SIR C. R. REDDY COLLEGE OF ENGINEERING

ELURU-534007, W.G.DIST, A.P.

(Approved by AICTE, accredited by NAAC and Affiliated to JNTUK) (2021-2022)

ELECTRONICS AND COMMUNICATION ENGINEERING

MACHINE LEARNING WITH PYTHON

A Project submitted in accordance with the internship program regarding ML

Submitted by:

G. Surendra Babu
(20B81A0456)

G.Vinodh Chowdary
(20B81A0456)

G. Chinmaih Ganga Devi
(20B81A0460)

G.Chitti Babu
(20B81A0452)

D.Sowjanya
(20B81A0446)

D.Pavani
(20B81A0445)

SIX WEEKS INTERNSHIP TRAINING ON MACHINE LEARNING WITH PYTHON

Date: 06-07-2022 To 28-08-2022

Course: B.Tech 3rd year 1st Semester (R20 Regulation)

Academic year: 2022-2023

ABSTRACT

Now-a-days obtaining loans from banks has become a very common phenomenon. The banks gain profits from the loans lent to their customers in the form of interest. While approving a loan, the banks should consider many factors such as credit history and score, reputation of the person, the location of the property and the relationship with the bank. Many people apply for loans in the name of home loan, car loan and many more. Everyone cannot be approved based on above mentioned conditions. There are so many cases where applicant's applications for loans are not approved by various finance companies. The right predictions whether to give a loan to a customer or not is very important for the banks to maximize the profits. The idea behind this project is to use machine learning techniques to predict whether a customer can get a loan from a bank or not

TABLE OF CONTENTS

LIST OF FIGURES

1.INTRODUCTION

Finance companies deals with all kinds of loans such as house loans, vehicle loans, educational loans, personal loans etc... And has a presence across areas such as cities, towns and village areas. A Customer-first requests for a loan and after that Finance Company validates the customer eligibility for the loan approvel. Details like marital status, gender, education, and number of dependents, Income, Loan Amount, credit history, and others are given in the form to fill up by the applicants. Therefore, a robust model is built taking those details as input to verify whether an applicant is eligible to apply for loan or not. The target variable here is Applicants "Loan Status" and the other variables are predictors. After building the Machine Learning model a Web Application is to be developed for a user interface that allows the user to see instantly if he/she is eligible to get a loan by entering the given details.

.

1.1 MOTIVATION

Loan prediction is a very common real-life problem thatevery finance company faces in their lending operations. If theloan approval process is automated, it can save a lot of manhours and improve the speed of service to the customers. Theincrease in customer satisfaction and savings in operationalcosts are significant[9]. However, the benefits can only bereaped if the bank has a robust model to accurately predictwhich customer's loan it should approve and which to reject, inorder to minimize the risk of loan default.

1.2 Existing System

Existing system uses machine learning algorithms like decision tree and random forest. Although we can get good efficiency using the Random Forest classifier and Decision Tree classifier, we can get great results by using Naïve Bayes classifier.The lender must manually review each application, supported the main points provided by the applicant like gender, legal status, education, number of dependents, income, loan amount, credit history, etc., and judge if they’re creditworthy or not but it gives less accurate

result.

1.2.1 Limitations of existing system

- less efficiency

1.3 Objectives

Prediction of granting the loan to the customers by the bank is the proposed model. Classification is the target for developing the model and hence using Logistic Regression with sigmoid function is used for developing the model. Preprocessing is the major area of the model where it consumes more time and then Exploratory Data Analysis which is followed by Feature Engineering and then Model Selection. Feeding thetwo separate datasets to the model, and then preceding the model.

1.4 Outcomes

The process of prediction starts from cleaning and processing of data, imputation of missing values, experimental analysis of data set and then model building to evaluation of model and testing on test data. On Data set, the best case accuracy obtained on the original data set is 0.811. The following conclusions are reached after analysis that thoseapplicants whose credit score was worstwill fail to get loan approval, due to a higher probability of not paying back the loan amount.

1.5Applications

This application can use in baking sector

1.6 STRUCTURE OF PROJECT (SYSTEMANALYSIS)

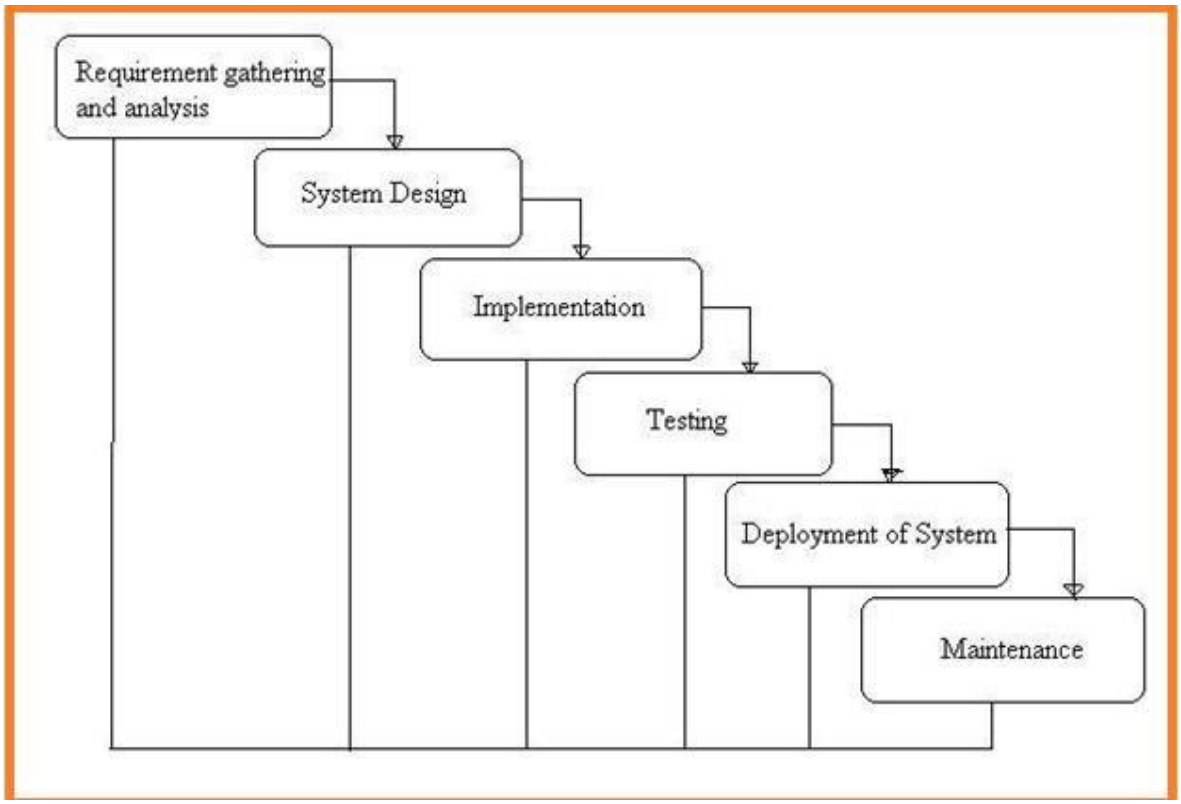


Fig: 1 Project SDLC

•Project Requisites Accumulating andAnalysis

•Application SystemDesign

•PracticalImplementation

•Manual Testing of MyApplication

•Application Deployment ofSystem

•Maintenance of theProject

1.6.1 REQUISITES ACCUMULATING AND ANALYSIS

It's the first and foremost stage of the any project as our is a an academic leave for requisites amassing we followed of IEEE Journals and Amassed so many IEEE Relegated papers and final culled a Paper designated "Individual web revisitation by setting and substance importance input and for analysis stage we took referees from the paper and did literature survey of some papers and amassed all the Requisites of the project in this stage

1.6.2 SYSTEMDESIGN

In System Design has divided into three types like GUI Designing, UML Designing with avails in development of project in facile way with different actor and its utilizer case by utilizer case diagram, flow of the project utilizing sequence, Class diagram gives information about different class in the project with methods that have to be utilized in the project if comes to our project our UML Will utilizable in this way The third and post import for the project in system design is Data base design where we endeavor to design data base predicated on the number of modules in ourproject

1.6.3 IMPLEMENTATION

The Implementation is Phase where we endeavor to give the practical output of the work done in designing stage and most of Coding in Business logic lay coms into action in this stage its main and crucial part of the project

1.6.4TESTING UNITTESTING

It is done by the developer itself in every stage of the project and fine-tuning the bug and module predicated additionally done by the developer only here we are going to solve all the runtime errors

MANUAL TESTING

As our Project is academic Leave, we can do any automatic testing so we follow manual testing by endeavor and error methods

1.6.4 DEPLOYMENT OF SYSTEM AND MAINTENANCE

Once the project is total yare, we will come to deployment of client system in genuinely world as its academic leave we did deployment i our college lab only with all need Software's withhaving Windows OS.

The Maintenance of our Project is one-time process only

1.7 FUNCTIONALREQUIREMENTS

1.Data Collection

2.DataPreprocessing

3.Training And Testing

4.Modiling

5.Predicting

1.8 NON FUNCTIONALREQUIREMENTS

NON-FUNCTIONAL REQUIREMENT (NFR) specifies the quality attribute of a software system. They judge the software system based on Responsiveness, Usability, Security, Portability and other non-functional standards that are critical to the success of the software system. Example of nonfunctional requirement, *"how fast does the website load?"* Failing to meet non-functional requirements can result in systems that fail to satisfy user needs. Non- functional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs. Example, the site should load in 3 seconds when the number of simultaneous users are> 10000. Description of non-functional requirements is just as critical as a functionalrequirement.

- Usabilityrequirement
- Serviceabilityrequirement
- Manageabilityrequirement
- Recoverabilityrequirement
- Securityrequirement
- Data Integrity requirement
- Capacityrequirement
- Availabilityrequirement
- Scalabilityrequirement
- Interoperabilityrequirement
- Reliabilityrequirement
- Maintainabilityrequirement
- Regulatoryrequirement
- Environmentalrequirement

1.8.1 EXAMPLES OF NON-FUNCTIONALREQUIREMENTS

Here, are some examples of non-functional requirement:

1.8.1.1 Users must upload dataset

1.8.1.2 The software should be portable. So moving from one OS to other OS does not createany problem.

1.8.1.3 Privacy of information, the export of restricted technologies, intellectual property rights, etc. should beaudited.

1.8.2 ADVANTAGES OF NON-FUNCTIONALREQUIREMENT

Benefits/pros of Non-functional testing are:

- The nonfunctional requirements ensure the software system follow legal andcompliance rules.
- They ensure the reliability, availability, and performance of the softwaresystem
- They ensure good user experience and ease of operating thesoftware.
- They help in formulating security policy of the softwaresystem.

1.8.3 DISADVANTAGES OF NON-FUNCTIONALREQUIREMENT

Cons/drawbacks of Non-function requirement are:

- None functional requirement may affect the various high-level softwaresubsystem
- They require special consideration during the software architecture/high-level design phase which increasescosts.
- Their implementation does not usually map to the specific softwaresub-system,
- It is tough to modify non-functional once you pass the architecturephase.

1.8.4 KEYLEARNING

The character of the time period, the length of road, the weather, the bus speed and the rate of road usage are adopted as input vectors in Support Vector Machine

2.LITERATURESURVEY

[1] Raj, J. S., &Ananthi, J. V., "Recurrent neural networks and nonlinear prediction in support vector machine" Journal of Soft Computing Paradigm (JSCP), 1(01), 33-40, 2019.

The detection of edges is the one of the important stage in the application, associated with the machine vision, computer vision and the image processing. It is most commonly and highly preferred in the area were the extraction or the detection of the attribute are necessary. As the manual methods of diagnosis in the medical images acquired from the CT (computed tomography) and the MRI (magnetic resonance images) are very tedious and as well as time consuming, the paper puts forth the methodology to detect the edges in the CT and the MRI by employing Gabor Transform as well as the soft and the hard clustering. This proposed method is highly preferred among the image with dynamic variations. The technique used in the paper is evaluated using 4500 instance of the MRI and 3000 instance of CT. The results on the basis of the figure of merit (FOM) and Misclassification rate (MCR) are compared with other standard approaches and the performance was evinced.

[2] X.FrencisJensy, V.P.Sumathi,Janani Shiva Shri, "An exploratory Data Analysis for Loan Prediction based on nature of clients", International Journal of Recent Technology and Engineering (IJRTE),Volume-7 Issue-4S, November 2018.

The term banking can be defined as receiving and protecting money that is deposited by the individual or the entities. This also includes lending money to the people which will be repaid within the given time. Banking sector is regulated in most of the countries as it is the important factor in determining the financial stability of the country. The provision of banking regulation act allows public to obtain loans.Loans are good sum of money borrowed for a

period and expected to be paid back at given interest rate. The purpose of the loan can be anything based on the customer requirements. Loans are broadly divided as openended and close-ended loans. Open-ended loans are the loans for which the client has approval for a specific amount. Examples of open-end loans are credit cards and a home equity line of credit (HELOC). Close-ended loans decreases with each payment. In other words, it is a legal term that cannot be modified by the borrower. Personal loans, mortgages, auto payments, instalment loan and student loans are the most common examples of close-ended loans. Secured or collateral loan are those loans that are protected by an asset. Houses, Vehicles, Savings accounts are the personal properties used to secure the loan. Unsecured loans are also known as personal or signature loans. Here the lender believes that the borrower can repay the loan based on financial resources possessed by the borrower. Liquidity risk is the risk that arises from the lackof marketability of an investment that cannot be bought or sold quickly enough to prevent or minimize a loss. The interest rate risk is the risk in which the interest rates priced on loans will be too low to earn the bank money. Revised Version Manuscript Received on 25 November, 2018. Ms.X.FrancisJency, CSE Department, Kumaraguru College of Technology, Coimbatore, India Ms.V.P.Sumathi, CSE Department, Kumaraguru College of Technology, Coimbatore, India Janani Shiva Sri,C S Department, Kumaraguru College of Technology, Coimbatore, India The primary objective of the bank is to provide their wealth in the safer hands. In recent times, banks approve loan after verifying and validating the documents provided by the customer. Yet there is no guarantee whether the applicant is deserving or not. This paper classifies the customers based on certain criteria. The classification is done using Exploratory Data Analysis. Exploratory Data Analysis (EDA) is an approach to analyse the datasets that summarizes the main characteristics with visual methods. The purpose of using EDA is to uncover the underlying structure of a relatively larger set of variables using visualizing techniques.

[3] PidikitiSupriya, MyneediPavani, NagarapuSaisushma,NamburiVimalaKumari, k Vikash,"Loan Prediction by using Machine Learning Models", International Journal of Engineering and Techniques.Volume 5 Issue 2, Mar-Apr 2019

With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this project we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this project is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing. In this paper we are predict the loan data by using some machine learning algorithms they are classification, logic regression, Decision Tree and gradient boosting.

[4]Shiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng,“ CNN for situat ions understanding based on sent iment analysis of twit ter data,” Procedia computer science, 111:376–381, 2017.CrossRef.

In this paper, we propose an approach to understand situations in the real world with the sentiment analysis of Twitter data base on deep learning techniques. With the proposed method, it is possible to predict user satisfaction of a product, happiness with some particular environment or destroy situation after disasters. Recently, deep learning is able to solve problems in computer vision or voice recognition, and convolutional neural network (CNN) works good for image analysis and image classification. The biggest reason to adopt CNN in image analysis and classification is due to CNN can extract an area of features from global information, and it is able to consider the relationship among these features. The above solution can achieve a higher accuracy in analysis and classification. For natural language processing, texts data features also can be extracted piece by piece and to consider the relationship among these features, but without the consideration of context or whole sentence, the sentiment might be understood wrong. And currently, convolutional neural network is one of the most effective methods

to do image classification, CNN has a convolutional layer to extract information by a larger piece of text, so we work for sentiment analysis with convolutional neural network, and we design a simple convolutional neural network model and test it on benchmark, the result shows that it achieves better accuracy performance in twitter sentiment classification than some of traditional method such as the SVM and Naive Bayes methods. In this paper, we propose an approach to understand situations in the real world with the sentiment analysis of Twitter data base on deep learning techniques. With the proposed method, it is possible to predict user satisfaction of a product, happiness with some particular environment or destroy situation after disasters. Recently, deep learning is able to solve problems in computer vision or voice recognition, and convolutional neural network (CNN) works good for image analysis and image classification. The biggest reason to adopt CNN in image analysis and classification is due to CNN can extract an area of features from global information, and it is able to consider the relationship among these features. The above solution can achieve a higher accuracy in analysis and classification. For natural language processing, texts data features also can be extracted piece by piece and to consider the relationship among these features, but without the consideration of context or whole sentence, the sentiment might be understood wrong. And currently, convolutional neural network is one of the most effective methods to do image classification, CNN has a convolutional layer to extract information by a larger piece of text, so we work for sentiment analysis with convolutional neural network, and we design a simple convolutional neural network model and test it on benchmark, the result shows that it achieves better accuracy performance in twitter sentiment classification than some of traditional method such as the SVM and Naive Bayes methods.

Shiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng,“ CNN for situat ions understanding based on sent iment analysis oftwit ter data,” Procedia computer science, 111:376–381, 2017.CrossRef.

In this paper, we propose an approach to understand situations in the real world with the sentiment analysis of Twitter data base on deep learning techniques. With the proposed method, it is possible to predict user

satisfaction of a product, happiness with some particular environment or destroy situation after disasters. Recently, deep learning is able to solve problems in computer vision or voice recognition, and convolutional neural network (CNN) works good for image analysis and image classification. The biggest reason to adopt CNN in image analysis and classification is due to CNN can extract an area of features from global information, and it is able to consider the relationship among these features. The above solution can achieve a higher accuracy in analysis and classification. For natural language processing, texts data features also can be extracted piece by piece and to consider the relationship among these features, but without the consideration of context or whole sentence, the sentiment might be understood wrong. And currently, convolutional neural network is one of the most effective methods to do image classification, CNN has a convolutional layer to extract information by a larger piece of text, so we work for sentiment analysis with convolutional neural network, and we design a simple convolutional neural network model and test it on benchmark, the result shows that it achieves better accuracy performance in twitter sentiment classification than some of traditional method such as the SVM and Naive Bayes methods.

Social media has become a source of varied kind of information., and the new type information could be harvested from social media. As one of the most popular social media, Twitter has at least 100 million active users, furthermore, 572,000 new accounts has been created on a single day (March 12, 2011, the day after the Sendai earthquake and resulting nuclear disaster), while an average of 140 million tweets are sent daily14. Valuable knowledge is often hidden behind Twitter contents and cannot be easily processed through automation12. Twitter is an ideal social media for the extraction of general public opinion on specific issues7 . Twitter data is useful for sentiment analysis, such as opinion mining or natural language processing10. There are several approaches for sentiment analysis on Twitter, one of them

is machine learning. Deep learning models have achieved great results in computer vision6 and speech recognition3 in recent years. To solve NLP (Natural Language Processing) problems, machine learning is also useful by using a general learning algorithm combined with a large sample of data to learn the classification rules. Several methods do it with traditional algorithm such as SVM or Naïve Bayes, most of such methods consider text word by word, classify a sentence to positive or negative by analyzing the word in the text, sometimes information lost by extracting key word without other word; CNN (Convolutional Neural Networks) is one of machine learning models which has archived impressive result in image recognition several years ago and has archived remarkable results in natural language processing recently, there is a convolutional layer to make a piece of words can be considered together. In this paper, we propose an approach to parsing Twitter data to understand situation in the real world based on a CNN model to do the sentiment analysis. We adopt convolutional neural network as our sentiment analysis model because in image analysis and classification field, CNN can extract an area of features from global information, with the convolution operation, a piece of data information can be extract together as the features, and it is able to consider the relationship among these features. For computer vision, such as image analysis, it is able to extract a part of pixel data information, not only extract the pixels one by one, the features information can be extracted piece by piece, the piece contains multi pixels data information; when we transfer the text into matrix, it can also be considered as same as an image pixels' matrix, so we can do the same operation to the text data to make the input features to the model can be trained in another effective way. The paper is organized as follow. In section 2 we introduce the approach we propose, and the key model we have choose in this approach; section 3 describe the experiment data, experiment method and experiment results, also include discussion about the experiments and the results; section 4 is the conclusion of this paper and future plan. 2. Approach The structure of this approach is shown in Figure 1. There is a convolutional neural network on the right side of the figure, which will be introduced in detail at the next sub section of this paper. Our approach is based on machine learning approach. A sentiment can be simply categorized into two groups13, so we choose MR1 and STS Gold

Dataset16 which has been labelled into two groups, positive and negative, as training dataset9 . Here the MR is a set of movie reviews with one sentence per review, the reviews are from Internet users and are similar to Twitter data; the STS Gold Dataset is an essential collection of real Twitter dataset. After training of the CNN with dataset, we input Twitter data we have gotten by hashtag and stored in MongoDB. The convolutional neural network will output the sentiment. With the sentiment and the geo-tag or other information included in Twitter, we can do further research.

PidikitiSupriya, MyneediPavani, NagarapuSaisushma,NamburiVimalaKumari, k Vikash,"Loan Predict ion by using Machine LearningModels", International Journal of Engineering and Techniques.Volume 5Issue 2, Mar-Apr 2019

With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this project we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this project is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing. In this paper we are predict the loan data by using some machine learning algorithms they are classification, logic regression, Decision Tree and gradient boosting. Keywords: Machine learning, Decision Tree, prediction, Python. I. INTRODUCTION This Problem is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this paper is to predict whether assigning the loan to a particular person will be safe or not. We have implemented this loan prediction

problem using Decision tree algorithm and data cleaning in Python as there are missing values in the dataset. We use map function for the missing values. The aim of this paper is to apply machine learning technique on dataset which has 1000 cases and 7 numerical and 6 categorical attributes. The creditability of a customer for sanctioning loan depend on several parameters, such as credit history, Installment etc. [2]. 2 LITERATURE REVIEW Data mining is the process of analyzing data from different perspectives and extracting useful knowledge from it[3]. It is the core of knowledge discovery process. The various steps involved in extracting knowledge from raw data as depicted in figure-1. Different data mining techniques include classification, clustering, association rule mining, prediction and sequential patterns, neural networks, regression etc. Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to classification technique. This approach frequently employs Decision tree based classification Algorithm. In classification, a training set is used to build the model as the classifier which can classify the data items into its appropriate classes. A test set is used to validate the model. RESEARCH ARTICLE OPEN ACCESS International Journal of Engineering and Techniques - Volume 5 Issue 2, Mar-Apr 2019 ISSN: 2395-1303 http://www.ijetjournal.org Page 145 Fig.1: Steps in knowledge extraction 2.1 Data Mining in Banking Due to tremendous growth in data the banking industry deals with, analysis and transformation of the data into useful knowledge has become a task beyond human ability. Data mining techniques can be adopted in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the data bases. By using data mining techniques to analyze patterns and trends, bank executives can predict, with increased accuracy, how customers will react to adjustments in interest rates, which customers are likely to accept new product offers, which customers will be at a higher risk for defaulting on a loan, and how to make customer relationships more profitable[5]. Globalization and the stiff competition had led the banks focus towards customer retention and fraud prevention. To help

them for the same, data mining is used. By analyzing the past data, data mining can help banks to predict credible customers. Thus they can prevent frauds, they can also plan for launching different special offers to retain those customers who are credible. Certain areas that effectively utilize data mining in banking industry are marketing, risk management and customer relationship management. Marketing:It is one of the most widely used areas of data mining in the banking industry. The consumer behavior with reference to product, price and distribution channel can be analyzed by the marketing department. The reaction of the customers to the existing and new products can also be known. This information can be used by the banks to promote the products, improve quality of products and services, and gain competitive advantages. Bank analysts can also analyze the past trends, determine the present demands and forecast the customer behavior of various products and services, in order to grab more business opportunities Risk Management: It is widely used for managing risks in the banking industry. Bank executives need to know the credibility of customers they are dealing with. Offering new customers credit cards, extending existing customers' lines of credit, and approving loans can be risky decisions for banks, if they do not know anything about their customers. Banks provide loans to their customers by verifying the various details relating to the loan, such as amount of loan, lending rate, repayment period etc. Even though, banks are cautious while providing loan, there are chances of loan repaying defaults by customers. Data mining technique helps to distinguish borrowers who repay loans promptly from those who default. Customer Relationship Management: Data mining can be useful in all the three phases of a customer relationship cycle such as customer acquisition, increasing value of the customer and customer retention. Customer acquisition and retention are very important concerns of any industry, especially the banking industry. Banks have to cater the needs of the customers by providing the services they prefer. This will ultimately lead to customer loyalty and customer retention. Data mining techniques help to analyze the customers who are loyal from those who shift to other banks for better services. If the customer is shifting from his bank to another, reasons for such shifting and the last transaction performed before shifting can be known, and this will help the banks to perform better and

retain their customers. 3. Proposed Model 3.1 Machine learning: Decision Tree Decision tree algorithm in machine learning methods which efficiently performs both classification and regression tasks[2]. It creates decision trees. Decision trees are widely used in the banking International Journal of Engineering and Techniques - Volume 5 Issue 2, Mar-Apr 2019 ISSN: 2395-1303 http://www.ijetjournal.org Page 146 industry due to their high accuracy and ability to formulate a statistical model in plain language. In Decision tree each node represents a feature (attribute), each link (branch) represents a decision (rule) and each leaf represents an outcome (categorical or continues value).

X.FrencisJensy, V.P.Sumathi,Janani Shiva Shri, "An exploratory DataAnalysis for Loan Predict ion based on nature of client s", InternationalJournal of Recent Technology and Engineering (IJRTE),Volume-7Issue-4S, November 2018.

In India, the number of people applying for the loans gets increased for various reasons in recent years. The bank employees are not able to analyse or predict whether the customer can payback the amount or not (good customer or bad customer) for the given interest rate. The aim of this paper is to find the nature of the client applying for the personal loan. An exploratory data analysis technique is used to deal with this problem. The result of the analysis shows that short term loans are preferred by majority of the clients and the clients majorly apply loans for debt consolidation. The results are shown in graphs that helps the bankers to understand the client's behaviour. Keywords - Loan analysis, exploratory data analysis technique, client's analysis, financial categories analysis The term banking can be defined as receiving and protecting money that is deposited by the individual or the entities. This also includes lending money to the people which will be repaid within the given time. Banking sector is regulated in most of the countries as it is the important factor in determining the financial stability of the country. The provision of banking regulation act allows public to obtain loans.Loans are good sum of money borrowed for a period and expected to be paid back at given interest rate. The purpose of the loan can be anything based on the customer requirements. Loans are broadly divided as openended and close-ended loans. Open-ended loans are the loans for which the client has

approval for a specific amount. Examples of open-end loans are credit cards and a home equity line of credit (HELOC). Close-ended loans decreases with each payment. In other words, it is a legal term that cannot be modified by the borrower. Personal loans, mortgages, auto payments, instalment loan and student loans are the most common examples of close-ended loans. Secured or collateral loan are those loans that are protected by an asset. Houses, Vehicles, Savings accounts are the personal properties used to secure the loan. Unsecured loans are also known as personal or signature loans. Here the lender believes that the borrower can repay the loan based on financial resources possessed by the borrower. Liquidity risk is the risk that arises from the lackof marketability of an investment that cannot be bought or sold quickly enough to prevent or minimize a loss. The interest rate risk is the risk in which the interest rates priced on loans will be too low to earn the bank money. Revised Version Manuscript Received on 25 November, 2018. Ms.X.FrancisJency, CSE Department, Kumaraguru College of Technology, Coimbatore, India Ms.V.P.Sumathi, CSE Department, Kumaraguru College of Technology, Coimbatore, India Janani Shiva Sri,C S Department, Kumaraguru College of Technology, Coimbatore, India The primary objective of the bank is to provide their wealth in the safer hands. In recent times, banks approve loan after verifying and validating the documents provided by the customer. Yet there is no guarantee whether the applicant is deserving or not. This paper classifies the customers based on certain criteria. The classification is done using Exploratory Data Analysis. Exploratory Data Analysis (EDA) is an approach to analyse the datasets that summarizes the main characteristics with visual methods. The purpose of using EDA is to uncover the underlying structure of a relatively larger set of variables using visualizing techniques. In [1] the researchers analyse the data set using data mining technique. Data mining procedure provides a great vision in loan prediction systems, since this will promptly distinguish the customers who are able to repay the loan amount within a period. Algorithms like “J48 algorithm”, “Bayes net”, Naive Bayes” are used. On applying these algorithms to the datasets, it was shown that “J48 algorithm” has high accuracy (correct percent) of 78.3784% which provides the banker to decide whether the loan can be given to the costumer or not. In paper [2], “loan

prediction using Ensemble technique", used "Tree model", "Random forest", "svm model" and combined the above three models as Ensemble model. A prototype has been discussed in paper [2] so that the banking sectors can agree/reject the loan request from their customers. The main method used is real coded genetic algorithms. The combined algorithms from the ensemble model, loan prediction can be done in an easier way. It is found that tree algorithm provides high accuracy of 81.25%. In paper [3], using R-language, an improved risk prediction clustering algorithm is used to find the bad loan customers since probability of default (PD) is the critical step for the customers who comes for a bank loan. So, a frame work for finding PD in the data set is provided by data mining technique. R- Language has the technique called as KNN (K-nearest neighbour) algorithm and it is used for performing multiple imputation calculation when there are missing values seen in the data set. The paper [4] had used tree model. It helps to find whether the banking sector people will be able to overcome the loan problem with their customers. It provides a high accuracy of 80.87%. An Exploratory Data Analysis for Loan Prediction Based on Nature of the Clients 177 Published By: Blue Eyes Intelligence Engineering Retrieval Number: E1864017519 & Sciences Publication The paper [5] uses decision tree induction algorithm and found that the algorithm finds a best way to evaluate the credit risk. To avoid the credit risk, bankers holds the technique called as "credit score", where it helps the lenders to keep note on who are the applicants who will able to repay the amount or probability of going into the default risks. The input given for credit evaluation was customer data, WEKA software, cibil score. The methodology used in prediction system was problem and data understanding, data filtering, system modelling and finally system evaluation. This was done on the banks existing dataset containing 1140 records and 24 attributes. At last the system was tested and helps the bankers to make a correct decision on whether to accept or reject the loan approval. The paper [6] used predictive model technique and descriptive model technique to predict the loan approval in banks. In predictive model technique, classification and regression were used and in descriptive model technique clustering and association were used. Classifiers also implement several algorithms like naive Bayes, kNN algorithms of R

language and regressors implements several algorithms like decision trees, neural networks, etc., To undergo this prediction analysis, out of all these algorithms, naive Bayes produces a most accurate classifier and the algorithms like decision tree, neural network, K-NN algorithms will be more accurate regressors. The main goal of the paper is to predict the loan classification based on the type of loan, loan applicant and the assets (property) that loan applicant holds. It was found that the decision tree algorithm gave an improved accuracy of almost 85% on doing the analysis. Whenever the bank makes decision to give loan to any customers then it automatically exposes itself to several financial risks. It is necessary for the bank to be aware of the clients applying for the loan. This problem motivates to do an EDA on the given dataset and thus analysing the nature of the customer. The dataset that uses EDA undergoes the process of normalisation, missing value treatment, choosing essential columns using filtering, deriving new columns, identifying the target variables and visualising the data in the graphical format. Python is used for easy and efficient processing of data. This paper used the pandas library available in Python to process and extract information from the given dataset. The processed data is converted into appropriate graphs for better visualisation of the results and for better understanding. For obtaining the graph Matplot library is used. A. Annual Income Vs Purpose Of Loan In this Figure 1, the X axis represents the purpose of loan i.e. the purpose for which the loan is applied. Debt consolidation, home improvement is some of the purposes. High, moderate and low represents the annual income of people

3. PROBLEM ANALYSIS

3.1 EXISTINGAPPROACH:

Existing system uses machine learning algorithms like decision tree and random forest. Although we can get good efficiency using the Random Forest classifier and Decision Tree classifier, we can get great results by using Naïve Bayes classifier.The lender must manually review each application, supported the main points provided by the applicant like gender, legal status, education, number of dependents, income, loan amount, credit history, etc., and judge if they're creditworthy or not but it gives less accurate result.

3.11Drawbacks

1.less efficiency

3.2 Proposed System

The proposed system automates the method of determining the applicant's loan approvel status. It concentrates on a single or multiple datasets containing the details of the loan applicants. In the preposed system, Naïve Bayes Classification Machine Learning model is been used.Naïve Bayes algorithm is one in all the supervised learning algorithms, which is dependent on Bayes theorem which is applicable for solving classification problems.It is good for foretelling the right result in the current real-world case scenarios and also help the finance agency to offer the finances in the right hands and also help the applicants in obtaining loan in a much faster way.The key advantage of this model is, it gives more accuracy when compared to the existing system

3.3 Software And Hardware Requirements

SOFTWARE REQUIREMENTS

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.

The appropriation of requirements and implementation constraints gives the general overview of the project in regards to what the areas of strength and deficit are and how to tackle them.

- **Python idel 3.7 version (or)**
- **Anaconda 3.7 (or)**
- **Jupiter (or)**
- **Google colab**

HARDWARE REQUIREMENTS

Minimum hardware requirements are very dependent on the particular software being developed by a given Enthought Python / Canopy / VS Code user. Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

- **Operating system : windows, linux**
- **Processor : minimum intel i3**
- **Ram : minimum 4 gb**
- **Hard disk : minimum 250gb**

3.4 About Dataset

For loan prediction we use customers details. This dataset consists of 13 columns and 614 records. Out of 13 columns twelve columns are customer features and one column represents labels. Where Labels are classified into two categories, these are yes or no(1 or 0).

Load_ID, Gender, Married, Dependents, Education, self_Employee, applicanIncome, coapplicantIncome, loanAmount, loan_Amount_Term, Credit_History, property_area, Loan_Status

LP001002, Male, No, 0, Graduate, No, 5849, 0, 128, 360, 1, Urban, Y

LP001003,Female, yes, 1, Graduate, NO, 4583, 1508, 128, 360, 1, Rural, N

Above all bold values are column names and thin values are dataset values and last column contains either yes or No(0 or 1) .yes means the loan is approved and no means loan is rejected.

4. SYSTEMDESIGN

UML DIAGRAMS

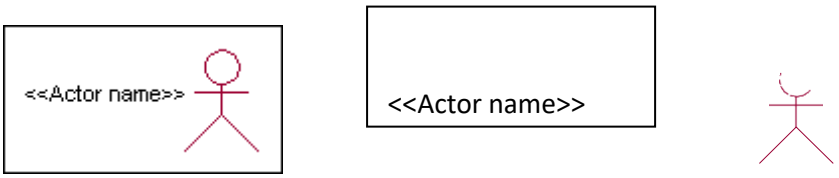
The System Design Document describes the system requirements, operating environment, system and subsystem architecture, files and database design, input formats, output layouts, human-machine interfaces, detailed design, processing logic, and externalinterfaces.

Global Use Case Diagrams:

Identification of actors:

Actor: Actor represents the role a user plays with respect to the system. An actor interacts with, but has no control over the use cases.

Graphical representation:



Actor

An actor is someone or something that:

Interacts with or uses the system.

Provides input to and receives information from thesystem.

Is external to the system and has no control over the use cases. Actors are discovered by examining:

- Who directly uses thesystem?
- Who is responsible for maintaining thesystem?
- External hardware used by thesystem.

- Other systems that need to interact with the system. Questions to identifyactors:
  - Who is using the system? Or, who is affected by the system? Or, which groups need help from the system to perform atask?

- Who affects the system? Or, which user groups are needed by the system to perform its functions? These functions can be both main functions and secondary functions such asadministration.
- Which external hardware or systems (if any) use the system to perform tasks?
- What problems does this application solve (that is, forwhom)?
- And, finally, how do users use the system (use case)? What are they doing with thesystem?

The actors identified in this system are:

a. System Administrator

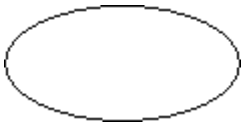
b. Customer

c. CustomerCare

Identification of usecases:

Usecase: A use case can be described as a specific way of using the system from a user's (actor's)perspective.

Graphical representation:



A more detailed description might characterize a use case as:

- Pattern of behavior the systemexhibits
- A sequence of related transactions performed by an actor and thesystem
- Delivering something of value to the actor Use cases provide a meansto:
- capture systemrequirements
- communicate with the end users and domainexperts
- test thesystem

Use cases are best discovered by examining the actors and defining what the actor will be able to do with thesystem.

Guide lines for identifying use cases:

- For each actor, find the tasks and functions that the actor should be able to perform or that the system needs the actor to perform. The use case should represent a course of events that leads to clear goal
- Name the usecases.
- Describe the use cases briefly by applying terms with which the user isfamiliar. This makes the description lessambiguous

Questions to identify use cases:

- What are the tasks of eachactor?
- Will any actor create, store, change, remove or read information in thesystem?
- What use case will store, change, remove or read thisinformation?
- Will any actor need to inform the system about sudden externalchanges?
- Does any actor need to inform about certain occurrences in thesystem?
- What usecases will support and maintains thesystem?

Flow of Events

A flow of events is a sequence of transactions (or events) performed by the system. They typically contain very detailed information, written in terms of what the system should do, not how the system accomplishes the task. Flow of events are created as separate files or documents in your favorite text editor and then attached or linked to a use case using the Files tab of a model element.

A flow of events should include:

- When and how the use case starts andends
- Use case/actorinteractions
- Data needed by the usecase
- Normal sequence of events for the usecase
- Alternate or exceptional flows Construction of Usecasediagrams:

Use-case diagrams graphically depict system behavior (use cases). These diagrams present a high level view of how the system is used as viewed from an outsider's (actor's) perspective. A use-case diagram may depict all or some of the use cases of asystem.

A use-case diagram can contain:

- actors ("things" outside thesystem)
- use cases (system boundaries identifying what the system should do)
- Interactions or relationships between actors and use cases in the system including the associations, dependencies, andgeneralizations.

Relationships in use cases:

1. Communication:

The communication relationship of an actor in a usecase is shown by connecting the actor symbol to the usecase symbol with a solid path. The actor is said to communicate with the usecase.

2. Uses:

A Uses relationship between the usecases is shown by generalization arrow from the usecase.

3. Extends:

The extend relationship is used when we have one usecase that is similar to another usecase but does a bit more. In essence it is like subclass.

SEQUENCE DIAGRAMS

A sequence diagram is a graphical view of a scenario that shows object interaction in a time- based sequence what happens first, what happens next. Sequence diagrams establish the roles of objects and help provide essential information to determine class responsibilities and interfaces. There are two main differences between sequence and collaboration diagrams: sequence diagrams show time-based object interaction while collaboration diagrams show how objects associate with each other. A sequence diagram has two dimensions: typically, vertical placement represents time and horizontal placement represents differentobjects.

Object:

An object has state, behavior, and identity. The structure and behavior of similar objects are defined in their common class. Each object in a diagram indicates some instance of a class. An object that is not named is referred to as a class instance.

The object icon is similar to a class icon except that the name is underlined:

An object's concurrency is defined by the concurrency of its class.

Message:

A message is the communication carried between two objects that trigger an event. A message carries information from the source focus of control to the destination focus of control. The synchronization of a message can be modified through the message specification. Synchronization means a message where the sending object pauses to wait forresults.

Link:

A link should exist between two objects, including class utilities, only if there is a relationship between their corresponding classes. The existence of a relationship between two classes symbolizes a path of communication between instances of the classes: one object may send messages to another. The link is depicted as a straight line between objects or objects and class instances in a collaboration diagram. If an object links to itself, use the loop version of the icon.

CLASS DIAGRAM:

Identification of analysis classes:

A class is a set of objects that share a common structure and common behavior (the same attributes, operations, relationships and semantics). A class is an abstraction of real-world items. There are 4 approaches for identifying classes:

a. Noun phrase approach:

b. Common class patternapproach.

c. Use case Driven Sequence or Collaborationapproach.

d. Classes , Responsibilities and collaboratorsApproach

1. Noun PhraseApproach:

The guidelines for identifying the classes:

- Look for nouns and noun phrases in the usecases.
- Some classes are implicit or taken from generalknowledge.
- All classes must make sense in the application domain; Avoid computer implementation classes – defer them to the designstage.
- Carefully choose and define the class names After identifying the classes we have to eliminate the following types ofclasses:
- Adjectiveclasses.

2. Common class patternapproach:

The following are the patterns for finding the candidate classes:

- Conceptclass.
- Eventsclass.
- Organizationclass
- Peoplesclass
- Placesclass
- Tangible things and devicesclass.

3. Use case drivenapproach:

We have to draw the sequence diagram or collaboration diagram. If there is need for some classes to represent some functionality then add new classes which perform those functionalities.

4. CRCapproach:

The process consists of the following steps:

- Identify classes' responsibilities (and identify the classes)
- Assign theresponsibilities
- Identify the collaborators. Identification of responsibilities of

eachclass:

The questions that should be answered to identify the attributes and methods of a class respectively are:

a. What information about an object should we keep trackof?

b. What services must a class provide? Identification of relationships among theclasses:

Three types of relationships among the objects are:

Association: How objects are associated?

Super-sub structure: How are objects organized into super classes and sub classes? Aggregation: What is the composition of the complex classes? Association:

The **questions** that will help us to identify the associations are:

a. Is the class capable of fulfilling the required task byitself?

b. If not, what does itneed?

c. From what other classes can it acquire what itneeds? Guidelines for identifying the tentativeassociations:

- A dependency between two or more classes may be an association. Association often corresponds to a verb or prepositionalphrase.
- A reference from one class to another is an association. Some associations are implicit or taken from generalknowledge.

Some common association patterns are:

Location association like part of, next to, contained in….. Communication association like talk to, order to ……

We have to eliminate the unnecessary association like implementation associations, ternary or n- ary associations and derived associations.

Super-sub class relationships:

Super-sub class hierarchy is a relationship between classes where one class is the parent class of another class (derived class).This is based on inheritance. Guidelines for identifying the super-sub relationship, a generalization are

1. **Top-down*:***

Look for noun phrases composed of various adjectives in a class name. Avoid excessive refinement. Specialize only when the sub classes have significant behavior.

2. **Bottom-up*:***

Look for classes with similar attributes or methods. Group them by moving the common attributes and methods to an abstract class. You may have to alter the definitions a bit.

3. **Reusability*:***

Move the attributes and methods as high as possible in the hierarchy.

4. **Multipleinheritances*:***

Avoid excessive use of multiple inheritances. One way of getting benefits of multiple inheritances is to inherit from the most appropriate class and add an object of another class as an attribute.

Aggregation or a-part-of relationship:

It represents the situation where a class consists of several component classes. A class that is composed of other classes doesn't behave like its parts. It behaves very difficultly. The major properties of this relationship are transitivity and antisymmetry.

The **questions** whose answers will determine the distinction between the part and whole relationships are:

- Does the part class belong to the problemdomain?
- Is the part class within the system'sresponsibilities?

- Does the part class capture more than a single value?(If not then simply include it as an attribute of the wholeclass)
- Does it provide a useful abstraction in dealing with the problemdomain? There are three types of aggregation relationships. Theyare:

Assembly:

It is constructed from its parts and an assembly-part situation physically exists.

Container:

A physical whole encompasses but is not constructed from physical parts.

Collection member:

A conceptual whole encompasses parts that may be physical or conceptual. The container and collection are represented by hollow diamonds but composition is represented by solid diamond.

USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

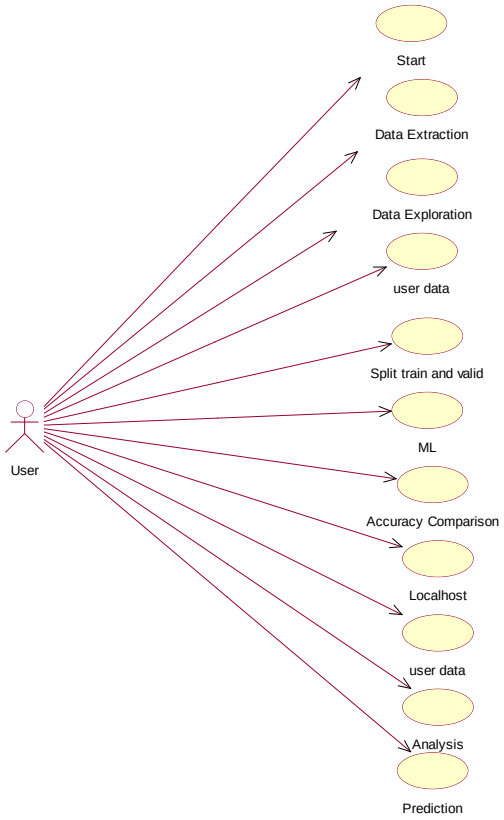


Fig 1: Use Case Diagram

CLASS DIAGRAM

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that

describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

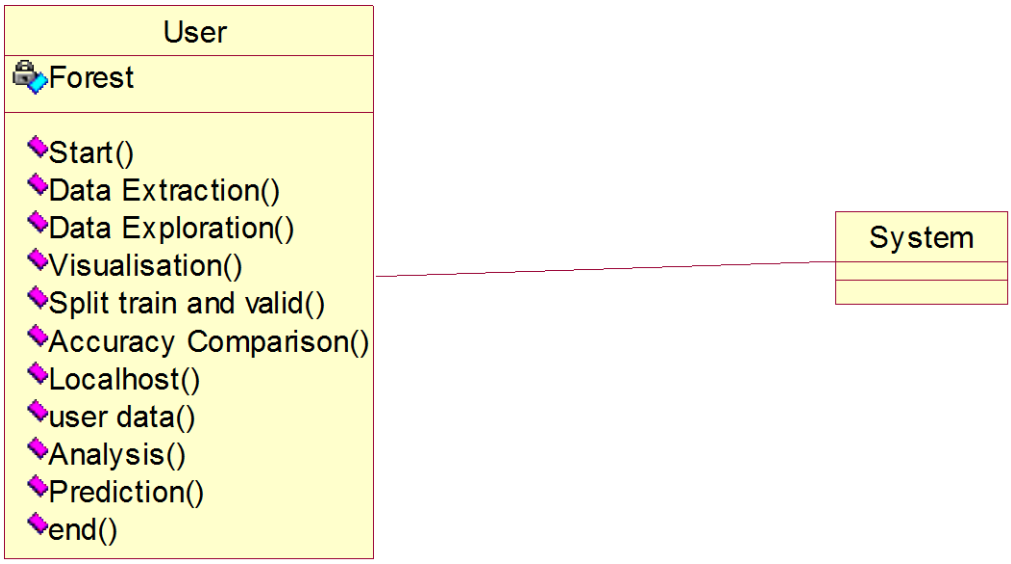


Fig 2:Class Diagram

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. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

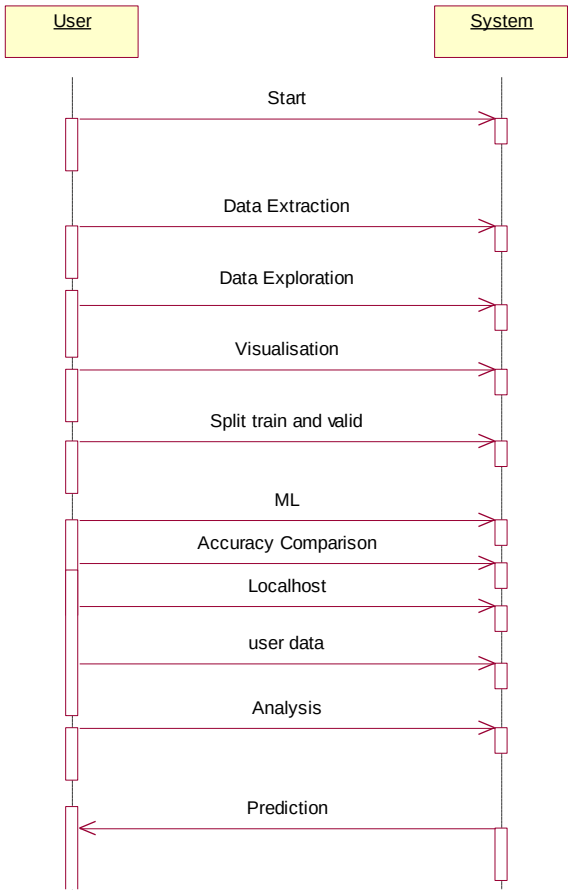
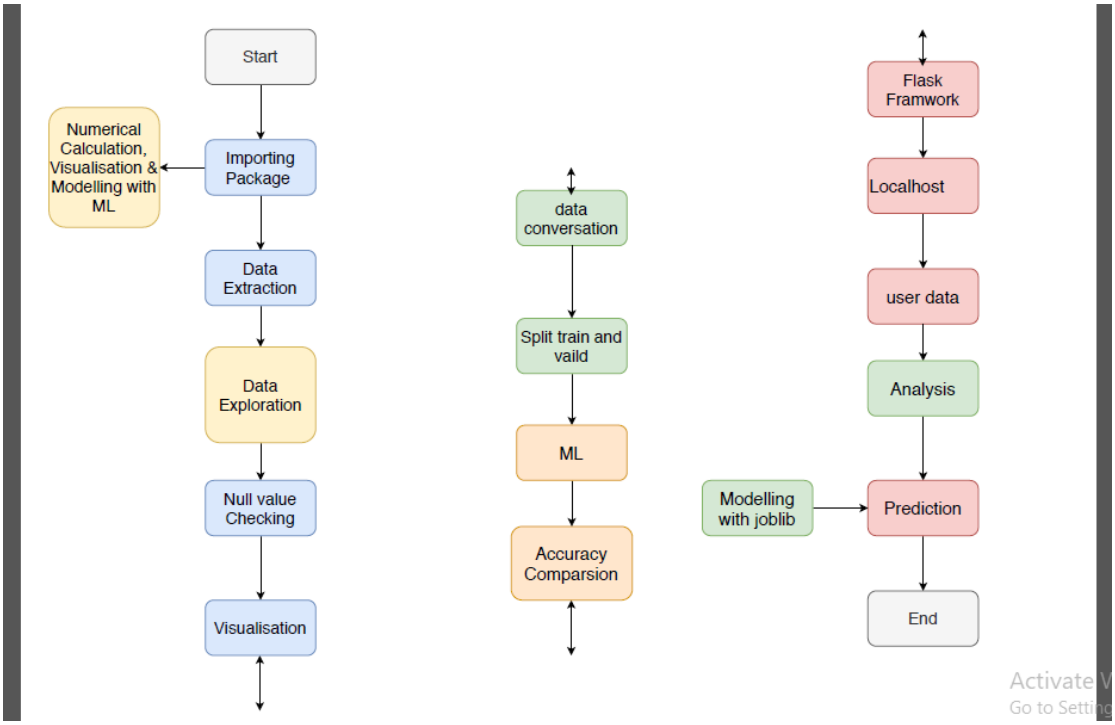


Fig 3: Sequence Diagram

5.IMPLEMENTATION

5.1 FLOW CHART:



5.2 Code

6.TESTING

6.1 SOFTWARETESTING

Testing

Testingisaprocessofexecutingaprogramwiththeaimoffindingerror.Tomakeour software perform well it should be error free. If testing is done successfully it will remove all the errors from thesoftware.

6.1.1 Types ofTesting

1. White BoxTesting
2. Black BoxTesting
3. Unit testing
4. IntegrationTesting
5. AlphaTesting
6. BetaTesting
7. Performance Testing and so on

White BoxTesting

Testing technique based on knowledge of the internal logic of an

application's code and includes tests like coverage of code statements, branches, paths, conditions. It is performedby softwaredevelopers

Black BoxTesting

A method of software testing that verifies the functionality of an application without having specificknowledgeoftheapplication'scode/internalstructure.Testsarebasedonrequirements andfunctionality.

Unit Testing

Software verification and validation method in which a programmer tests if individual unitsof source code are fit for use. It is usually conducted by the developmentteam.

IntegrationTesting

The phase in software testing in which individual software modules are combined and tested as a group. It is usually conducted by testing teams.

Alpha Testing

Type of testing a software product or system conducted at the developer's site. Usually it is performed by the end users.

BetaTesting

Final testing before releasing application for commercial purpose. It is typically done by end- users or others.

PerformanceTesting

Functional testing conducted to evaluate the compliance of a system or component with specified performance requirements. It is usually conducted by the performance engineer.

Black BoxTesting

Blackbox testing is testing the functionality of an application without knowing the details of itsimplementationincludinginternalprogramstructure,datastructuresetc.Testcasesforblack box testing are created based on the requirement specifications. Therefore, it is also called as specification-based testing. Fig.4.1 represents the black box testing:

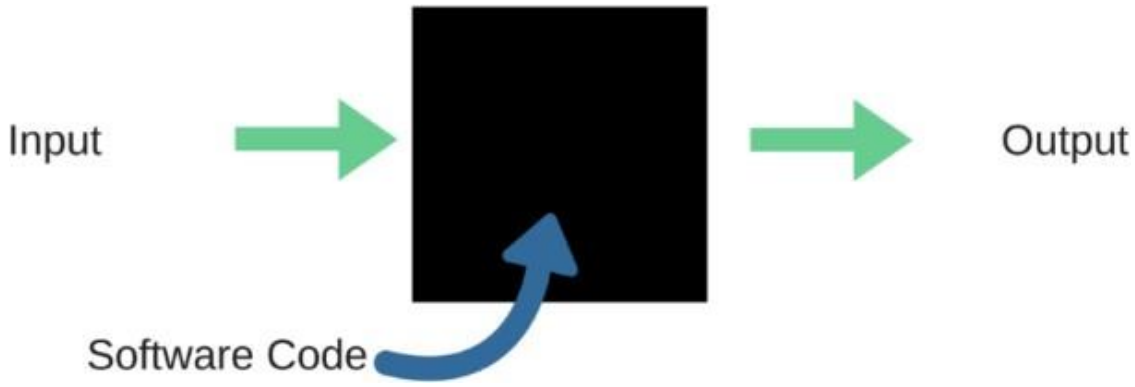


Fig.: Black Box Testing

When applied to machine learning models, black box testing would mean testing machine learning models without knowing the internal details such as features of the machine learning

model, the algorithm used to create the model etc. The challenge, however, is to verify the test outcome against the expected values that are known beforehand.

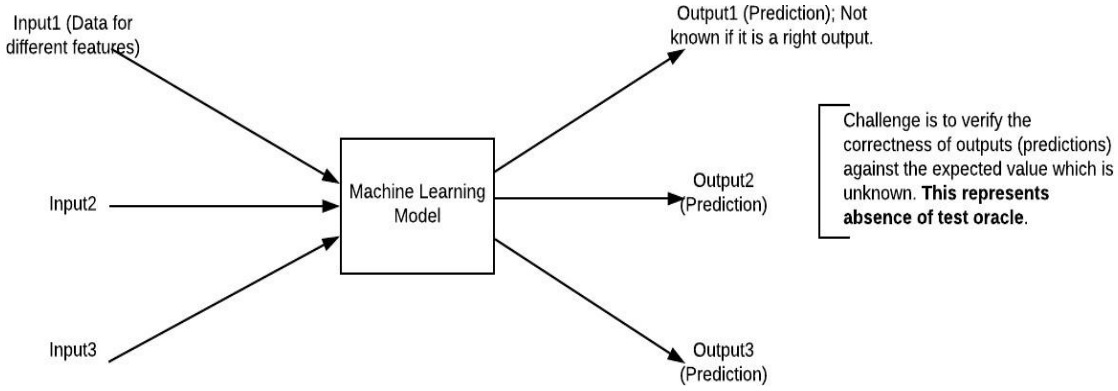


Fig.: Black Box Testing for Machine Learning algorithms

The above Fig.4.2 represents the black box testing procedure for machine learning algorithms.

Table.4.1: Black box Testing

Input	**Actual Output**	**Predicted Output**

[16,6,324,0,0,0,22,0,0,0,0,0,0]	0	0
[16,7,263,7,0,2,700,9,10,1153,832,9,2]	1	1

The model gives out the correct output when different inputs are given which are mentioned in Table 4.1. Therefore the program is said to be executed as expected or correct program

Testing

Testingisaprocessofexecutingaprogramwiththeaimoffindingerror.Tomakeour software perform well it should be error free. If testing is done successfully it will remove all the errors from thesoftware.

7.2.2 Types ofTesting

1. White BoxTesting
2. Black BoxTesting
3. Unit testing
4. IntegrationTesting
5. AlphaTesting
6. BetaTesting
7. Performance Testing and so on

White BoxTesting

Testing technique based on knowledge of the internal logic of an application's code and includes tests like coverage of code statements, branches, paths, conditions. It is performedby softwaredevelopers

Black BoxTesting

A method of software testing that verifies the functionality of an application without having specificknowledgeoftheapplication'scode/internalstructure.Testsarebasedonrequirements andfunctionality.

Unit Testing

Software verification and validation method in which a programmer tests if individual unitsof source code are fit for use. It is usually conducted by the developmentteam.

IntegrationTesting

The phase in software testing in which individual software modules are combined and tested as a group. It is usually conducted by testing teams.

Alpha Testing

Type of testing a software product or system conducted at the developer's site. Usually it is performed by the end users.

BetaTesting

Final testing before releasing application for commercial purpose. It is typically done by end- users or others.

PerformanceTesting

Functional testing conducted to evaluate the compliance of a system or component with specified performance requirements. It is usually conducted

by the performance engineer.

Black BoxTesting

Blackbox testing is testing the functionality of an application without knowing the details of itsimplementationincludinginternalprogramstructure,datastructuresetc.Testcasesforblack box testing are created based on the requirement specifications. Therefore, it is also called as specification-based testing. Fig.4.1 represents the black box testing:

Fig.: Black Box Testing

When applied to machine learning models, black box testing would mean testing machine learning models without knowing the internal details such as features of the machine learning

model, the algorithm used to create the model etc. The challenge, however, is to verify the test outcome against the expected values that are known beforehand.

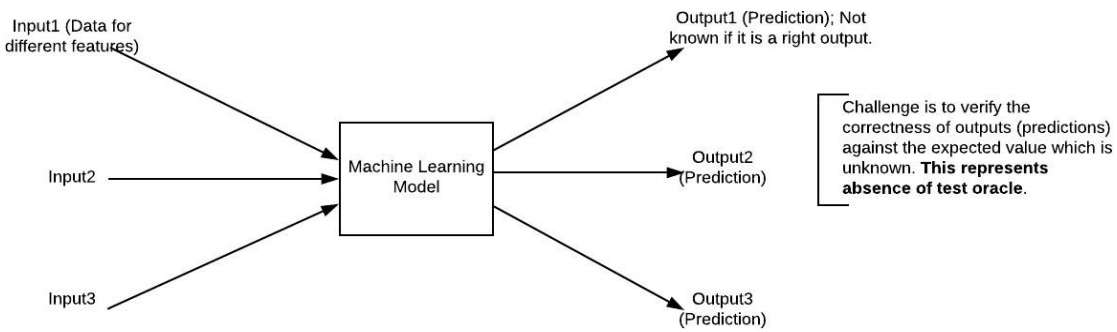


Fig.: Black Box Testing for Machine Learning algorithms

The above Fig.4.2 represents the black box testing procedure for machine learning algorithms.

Table.4.1: Black box Testing

Input	**Actual Output**	**Predicted Output**
[16,6,324,0,0,0,22,0,0,0,0,0,0]	0	0
[16,7,263,7,0,2,700,9,10,1153,832,9,2]	1	1

The model gives out the correct output when different inputs are given which are mentioned in Table 4.1. Therefore the program is said to be executed as expected or correct program

Test Case Id	Test Case Name	Test Case Description	Test Steps			Test Case Status	Test Priority
			Step	Expected	Actual		
01	Start the Application	Host the application and test if it starts making sure	If it doesn't Start	We cannot run the application.	The application hosts success.	High	High

		the required software is available					
02	Home Page	Check the deployment environment for properly loading the application.	If it doesn't load.	We cannot access the application.	The application is running successfully.	High	High
03	User Mode	Verify the working of the application in freestyle mode	If it doesn't Respond	We cannot use the Freestyle mode.	The application displays the Freestyle Page	High	High
04	Data Input	Verify if the application takes input and updates	If it fails to take the input or store in The Database	We cannot proceed further	The application updates the input to application	High	High

An Approach for Prediction of Loan Approval using Machine Learning Algorithm

7.RESULTS ANDDISCUSSIONS

Add screenshots

8.CONCLUSION

Therefore, the developed model automates the method of determining the applicant's creditworthiness. It focuses on an information containing the main points of the loan applicants. In this system Naïve Bayes Classification model is used. In Machine Learning, Naïve Bayes classification analysis is one of the supervised learning algorithm, which is dependent on Bayes theorem and used to solve classification problems. Hence, it is good for predicting the right result in the current world scenario and also help the bank to give the money in the right hands and also help the people in getting loan in a much faster way. The main advantage of this system is, it gives more accuracy.

8.BIBILOGRAPHY

[1] Toby Segaran, "Programming Collective Intelligence: Building Smart Web 2.0 Applications." O'Reilly Media.

[2] Drew Conway and John Myles White," Machine Learning for Hackers: Case Studies and Algorithms to Get you Started," O'Reilly Media.

[3] Trevor Hastie, Robert Tibshirani, and Jerome Friedman,"The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer ,Kindle

[4] PhilHyoJin Do ,Ho-Jin Choi, "Sentiment analysis of real-life situations using loca- tion, people and time as contextual features," International Conference on Big Data and Smart Computing (BIGCOMP), pp. 39–42. IEEE, 2015.

[5] Bing Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, May 2012.

[6] Bing Liu, "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions," Cambridge University Press, ISBN:978-1-107-01789-4.

[7] Shiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng, "CNN for situations understanding based on sentiment analysis of twitter data," Procedia computer science, 111:376–381, 2017.CrossRef.

[8] K I Rahmani, M.A. Ansari, Amit Kumar Goel, "An Efficient Indexing Algorithm for CBIR,"IEEE- International Conference on Computational Intelligence & Communication Technology ,13-14 Feb 2015.

[9] Gurlove Singh, Amit Kumar Goel ,"Face Detection and Recognition System using Digital Image Processing" , 2nd International conference on Innovative Mechanism for Industry Application ICMIA 2020, 5-7 March 2020, IEEE Publisher.

[10] Amit Kumar Goel, KalpanaBatra, Poonam Phogat," Manage big data using optical networks", Journal of Statistics and Management Systems "Volume 23, 2020, Issue 2, Taylors & Francis.