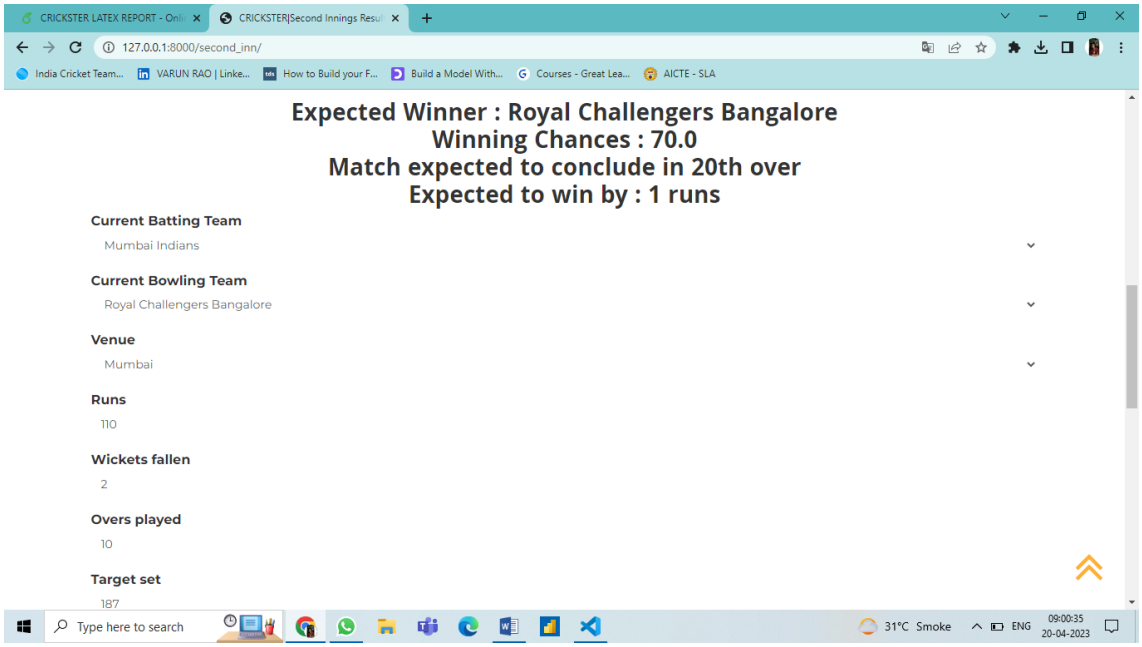

Figure 7.6: Winner Prediction

The above image shows us the result of a match after the target has been set by Team A and what are the chances that either of the team will win and by how many runs.

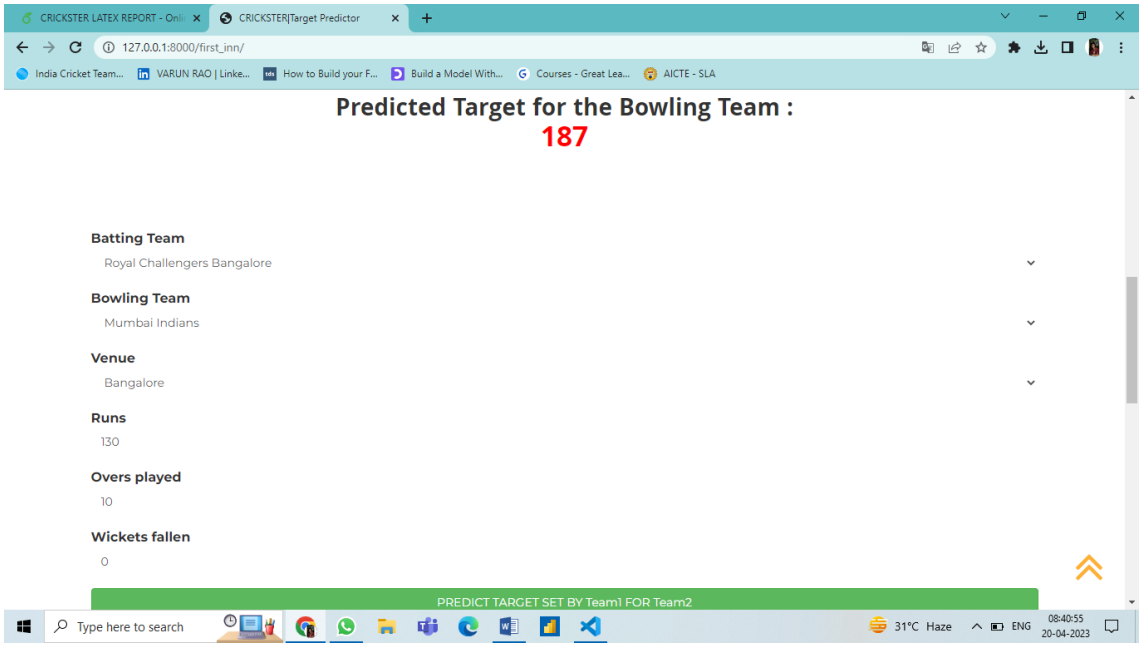

Figure 7.7: Target Prediction

In this Image we can see that the results of Team A's target after we put some minimum data. It shows us how muc a Team will score after the end of 20th over.

Chapter 8

SUMMARY:

8.1 CONCLUSION:

The best team being chosen for a cricket match is crucial to the team's success. This essay's major objective is to forecast the performance of the players by analysing cricket data. To identify the most precise algorithm, three classification algorithms are utilised and put side by side. With 89.15 percent accuracy in predicting the greatest player performance, Random Forest is shown to be the best classifier.In the future, the best team will be predicted using this information.

Let us now look into all the three outcomes of our projects.

1. Before Toss result showed us that how we can predict the results of a Team even before the matches has started. This prediction is done ne the basis of past 10 years data of a particular team and how different aspects have contributed to winning or losing if that team.

2. Target prediction showed us that how we can predict the score of a team based on simple data like giving runs and overs also giving wickets. Using such parameters we can easily guess how much a team can score at the end of 20th over.

3. The third and the last prediction parameters is Will the 2nd team be able to Chase the score given by Team 1. In this prediction we again assign some basic data such as runs, overs, wickets fallen and the Target given by Team 1.

These were some of the Prediction Parameters we tried to incorporate

in our project, but this is not the end. we aim to add more features to our project, we would like to expand this project by considering other cricket formats such as TESTS and ODI's. If given the right data to the model one can easily predict each and everything possible.

Chapter 9

REFERENCES:

[1] Passi, Kalpdrum and Pandey, Niravkumar. (2018) "Predicting Players' Performance in One Day International Cricket Matches Using Machine Learning" 111-126. 10.5121/csit.2018.80310.

[2] I. P. Wickramasinghe et. al, "Predicting the performance of batsmen in test cricket," Journal of Human Sport and Exercise", vol. 9, no. 4, pp. 744-751, May 2014.

[3] R. P. Schumaker, O. K. Solieman and H. Chen, "Predictive Modeling for Sports and Gaming" in Sports Data Mining, vol. 26, Boston, Massachusetts: Springer, 2010.

[4] J. McCullagh, "Data Mining in Sport: A Neural Network Approach," International Journal of Sports Science and Engineering, vol. 4, no. 3, pp. 131-138, 2012.

[5] Bunker, Rory and Thabtah, Fadi. (2017) "A Machine Learning Framework for Sport Result Prediction. Applied Computing and Informatics", 15. 10.1016/j.aci.2017.09.005.

[6] Ramon Diaz-Uriarte and Sara, "Gene selection and classification of microarray data using random forest, BMC Bioinformatics", doi:10.1186/1471-2105-7-3

[7] Rabindra Lamsal and AyeshaChoudhary, "Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning"

[8] Akhil Nimmagadda et. Al, "Cricket score and winning prediction using

data mining", IJARnD Vol.3, Issue3.

[9] Ujwal U J et. At, "Predictive Analysis of Sports Data using Google Prediction API" International Journal of Applied Engineering Research", ISSN 0973-4562 Volume 13, Number 5 (2018) pp. 2814-2816.

[10] Rameshwari Lokhande and P.M.Chawan, "Live Cricket Score and Winning Prediction", International Journal of Trend in Research and Development, Volume 5(1), ISSN: 2394-9333.