A Major Project Report

On

**Cricket Management System**

*Submitted to JNTU HYDERABAD*

*In Partial Fulfillment of the requirements for the Award of Degree of*

**BACHELOR OF TECHNOLOGY**   **IN COMPUTER SCIENCE AND ENGINEERING**

Submitted

By

**G. Vineeth Kumar (208R1A0521) K. Sridhar (208r1A0526)**

**N. Vikram Reddy (208R1A0541) P. Revanth (208R1A0545)**

**R. Venu (208R1A0548) R. Eeshwar (208R1A0549)**

Under the Esteemed guidance of

**Mrs. M. Ashwitha Reddy**

Assistant Professor, Department of CSE



**Department of Computer Science & Engineering**

**CMR ENGINEERING COLLEGE**

(Approved by AICTE, NEW DELHI, (Autonomous), Hyderabad)

Kandlakoya, Medchal Road, R.R. Dist. Hyderabad-501 401)

**2021-2022**

**CMR ENGINEERING COLLEGE**

*(Accredited by NBA,Approved by AICTE NEW DELHI, (Autonomous), Hyderabad)*

*Kandlakoya, Medchal Road, Hyderabad-501 401*

**Department of Computer Science & Engineering**



**CERTIFICATE**

This is to certify that the project entitled **“LOAN PREDICTION DATASET USING**

**MACHINE LEARNING WITH DATA ANALYSIS”** is a bonafide work carried out by

**G. Vineeth Kumar (208R1A0521) K. Sridhar (208r1A0526)**

**N. Vikram Reddy (208R1A0541) P. Revanth (208R1A0545)**

**R. Venu (208R1A0548) R. Eeshwar (208R1A0549)**

in partial fulfilment of the requirement for the award of the degree of **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE AND ENGINEERING** from CMR Engineering

College, (Autonomous), Hyderabad, under our guidance and supervision.

The results presented in this project have been verified and are found to be satisfactory. The results embodied in this project have not been submitted to any other university for the award of any other degree or diploma.

Internal Guide Major Project Coordinator Head of the Department

**Mrs. M.Ashwitha Reddy Mrs. G. Sumalatha Dr. Sheo Kumar**

Assistant Professor Associate Professor Professor & H.O.D

Department of CSE Department of CSE Department of CSE,

CMREC ,Hyderabad CMREC ,Hyderabad CMREC, Hyderabad

**DECLARATION**

This is to certify that the work reported in the present project entitled " **LOAN PREDICTION**

**DATASET USING MACHINE LEARNING WITH DATA ANALYSIS "** is a record of bonafide work done by us in the Department of Computer Science and Engineering, CMR Engineering College, JNTU Hyderabad. The reports are based on the project work done entirely by us and not copied from any other source .We submit our project for further development by any interested students who share similar interests to improve the project in the future.

The results embodied in this project report have not been submitted to any other University or Institute for the award of any degree or diploma to the best of our knowledge and belief.

**G. Vineeth Kumar (208R1A0521) K. Sridhar (208r1A0526)**

**N. Vikram Reddy (208R1A0541) P. Revanth (208R1A0545)**

**R. Venu (208R1A0548) R. Eeshwar (208R1A0549)**

**ACKNOWLEDGMENT**

We are extremely grateful to **Dr. A. Srinivasula Reddy**, Principal and **Dr. Sheo Kumar**, HOD, **Department of CSE, CMR Engineering College** fortheir constant support**.**

We are extremely thankful to **Mrs. M. Ashwitha Reddy,** Assistant Professor, Internal Guide, Department of CSE, for her constant guidance, encouragement and moral support throughout the project.

We will be failing in duty if we do not acknowledge with grateful thanks to the authors of the references and other literatures referred in this Project.

We express my thanks to all staff members and friends for all the help and co-ordination extended in bringing out this Project successfully in time.

Finally, we are very much thankful to our parents who guided me for every step.

**G. Vineeth Kumar (208R1A0521) K. Sridhar (208r1A0526)**

**N. Vikram Reddy (208R1A0541) P. Revanth (208R1A0545)**

**R. Venu (208R1A0548) R. Eeshwar (208R1A0549)**

**CONTENTS**

##### TOPIC PAGE NO

# ABSTRACT i

# LIST OF FIGURES ii

# 1.INTRODUCTION

1.1 Introduction 1

1.2 Purpose of System 3

1.3 Existing System 3

1.4 Proposed System 4

2.LITERATURE SURVEY 6

3.SOFTWARE REQUIREMENT ANALYSIS

3.1 Software Development life Cycle 7

3.2 Study of System 9

3.3 Fundamental concepts on (Domain) 10

3.4 System Analysis 13

3.5 Functional Requirements 14

3.6 Feasible Study 14

4.SOFTWARE AND HARDWARE REQUIREMENTS

4.1 Software Requirements 21

4.2 Hardware Requirements 21

##### TOPIC PAGE NO

5.SOFTWARE DESIGN

5.1 UML Diagrams Overview 22

5.2 Use Case Diagram 23

5.3 Class Diagram 24

5.4 Activity Diagram 25

5.5 Sequence Diagram 26

5.6 Component Diagram 27

5.7 Deployment Diagram 28

1. CODING

6.1 Coding 29

7.TESTING 35

8.OUTPUT SCREENS 38

9.CONCLUSION 43

10.REFERENCE 44

**ABSTRACT**

Online cricket management system is used for providing a web interface with the complete information of the National and International statistics of cricket. The information is available country wise and player wise.

By entering the data of each match, we can get all type of reports instantly, which will be useful to call back history of each player. Also the team performance in each match can be obtained. We can get a report on number of matches, wins and lost.

This web interface is useful for cricket lovers. Since they can instant reports generated within few seconds of time. And steps for accessing those reports is also very simple all the credentials they need to provide is a phone number or an effective username for login this will generate user to a home page that is defined with different set of options.

The information regarding the players and their teams is stored on a database and that information includes players personal details along with the achievements in their cricket career.

The admin here have access to edit player records, team records and other major information related to cricket management System for effective application development approach.

i

##### LIST OF FIGURES

###### S.NO DESCRIPTION PAGENO

3.1 Software Development Life Cycle 7

3.2 Study of the system 9

3.3 Model Performance with Cross-Validation (CV) 19

3.4 Support Vector Machine 20

5.1 UML Diagram Overview 22

5.2 Use Case Diagram 23

5.3 Class Diagram 24

5.4 Activity Diagram 25

5.5 Sequence Diagram 26

5.6 Component Diagram 27 5.7 Deployment Diagram 28

ii

#### 1.INTRODUCTION

##### INTRODUCTION

The aim of this project is to provide the complete information of the National and International statistics. The information is available country wise and player wise. By entering the data of each match, we can get all type of reports instantly, which will be useful to call back history of each player. Also the team performance in each match can be obtained. We can get a report on number of matches, wins and lost.

The project will enable the user to view:

 Every Cricketer’s Information

 Every Cricketer’s Personal Achievements

 International Travel News

 Upcoming Series

 Every Team’s Information

 ODI and test ranking

 Increment of each ball ,over, run is done automatically

 Commentary of each ball

All the users will see the same page when they enter CMS(Cricket Management System).There will be 2 options: One for the user and one for the admin. The admin is required to login through an ID and a password so as to make changes in the database or to add new information to it.

On the other hand users can retrieve the information about the different players, along with their personal details and team details as well after getting logged in to their respective accounts.

The admin here have access to edit player records, team records and other major information related to cricket management System for effective application development approach.

#### PURPOSE OF THE SYSTEM

This document aims to give a brief description about the Cricket Management System Project.

This project is very use for Cricket match broadcasters to get information quickly. Also for Cricket lovers who are very much interested in Cricket Statistics. In other words this document will provide a basis for validation and verification

With the CMS (Cricket Management System) we will provide them with capabilities and properties organized neatly. CMS which is an online intranet System will be used by anyone who wishes to find information for the players and teams.

Admin will basically query and edit the database via CMS. It will also calculate the rank of each player depending on their Strike Rate and no. of matches played.

#### 1.3 EXISTING SYSTEM

K-Nearest Neighbor:

K-Nearest neighbor is a lazy learner technique. This algorithm depends on learning by analogy. It

is a supervised classification method. This classifier is used extensively for classification purpose. This classifier waits till the last minute prior to build some model on a specified tuple as compared to earlier classifiers. The training tuples are characterized in N-dimensional space in this classifier. This classification model looks for the k training tuples nearest to the indefinite sample in case of an indefinite tuple. Then, this classifier puts the sample in the closest class.

**Disadvantages:**

Results with less accuracy as low as 50% due.

1. **Does not work well with large dataset:** In large datasets, the cost of calculating the distance between the new point and each existing point is huge which degrades the performance of the algorithm.

1. **Does not work well with high dimensions:** The KNN algorithm doesn't work well with high dimensional data because with large number of dimensions, it becomes difficult for the algorithm to calculate the distance in each dimension.

1. **Need feature scaling:** We need to do feature scaling (standardization and normalization) before applying KNN algorithm to any dataset. If we don't do so, KNN may generate wrong predictions.

1. **Sensitive to noisy data, missing values and outliers**: KNN is sensitive to noise in the dataset.

We need to manually impute missing values and remove outliers.

#### 1.4 PROPOSED SYSTEM

“Support Vector Machine” (SVM) is a supervised [machine learning algorithm w](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle)hich can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in ndimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate.

Results with accuracy are 82%.

**Advantages:**

1. SVM works relatively well when there is clear margin of separation between classes.
2. SVM is more effective in high dimensional spaces.
3. SVM is effective in cases where number of dimensions is greater than the number of samples. 4. SVM is relatively memory efficient

# 2. LITERATURE SURVEY

[1] It took almost 400 years for test cricket to start off, and thereafter it just took 94, 34 and 13 years for the introduction of ODI, T20I and T10 formats, respectively. Length of the formats is shrinking over time at an alarming rate. If the trend continues, the format of five overs or less may not be far away and cricket pundits are fearing that the longest format, that is, test cricket, may lose its relevance.

[2] Unfortunately, the management part of the cricketing ecosystem has not drawn much attention from the academic world. Since market dynamics is playing a vital role in decision-making post commercialization of the cricket, this study applies stakeholder analysis and identifies three key stakeholders, namely administrators, players and spectators, and their aspirations.

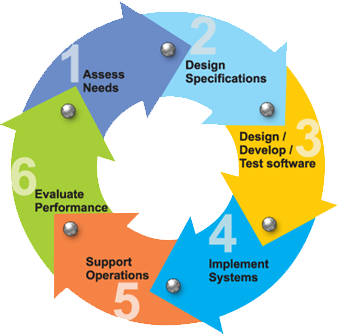
[3] It suggests a shift from the ICC-led hierarchical model to the horizontal and more democratic model for sports governance during the post-commercialization stage. The results of the study indicate that even though ICC had taken a number of measures for advancement of the sport during 2015–2018, many of them are criticized like the world test championship or day-night test matches. The study scrapes through opinions of cricket-related professionals from open sources, applies sentiment analytics to classify them, uses text summarization to extract summary viewpoints and rates them on unanimity scale.

[4] A comparison is made between the actions taken by ICC and the pool of unanimous viewpoints using evidence-based assessment (EBA). Finally, it develops a framework of six research paradigms, taking into account the stakeholders’ aspirations and the EBA outcome. If these paradigms are acted upon, it can ensure convergence of stakeholders’ goals and balance in the cricket mix. Additionally, if ICC can make unanimous viewpoints from its stakeholders a part of its future plan, acceptability of its decisions will be more.

# 3.SOFTWARE REQUIREMENT ANALYSIS

### 3.1 Software Development Life Cycle

There is various software development approaches defined and designed which are used/employed during development process of software, these approaches are also referred as "Software Development Process Models". Each process model follows a particular life cycle in order to ensure success in process of software development.



**Fig:3.1. Software Development Life Cycle.**

##### Requirements

Business requirements are gathered in this phase. This phase is the main focus of the project managers and stake holders. Meetings with managers, stake holders and users are held in order to determine the requirements. Who is going to use the system? How will they use the system? What data should be input into the system? What data should be output by the system? These are general questions that get answered during a requirement gathering phase. This produces a nice big list of functionalities that the system should provide, which describes functions the system should perform, business logic that processes data, what data is stored and used by the system, and how the user interface should work. The overall result is the system as a whole and how it performs, not how it is actually going to do it.

##### Design

The software system design is produced from the results of the requirements phase. Architects have the ball in their court during this phase and this is the phase in which their focus lies. This is where the details on how the system will work is produced. Architecture, including hardware and software, communication, software design (UML is produced here) are all part of the deliverables of a design phase.

##### Implementation

Code is produced from the deliverables of the design phase during implementation, and this is the longest phase of the software development life cycle. For a developer, this is the main focus of the life cycle because this is where the code is produced. Implementation my overlap with both the design and testing phases. Many tools exists (CASE tools) to actually automate the production of code using information gathered and produced during the design phase.

##### Testing

During testing, the implementation is tested against the requirements to make sure that the product is actually solving the needs addressed and gathered during the requirements phase. Unit tests and system/acceptance tests are done during this phase. Unit tests act on a specific component of the system, while system tests act on the system as a whole. So, in a nutshell, that is a very basic overview of the general software development life cycle model. Now let’s delve into some of the traditional and widely used variations.

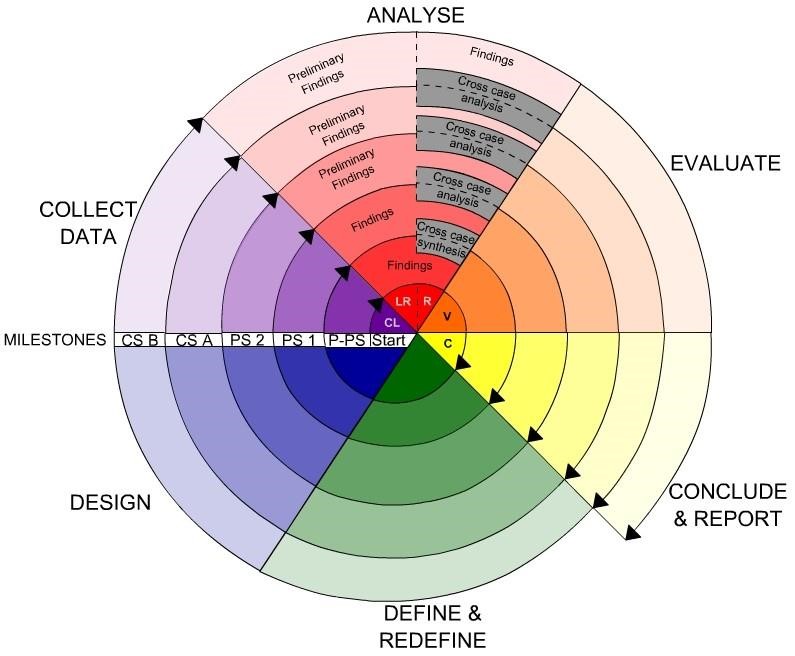
### 3.2 Study of the system

In the flexibility of uses the interface has been developed a graphics concepts in mind, associated through a browser interface. The GUI’s at the top level has been categorized as follows

1. Administrative User Interface Design
2. The Operational and Generic User Interface Design

The administrative user interface concentrates on the consistent information that is practically, part of the organizational activities and which needs proper authentication for the data collection. The Interface helps the administration with all the transactional states like data insertion, data deletion, and data updating along with executive data search capabilities.

The operational and generic user interface helps the users upon the system in transactions through the existing data and required services. The operational user interface also helps the ordinary users in managing their own information helps the ordinary users in managing their own information in a customized manner as per the assisted flexibility.



**Fig:3.2. Study of the system.**

### 3.3 Fundamental Concepts on (Domain)

##### Relational Database

A database system, also called a database management system (DBMS), consists of a collection of interrelated data, known as a database, and a set of software programs to manage and access the data A relational database is a collection of tables, each of which is assigned a unique name. Each table consists of a set of attributes (columns or fields) and usually stores a large set of tuples (records or rows).Each tuple in a relational table represents an object identified by a unique key and described by a set of attribute values A semantic data model, such as an entity-relationship (ER) data model, is often constructed for relational databases. An ER data model represents the database as a set of entities and their relationships Relational data can be accessed by database queries written in a relational query language, such as SQL.

##### Data Warehouse

A data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and that usually resides at a single site Data warehouses are constructed via a process of data cleaning, data integration, data transformation, data loading, and periodic data refreshing.

##### Object-Relational Database

Based on an object-relational data model Extends the relational model by providing a rich data type for handling complex objects and object orientation Objects that share a common set of properties can be grouped into an object class. Each object is an instance of its class. Object classes can be organized into class/subclass hierarchies.

##### Advance Data and Information System

With the progress of database technology, various kinds of advanced data and information systems have emerged and are undergoing development to address the requirements of new applications handling spatial/temporal data (such as maps) engineering design data (such as the design of buildings, system components, or integrated circuits) hypertext and multimedia data

(including text, image, video, and audio data) time-related data (such as historical records or stock exchange data) stream data (such as video surveillance and sensor data, where data flow in and out like streams) the World Wide Web (a huge, widely distributed information repository made available by the Internet).

##### The World Wide Web

The World Wide Web and its associated distributed information services, such as Yahoo!

and Google provide rich, worldwide, on-line information services, where data objects are linked together to facilitate interactive access Capturing user access patterns in such distributed information environments is called Web usage mining (or Weblog mining).

Database or data warehouse server responsible for fetching the relevant data, based on the user’s data mining request can be decouples/loose coupled/tightly coupled with the database layer. Knowledge base the domain knowledge that is used to guide the search or evaluate the interestingness of resulting patterns interestingness constraints or thresholds, metadata, concept hierarchies, etc.

Data mining engine this is essential to the data mining system and ideally consists of a set of functional modules for tasks such as characterization, association and correlation analysis, classification, prediction, cluster analysis, outlier analysis, and evolution analysis query languages (DMQL) based on mining primitives to access the data.

Pattern evaluation module interacts with the data mining modules so as to focus the search toward interesting patterns may use interestingness thresholds to filter out discovered patterns may be integrated with the mining module.

User interface communicates between users and the data mining system allows the user to interact with the system by specifying a data mining query or task, providing information to help focus the search, and performing exploratory data mining based on the intermediate data mining results allows the user to browse database and data warehouse schemas or data structures, evaluate mined patterns, and visualize the patterns in different forms.

### 3.4 System Analysis

The Systems Development Life Cycle (SDLC), or *Software Development Life Cycle* in [systems engineering](http://en.wikipedia.org/wiki/Systems_engineering)[, information systems a](http://en.wikipedia.org/wiki/Information_systems)nd [software engineering,](http://en.wikipedia.org/wiki/Software_engineering) is the process of creating or altering systems, and the models and [methodologies t](http://en.wikipedia.org/wiki/Methodologies)hat people use to develop these systems.

In software engineering the SDLC concept underpins many kinds of [software development methodologies.](http://en.wikipedia.org/wiki/Software_development_methodologies) These methodologies form the framework for planning and controlling the creation of an information system the [software development process.](http://en.wikipedia.org/wiki/Software_development_process)

**Software model and Architecture Analysis:**

Structured project management techniques (such as an SDLC) enhance management’s control over projects by dividing complex tasks into manageable sections. A software life cycle model is either a descriptive or prescriptive characterization of how software is or should be developed. But none of the SDLC models discuss the key issues like Change management, Incident management and Release management processes within the SDLC process, but, It is addressed in the overall project management. In the proposed hypothetical model, the concept of user-developer interaction in the conventional SDLC model has been converted into a threedimensional model which comprises of the user, owner and the developer. In the proposed hypothetical model, the concept of user-developer interaction in the conventional SDLC model has been converted into a three-dimensional model which comprises of the user, owner and the developer. The ―one size fits all‖ approach to applying SDLC methodologies is no longer appropriate. We have made an attempt to address the above-mentioned defects by using a new hypothetical model for SDLC described elsewhere. The drawback of addressing these management processes under the overall project management is missing of key technical issues pertaining to software development process that is, these issues are talked in the project management at the surface level but not at the ground level.

### 3.5 Functional requirements

Outputs from computer systems are required primarily to communicate the results of processing to users. They are also used to provide a permanent copy of the results for later consultation. The various types of outputs in general are:

* External Outputs, whose destination is outside the organization,.
* Internal Outputs whose destination is within organization and they are the ➢ User’s main interface with the computer.
* Operational outputs whose use is purely within the computer department.
* Interface outputs, which involve the user in communicating directly.
* Understanding user’s preferences, expertise level and his business requirements through a friendly questionnaire.
* Input data can be in four different forms - Relational DB, text files, .xls and xml files. For testing and demo you can choose data from any domain. User-B can provide business data as input.

## 3.6 Feasible Study

The project will enable the user to view:

Every Cricketer’s Information, Every Cricketer’s Personal Achievements, International Travel News, Upcoming Series, Every Team’s Information, ODI and test ranking, Increment of each ball ,over, run is done automatically, Commentary of each ball

**New Unproven Technology**

The Cricket Management System is developed on Android Studio. One of the biggest risks involved with our project is the advancement in technology. As soon as new tools and better technology will come into play in the near future, the software might become outdated and hence would not be able to meet the user expectations. Thus, the software will not be able to compete with other better software available in the market and henceforth will ultimately fail.

**o**

**Machine Learning**

There are several excellent e-books to study Machine Learning competition entries. However many will skip some of the explanation on how the solution is developed as these ebooks are developed by experts for experts. The objective of this book is to follow a step-bystep workflow, explaining each step and rationale for every decision we take during solution development.

##### How Machine Learning Beat the Odds

It's the classical problem; predict the outcome of a binary event. In laymen terms this means, it either occurred or did not occur.

For example, you won or did not win, you passed the test or did not pass the test, you were accepted or not accepted, and you get the point.

A common business application is churn or customer retention. Another popular use case is healthcare's mortality rate or survival analysis. Binary events create an interesting dynamic, because we know statistically, a random guess should achieve a 50% accuracy rate, without creating one single algorithm or writing one single line of code. However, just like autocorrect spell-check technology, sometimes we humans can be too smart for our own good and actually underperform a coin flip.

##### A Machine Learning Framework

The competition solution workflow goes through seven stages described as

1. Question or problem definition.
2. Acquire training and testing data.
3. Wrangle, prepare, and cleanse the data.
4. Analyze, identify patterns, and explore the data.
5. Model, predict and solve the problem.
6. Visualize, report, and present the problem solving steps and final solution.
7. Supply or submit the results.

The workflow indicates general sequence of how each stage may follow the other. However there are use cases with exceptions.

* We may combine multiple workflow stages. We may analyze by visualizing data.
* Perform a stage earlier than indicated. We may analyze data before and after wrangling.
* Perform a stage multiple times in our workflow. Visualize stage may be used multiple times.
* Drop a stage altogether. We may not need supply stage to productize or service enable our dataset for a competition.

1. **Define the Problem:** If data science, big data, machine learning, predictive analytics, business intelligence, or any other buzzword is the solution, then what is the problem? As the saying goes, don't put the cart before the horse. Problems before requirements, requirements before solutions, solutions before design, and design before technology. Too often we are quick to jump on the new shiny technology, tool, or algorithm before determining the actual problem we are trying to solve.
2. **Gather the Data:** John Naisbitt wrote in his 1984 (yes, 1984) book Megatrends, we are “drowning in data, yet staving for knowledge." So, chances are, the dataset(s) already exist somewhere, in some format. It may be external or internal, structured or unstructured, static or streamed, objective or subjective, etc. As the saying goes, you don't have to reinvent the wheel, you just have to know where to find it. In the next step, we will worry about transforming "dirty data" to "clean data."
3. **Prepare Data for Consumption:** This step is often referred to as data wrangling, a required process to turn “wild” data into “manageable” data. Data wrangling includes implementing data architectures for storage and processing, developing data governance standards for quality and control, data extraction (i.e. ETL and web scraping), and data cleaning to identify aberrant, missing, or outlier data points.
4. **Perform Exploratory Analysis:** Anybody who has ever worked with data knows garbage-in, garbage-out (GIGO). Therefore, it is important to deploy descriptive and graphical statistics to look for potential problems, patterns, classifications, correlations and comparisons in the dataset. In addition, data categorization (i.e. qualitative vs quantitative) is also important to understand and select the correct hypothesis test or data model.
5. **Model Data:** Like descriptive and inferential statistics, data modeling can either summarize the data or predict future outcomes. Your dataset and expected results, will determine the algorithms available for use. It's important to remember, algorithms are tools and not magical wands or silver bullets. You must still be the master craft (wo)man that knows how-to select the right tool for the job. An analogy would be asking someone to hand you a Philip screwdriver, and they hand you a flathead screwdriver or worst a hammer. At best, it shows a complete lack of understanding. At worst, it makes completing the project impossible. The same is true in data modeling.

The wrong model can lead to poor performance at best and the wrong conclusion (that’s used as actionable intelligence) at worst.

1. **Validate and Implement Data Model:** After you've trained your model based on a subset of your data, it's time to test your model. This helps ensure you haven't over fit your model or made it so specific to the selected subset, that it does not accurately fit another subset from the same dataset. In this step we determine if our [model over fit, generalize, or under fit our dataset.](http://docs.aws.amazon.com/machine-learning/latest/dg/model-fit-underfitting-vs-overfitting.html)
2. **Optimize and Strategize:** This is the "bionic man" step, where you iterate back through the process to make it better...stronger...faster than it was before. As a machine learning engineer or data scientist, your strategy should be to outsource developer operations and application plumbing, so you have more time to focus on recommendations and design.

Once you're able to package your ideas, this becomes your “currency exchange" rate.

**Split Training and Testing Data:**

As mentioned previously, the test file provided is really validation data for competition submission. So, we will use sklearn function to split the training data in two datasets; 75/25 split. This is important, so we don't [over fit our model.](https://www.coursera.org/learn/python-machine-learning/lecture/fVStr/overfitting-and-underfitting) Meaning, the algorithm is so specific to a given subset, it cannot accurately generalize another subset, from the same dataset. It's important our algorithm has not seen the subset we will use to test, so it doesn't "cheat" by memorizing the answers. We will use [sklearn's train\_test\_split function.](http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html) In later sections we will also use [sklearn's cross validation functions,](http://scikit-learn.org/stable/modules/cross_validation.html#cross-validation) that splits our dataset into train and test for data modeling comparison.

**Model Data:**

Data Science is a multi-disciplinary field between mathematics (i.e. statistics, linear algebra, calculus, etc.), computer science (i.e. programming languages, computer systems, etc.) and business management (i.e. communication, subject-matter knowledge, etc.). Most data scientist comes from one of the three fields, so they tend to lean towards that discipline. However, data science is like a three-legged stool, with no one leg being more important than the other. So, this step will require advanced knowledge in mathematics

Machine Learning (ML), as the name suggest, is teaching the machine how-to think and not what to think. While this topic and big data has been around for decades, it is becoming more popular than ever because the barrier to entry is lower, for businesses and professionals alike.

This is both good and bad. It’s good because these algorithms are now accessible to more people that can solve more problems in the real-world

First, you must understand, that the purpose of machine learning is to solve human problems. Machine learning can be categorized as: supervised learning, unsupervised learning, and reinforced learning.

Supervised learning is where you train the model by presenting it a training dataset that includes the correct answer.

Unsupervised learning is where you train the model using a training dataset that does not include the correct answer.

And reinforced learning is a hybrid of the previous two, where the model is not given the correct answer immediately, but later after a sequence of events to reinforce learning.

We are doing supervised machine learning, because we are training our algorithm by presenting it with a set of features and their corresponding target. We then hope to present it a new subset from the same dataset and have similar results in prediction accuracy.

There are many machine learning algorithms, however they can be reduced to four categories: classification, regression, clustering, or dimensionality reduction, depending on your target variable and data modeling goals. We can generalize that a continuous target variable requires a regression algorithm and a discrete target variable requires a classification algorithm. One side note, logistic regression, while it has regression in the name, is really a classification algorithm. We will use cross validation and scoring metrics, discussed in later sections, to rank and compare our algorithms’ performance.

**Machine Learning Selection:**

* [Sklearn Estimator Overview](http://scikit-learn.org/stable/user_guide.html)
* [Sklearn Estimator Detail](http://scikit-learn.org/stable/modules/classes.html)
* [Choosing Estimator Mind Map](http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html)

Now we identified that our solution is supervised learning classification algorithm. We can narrow our list of choices.

**Machine Learning Classification Algorithms:**

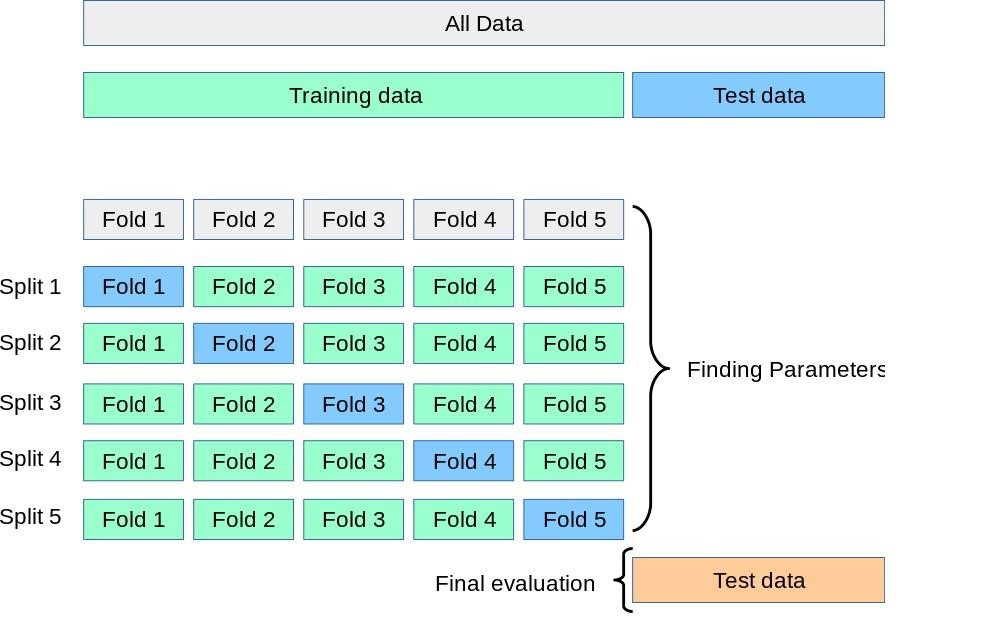
* [Ensemble Methods](http://scikit-learn.org/stable/modules/classes.html#module-sklearn.ensemble)
* [Generalized Linear Models (GLM)](http://scikit-learn.org/stable/modules/classes.html#module-sklearn.linear_model)
* [Naive Bayes](http://scikit-learn.org/stable/modules/classes.html#module-sklearn.naive_bayes)
* [Nearest Neighbors](http://scikit-learn.org/stable/modules/classes.html#module-sklearn.neighbors)
* [Support Vector Machines (SVM)](http://scikit-learn.org/stable/modules/classes.html#module-sklearn.svm)
* [Decision Trees](http://scikit-learn.org/stable/modules/classes.html#module-sklearn.tree)
* [Discriminant Analysis](http://scikit-learn.org/stable/modules/classes.html#module-sklearn.discriminant_analysis)

##### Model Performance with Cross-Validation (CV)

We used [sklearn cross\_validate f](http://scikit-learn.org/stable/modules/cross_validation.html#multimetric-cross-validation)unction to train, test, and score our model performance.

Remember, it's important we use a different subset for train data to build our model and test data to evaluate our model. Otherwise, our model will be overfitted. Meaning it's great at "predicting" data it's already seen, but terrible at predicting data it has not seen; which is not prediction at all. It's like cheating on a school quiz to get 100%, but then when you go to take the exam, you fail because you never truly learned anything. The same is true with machine learning.

In addition to CV, we used a customized [sklearn train test splitter,](http://scikit-learn.org/stable/modules/classes.html#module-sklearn.model_selection) to allow a little more randomness in our test scoring. Below is an image of the default CV split.



**Fig:3.3. Model Performance with Cross-Validation (CV)**

##### Validate and Implement

The next step is to prepare for submission using the validation data. Now we are ready to train a model and predict the required solution. There are 60+ predictive modeling algorithms to choose from. We must understand the type of problem and solution requirement to narrow down to a select few models which we can evaluate. Our problem is a classification and regression problem. We want to identify relationship between output with other variables or features. We are also performing a category of machine learning which is called supervised learning as we are training our model with a given dataset. With these two criteria - Supervised Learning plus Classification and Regression, we can narrow down our choice of models to a few. These include:

* Logistic Regression
* KNN or k-Nearest Neighbors
* Support Vector Machines
* Naive Bayes classifier
* Decision Tree
* Random Forrest
* Perceptron
* Artificial neural network

Linear Regression is a useful model to run early in the workflow. Linear regression measures the relationship between the numerical dependent variable (feature) and one or more independent variables (features) by estimating values, which is based on straight line equation.

Note the confidence score generated by the model based on our training dataset.

The next model Random Forests is one of the most popular. Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees (n\_estimators=100) at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

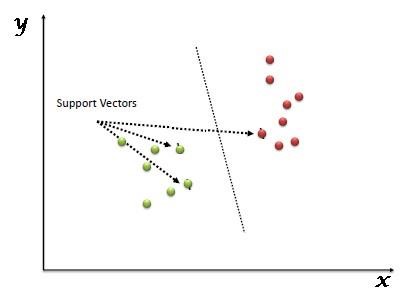
The model confidence score is the highest among models evaluated so far. We decide to use this model's output (Y\_pred) for creating our competition submission of results.

##### Model evaluation

We can now rank our evaluation of all the models to choose the best one for our problem. While both Decision Tree and Random Forest score the same, we choose to use Random Forest as they correct for decision trees' habit of over fitting to their training set.

##### Support Vector Machine Algorithm

“Support Vector Machine” (SVM) is a supervised [machine learning algorithm w](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle)hich can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in ndimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).



**Fig: 3.4 Support Vector Machine**

# 4.SOFTWARE AND HARDWARE REQUIREMENTS

### 4.1. SOFTWARE REQUIREMENTS

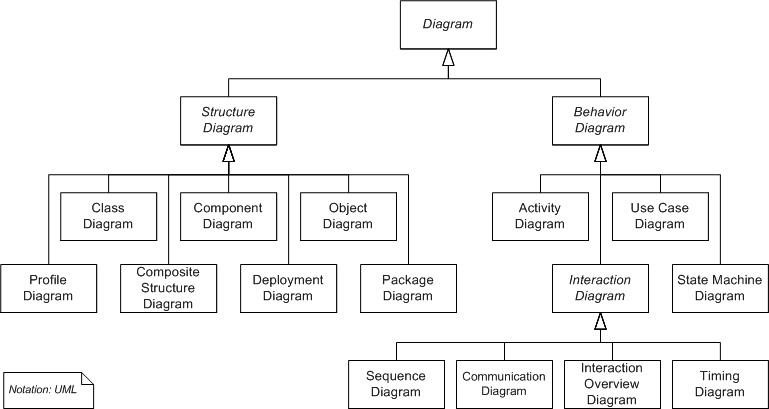
* Operating System: Windows 7 Ultimate.
* Coding Language: Python.
* Front-End: Python.
* Designing: Html, CSS, JavaScript.
* Data Base: MySQL

### 4.2. HARDWARE REQUIREMENTS

* System: Pentium IV 2.4 GHz
* Hard Disk: 40GB
* Floppy Drive: 1.44 MB
* RAM: 512 MB

# 5.SOFTWARE DESIGN

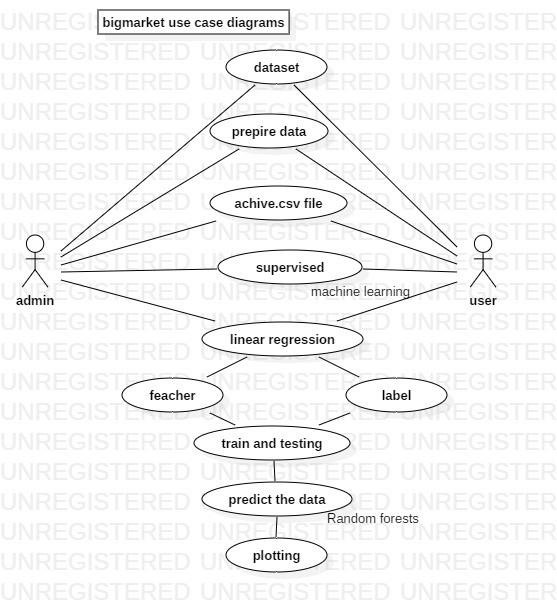
### 5.1 UML Diagrams Overview



##### Fig: 5.1 UML Diagram Overview

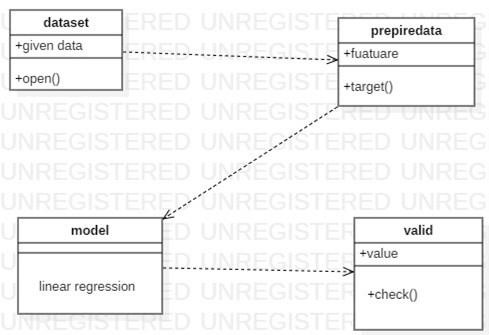
UML combines best techniques from data modeling (entity relationship diagrams), business modeling (work flows), object modeling, and component modeling. It can be used with all processes, throughout the software development life cycle, and across different implementation technologies. UML has synthesized the notations of the Booch method, the Object-modeling technique (OMT) and Object-oriented software engineering (OOSE) by fusing them into a single, common and widely usable modeling language. UML aims to be a standard modeling language which can model concurrent and distributed systems.

**5.2 USE CASE DIAGRAM:**



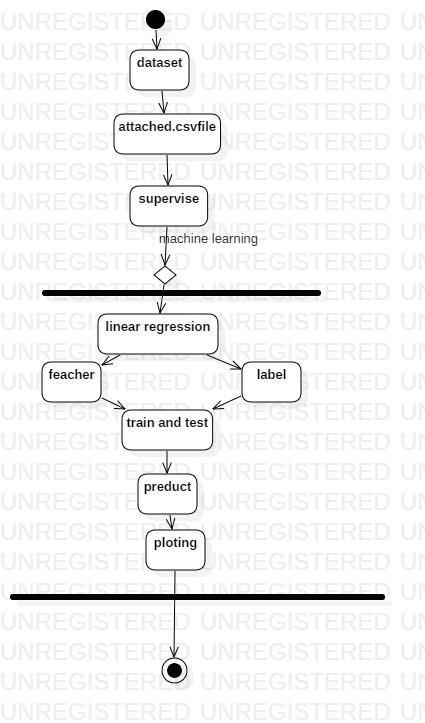
### Fig.5.2 Use Case Diagram

**5.3 CLASS DAIGRAM:**



### Fig: 5.3 Class diagram

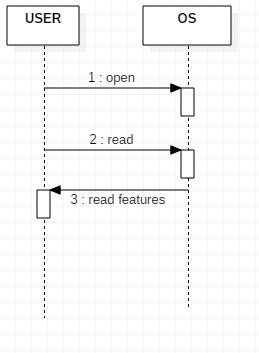
**5.4 ACTIVITY DIAGRAM:**



### Fig: 5.4 Activity diagram

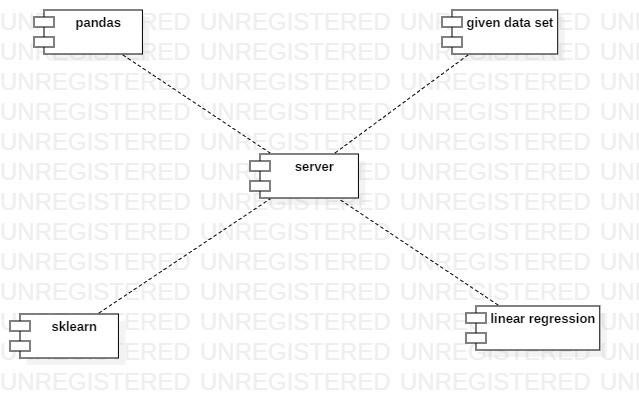
**5.5 SEQUENCE DIAGRAM:**

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows, as parallel vertical lines ("lifelines"), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner.



## Fig: 5.5 Sequence diagram

**5.6 COMPONENT DAIGRAM:**

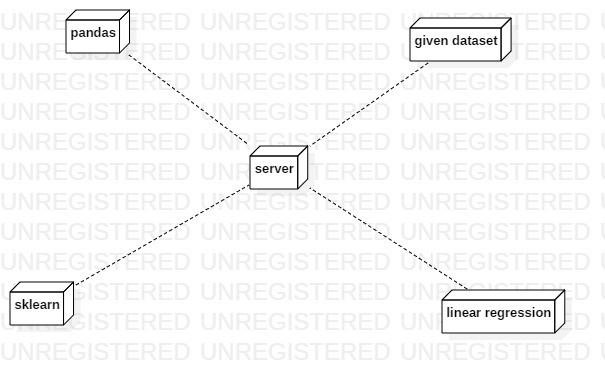


### Fig: 5.6 Component diagram

### 5.7 DEPLOYMENT DIAGRAM:



**Fig:5.7 Deployment diagram**



### 6.CODING

**6.1 Coding:**

1. #include<iostream.h>

2. #include<conio.h>

3. Using namespace std;

4. void add\_player() 5. {

6. int pid, dob, no\_of\_centuries;

7. char pname, bat\_style, bowl\_style;

8. cout<<”enter player id”;

9. cin>>pid;

10. cout<<”enter player name”;

11. cin>>pname;

12. cout<<”enter dob of player”;

13. cin>>dob;

14. cout<<”enter bat style”;

15. cin>>bat\_style;

16. cout<<”enter bowl tyle”;

17. cin>>bowl\_style;

18. cout<<”enter no. of centuries”;

19. cin>>no\_of\_centuries;

20. }

21. void del\_player(int y)

22. {

23. if(pid==y)

24. cout<<”player details deleted successfully”;

25. }

26. void add\_team()

27. {

28. int tid, bats, bowls;

29. char tname, captain, tcol, cname;

30. cout<<”enter team id”;

31. cin>>tid;

32. cout<<”enter team name”;

33. cin>>tname;

34. cout<<”enter name of captain”;

35. cin>>captain;

36. cout<<”enter no. of batsmen”;

37. cin>>bats;

38. cout<<”enter no. of bowlers”;

39. cin>>bowls;

40. cout<<”enter team color”;

41. cin>>tcol;

42. cout<<”enter coach name”;

43. cin>>cname;

44. }

45. void del\_team(int x)

46. {

47. if(tid==x)

48. cout<<”team details deleted successfully”;

49. }

50. void fetch\_player()

51. {

52. int pid, dob, no\_of\_centuries;

53. char pname, bat\_style, bowl\_style;

54. cout<<” player id is:”;

55. cout<<pid;

56. cout<<”player name is:”;

57. cout<<pname;

58. cout<<” dob of player is:”;

59. cout<<dob;

60. cout<<” bat style is:”;

61. cout<<bat\_style;

62. cout<<”bowl style is:”;

63. cout<<bowl\_style;

64. cout<<”no. of centuries are:”;

65. cout<<no\_of\_centuries;

66. }

67. void fetch\_team()

68. {

69. int tid, bats, bowls;

70. char tname, captain, tcol, cname;

71. cout<<”team id is:”;

72. cout<<tid;

73. cout<<” team name is:”;

74. cout<<tname;

75. cout<<”name of captain is:”;

76. cout<<captain;

77. cout<<”no. of batsmen are:”;

78. cout<<bats;

79. cout<<”no. of bowlers is:”;

80. cout<<bowls;

81. cout<<”team color is:”;

82. cout<<tcol;

83. cout<<”coach name is:”;

84. cout<<cname;

85. }

86. int main()

87. {

88. int choice, ch, ch1;

89. Cout<<”1. LOGIN AS USER”;

90. Cout<<”2. LOGIN AS ADMIN”;

91. Cout<<”enter your choice”;

92. Cin>>choice;

93. switch(choice)

94. {

95. case 1: cout<<”entering as user”;

96. cout<<”1. fetch player details”;

97. Cout<<”2. fetch team details”;

98. Cout<<”enter your choice”;

99. Cin>>ch;

100. switch(ch)

101. {

102. case 1: cout<<”\*\*\*\*\*Displaying player details\*\*\*\*\*”;

103. fetch\_player();

104. break;

105. case 2: cout<<”\*\*\*\*\*Displaying team details\*\*\*\*\*”;

106. fetch\_team();

107. break;

108. default : cout<<”wrong choice”;

109. break;

110. }

111. case 2: cout<<”entering as admin”;

112. cout<<”1. Add player details”;

113. cout<<”2. Delete player details”;

114. cout<<”3. Add team details”;

115. cout<<”4. Delete team details”;

116. cout<<”enter your choice”;

117. cin>>ch1;

118. switch(ch1)

119. {

120. case 1: cout<<”\*\*\*\*\*Adding player details\*\*\*\*\*”;

121. add\_player();

122. break;

123. case 2: cout<<”\*\*\*\*\*Deleting player details\*\*\*\*\*”;

124. del\_player();

125. break;

126. case 3: cout<<”\*\*\*\*\*Adding team details\*\*\*\*\*”;

127. add\_team();

128. break;

129. case 4: cout<<”\*\*\*\*\*Deleting team details\*\*\*\*\*”;

130. del\_team();

131. break;

132. default : cout<<”wrong choice”;

133. break;

134. }

135. }

136. }

# 7.TESTING

**SOFTWARE TESTING**

Testing is done with an objective of finding most errors with minimum amount of time and effort.

WHITE BOX testing sometimes called glass-box testing, is a test-case design philosophy

that uses the control structure described as part of component-level design to derive test cases. Using white-box testing methods, you can derive test cases that :

(1) Guarantee that all independent paths within a module have been exercised at least once.

(2) Exercise all logical decisions on their true and false sides.

(3) Execute all loops at their boundaries and within their operational bounds.

(4) Exercise internal data structures to ensure their validity.

Basic Path Testing is a White Box testing technique that enables to derive logical complexity and defines basic test of execution paths. The test cases are prepared so that each execution path will occur at least once.

• To Design Program Graph

1. int main(intballs\_played , inttot\_runs) 2. {

3. float strike\_rate ;

4. char rank;

5. strike\_rate=(tot\_runs/balls\_played)\*100;

6. cout<<”the strike rate of the player is:”<<strike\_rate;

7. if((strike\_rate<=300 &&strike\_rate>=200)&&tot\_runs>3000)

8. {

9. rank=’H’;

10. cout<<”the player has the highest rank”<<rank;

11. }

12. else if((strike\_rate<=200 &&strike\_rate>=100)&&tot\_runs>2000)

13. {

14. rank=’A’;

15. cout<<”the player has an average rank”<<rank;

16. }

17. else if((strike\_rate<=100 &&strike\_rate>=50)&&tot\_runs>1000)

18. {

19. rank=’B’;

20. cout<<”the player has a below average rank”<<rank;

21. }

22. Else

23. {

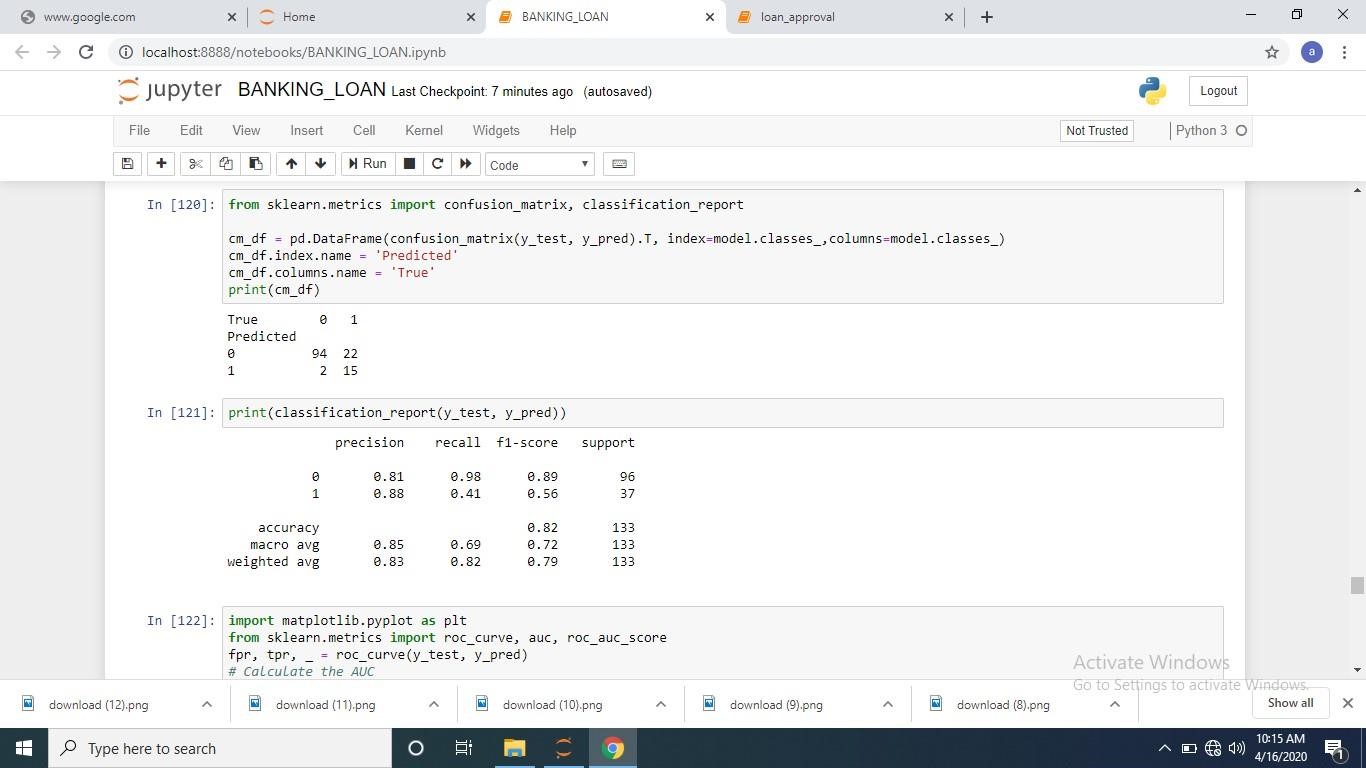
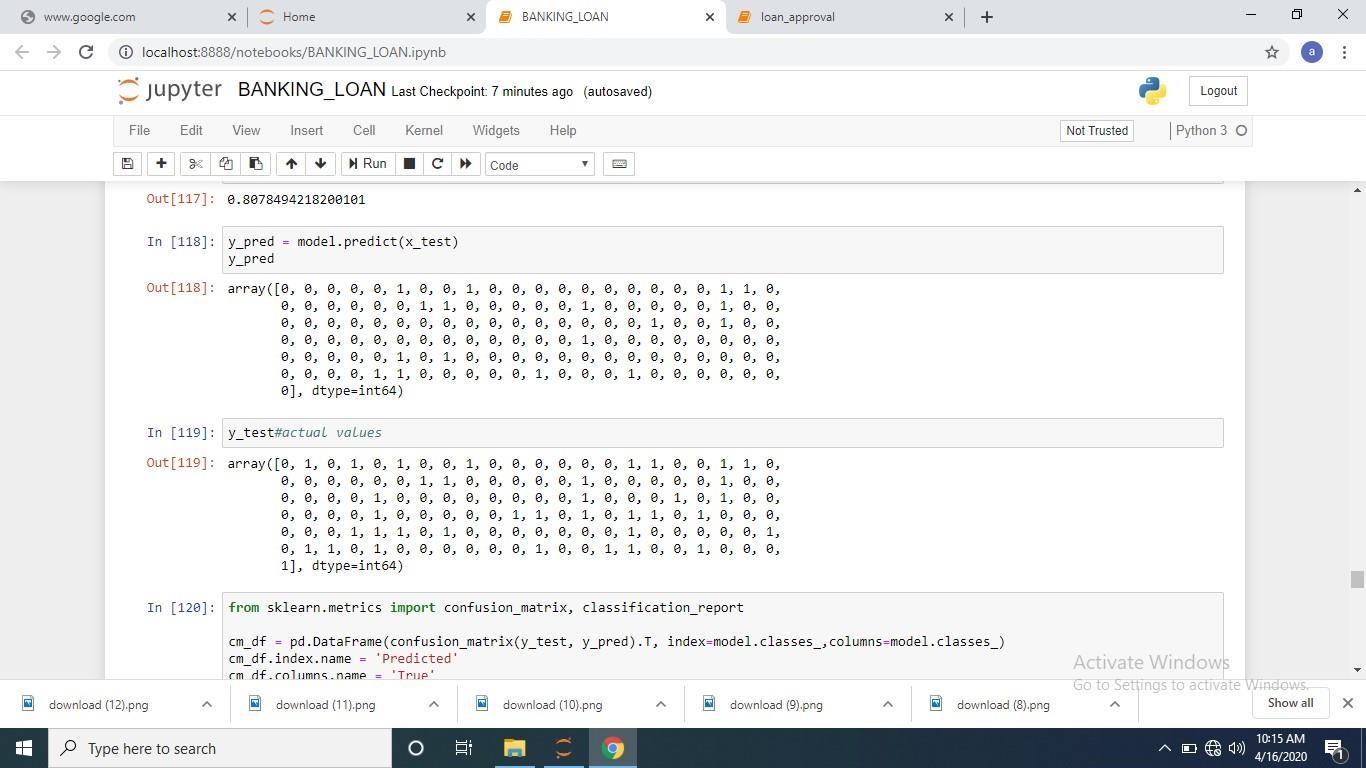
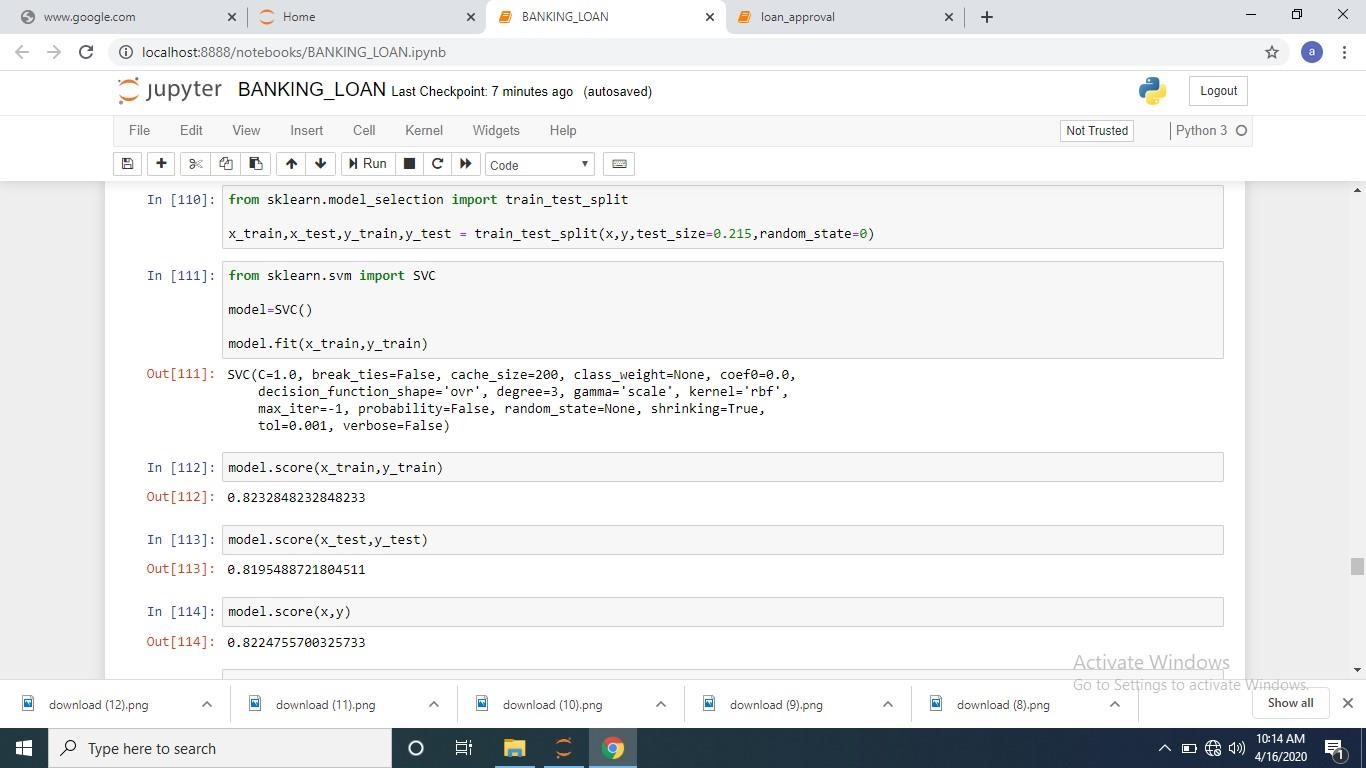
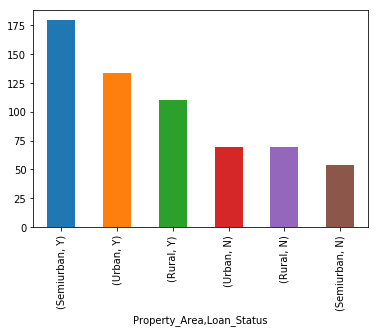
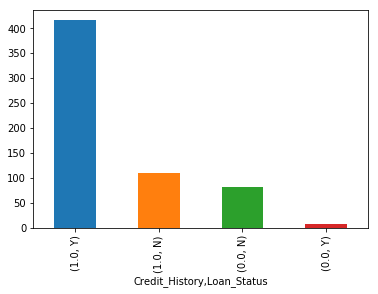
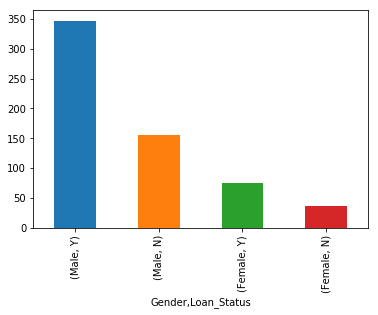
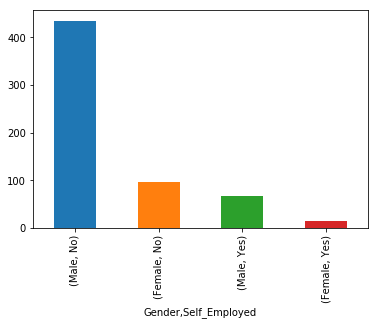
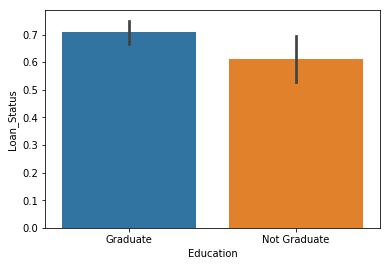
24. rank=’L’;

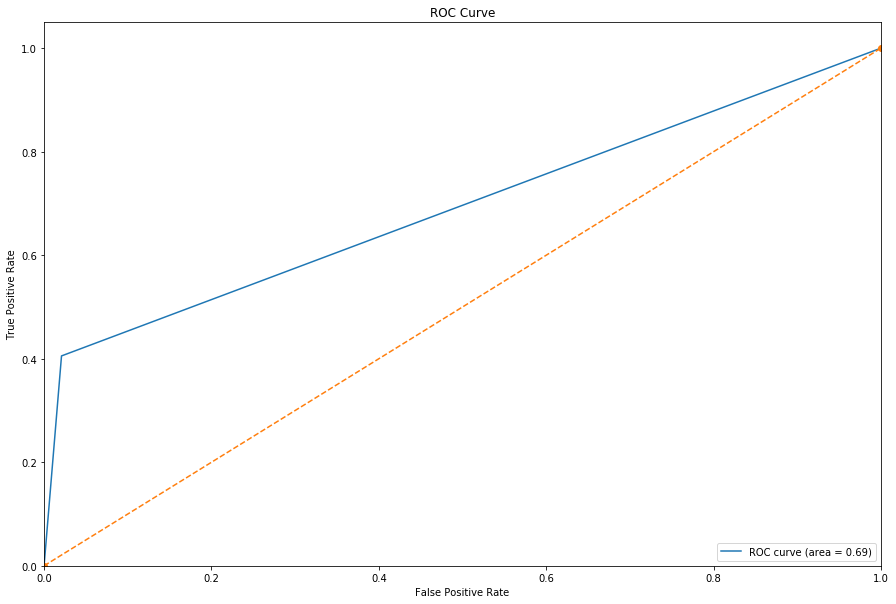
25. cout<<”the player has the least rank”<<rank;

26. }

27. }

**8.OUTPUT SCREEN**





# 9.CONCLUSION

We can conclude that the Cricket Management System can be used by people all across the globe to access the information of various players and teams to retrieve details along with the player personal details also. It can help in accessing new data and news about favourite players and teams. It is a modern approach to fetch data of players and teams, along with live scores.

It is based on advanced technology as it is compatible with all mobile sets having android version 4.4 or higher. Some of the advantages of CMS (Cricket Management System) are as follows:-

• Can ODI Tournament details on Internet.

• Players, Organizers, Selectors, Fan and Followers can access it from anywhere and anytime.

• The ODI Tournament can reach thousands of online users and gain major popularity.

• User-Friendly admin and easy to manage.

• Admin can add and edit information at any point of time and also from any location.

• Can share ODI Details through Social Media: Comments, posting, etc.

• Latest cricket news updates, ongoing ODI Events information is easy to retrieve.

**10.REFERENCES**

The successful completion of this project has been achieved by the assistance from various resources which include:

1. https://www.tutorialsduniya.com

2. Software Engineering: A Practitioners Approach by Roger S. Pressman

3. An Integrated Approach to Software Engineering by P. Jalote

4. Software Engineering by K.K. Aggarwal

5. https://www.tutorialsduniya.com