A

Major

Project On

PREDICTION OF LOAN USING MACHINE LEARNING ALGORITHM

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

BY

Rachana Arelli

(177R1A0504)

Nikhitha Maduri

(177R1A0536)

Rachana pallikonda

(177R1A0543)

Under the Guidance of

M.CHALAPATHI RAO

(Associate Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CMR TECHNICAL CAMPUS
UGC AUTONOMOUS
(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi)
Recognized Under Section 2(f) & 12(B) of the UGC Act.1956,
Kandlakoya (V), Medchal Road, Hyderabad-501401.
2017-2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled "**PREDICTION OF LOAN USING MACHINE LEARNING ALGORITHM**" being submitted by **Rachana Arelli, Nikhitha Maduri, Rachana Pallikonda** bearing the **177R1A0504, 177R1A0536, 177R1A0543** roll number in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by him/her under our guidance and supervision during the year 2020-21.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Mr.M.Chalapathi Rao   **Dr. A. Raji Reddy**
Associate Professor   **DIRECTOR**
INTERNAL GUIDE

Dr. K. Srujan Raju
HoD

                                                                        EXTERNAL EXAMINER

Submitted for viva voice Examination held on_____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. We take this opportunity to express our profound gratitude and deep regard to our guide

Mr.M..Chalapathi rao , Associate Professor for her exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by her shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to Project Review Committee (PRC) Coordinators: **Mr. J. Narasimha Rao, Mr. B. P. Deepak Kumar,   Mr. K. Murali, Dr. Suwarna Gothane and Mr. B. Ramji** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to the Head of the Department **Dr. K. Srujan Raju** for providing excellent infrastructure and a nice atmosphere for completing this project successfully. We are obliged to our Director **Dr. A. Raji Reddy** for being cooperative throughout the course of this project. We would like to express our sincere gratitude to our Chairman Sri. **Ch. Gopal Reddy** for his encouragement throughout the course of this project The guidance and support received from all the members of **CMR TECHNICAL CAMPUS** who contributed and who are contributing to this project, was vital for the success of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement without which this assignment would not be possible. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project.

<div align="right">

Rachana   Arelli(177R1A0504)

Nikhitha Maduri(177R1A036)

Rachana Pallikonda(177R1A0543)

</div>

ABSTRACT

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 data of the previous records of the people to whom the loan was granted before. So the idea of this project is to gather loan data from multiple data sources and train machine learning algorithms on this data and to deploy this model we use flask framework which redirects it to a web browser that predicts loan approval/rejection. This model can be used by the organizations in making the right decision to approve or reject the applicants request.

LIST OF FIGURES

LIST OF SCREENSHOTS

TABLE OF CONTENT

1.INTRODUCTION

1.INTRODUCTION

Machine learning is an application of artificial intelligence (AI) that pro-videos systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The primary aim is to allow the computers to learn automatically without human intervention or assistance and adjust actions accordingly.

Machine learning in conjunction with big data can not only collect information but also find specific patterns. The main portion of the bank's assets comes directly from the profit earned from the loans distributed by the banks. The prime objective in the banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approve loans after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique. The aim of this system is to provide a quick, immediate and easy way to choose the deserving applicants.

This system is exclusively for the managing authority of a Bank/finance company, the whole process of prediction is done privately and no stakeholders would be able to alter the processing. Results against particular Loan Id can be sent to various departments of banks so that they can take appropriate action on application. This helps all other departments to carry out other formalities.

1.1 PROJECT PURPOSE

Here the purpose of this project is to train and test the data to Predict whether assigning the loan to a particular person or not. In this project we are going to predict the loan data by using some machine learning algorithms. They are classification, logic regression, Decision Tree Support vector machine and Naive Bayes.

1.2 PROJECT SCOPE

The scope of this paper is to implement and investigate how different supervised binary classification methods impact default prediction.This is done by mining the data of the previous records of the people to whom the loan was granted before. So the idea of this project is to gather loan data from multiple data sources and train machine learning algorithms on this data and to deploy this model we use flask framework which redirects it to a web browser that predicts loan approval or rejection.

1.3 PROJECT FEATURES

Data mining is the process of analyzing data from different perspectives and extracting useful knowledge from it. It is the core of knowledge discovery process.The various steps involved in extracting knowledge from the raw data.Different data mining techniques which it include classification,clustering,association the rule of mining,prediction and sequential patterns,neural networks,regression etc.Classification is the most commonly applied data mining technique which employs a set of classified examples to develop a model that can classify the population 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 set set is used to validate the model. 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 data of the previous records of the people to whom the loan was granted before. So the idea of this project is to gather loan data from multiple data sources and train machine learning algorithms on this data and to deploy this model we use flask framework which redirects it to a web browser that predicts loan approval/rejection. This model can be used by the organizations in making the right decision to approve or reject the applicant's request.

2.SYSTEM ANALYSIS

2. SYSTEM ANALYSIS

SYSTEM ANALYSIS

System Analysis is the important phase in the system development process. The System is studied to the minute details and analyzed. The system analyst plays an important role of an interrogator and dwells deep into the working of the present system. In analysis, a detailed study of these operations performed by the system and their relationships within and outside the system is done. A key question considered here is, "what must be done to solve the problem?" The system is viewed as a whole and the inputs to the system are identified. Once analysis is completed the analyst has a firm understanding of what is to be done.

Requirement Specification is the part of the project which gives the details about the hardware and software requirements of the project. It also details the features of the programming language used. In order to implement this project, the choice of processor with maximum possible speed is made. There should be sufficient memory to store data and software tools for efficient processing.

2.1 PROBLEM DEFINITION

A detailed study of the process of loan prediction using various machine learning algorithms.The data can be collected from various sources.First we have to take data set from bank which has various details about the accountant such as such as applicant personal details, income, co-applicant income, credit score ,urban, semi urban and rural areas etc.. Customers first apply for a loan after that company (bank) validates the customer eligibility for loan. However doing this manually takes a lot of time. Hence it wants to automate the loan eligibility process (real time) based on customer information.

So the final thing is to identify the factors/ customer segments that are eligible for taking loans. How will the company(bank) benefit if we give the customer segments is the immediate question that arises. The solution is ….Banks would give loans to only those customers that are eligible so that they can be assured of getting the money back.

Hence the more accurate we are in predicting the eligible customers by training the data with machine learning algorithms and deploying this model we use the flask framework which redirects it to a web browser that predicts loan approval or rejection..

2.2 EXISTING SYSTEM

Assessing the risk, which is involved in a loan application, is one of the most important concerns of the banks for survival in the highly competitive market and for profitability. These banks receive a number of loan applications from their customers and other people on a daily basis. Not everyone gets approved. Most of the banks use their own credit scoring and risk assessment techniques in order to analyze the loan application and to make decisions on credit approval. In spite of this, there are many cases happening every year, where people do not repay the loan amounts or they default, due to which these financial institutions suffer huge amounts of losses.

2.2.1 LIMITATIONS OF EXISTING SYSTEM

The disadvantage is that it delays the process of Loan request for an applicant.In addition to this, it can also cause internal corruption among employees.

2.3 PROPOSED SYSTEM

The primary goal of this project is to extract patterns from a common loan approved data set, and then build a model based on these extracted patterns, in order to predict the likely loan defaulters by using classification data mining algorithms. The historical data of the customers like their age, income, loan amount, employment length etc. will be used in order to do the analysis. Later on, some analysis will also be done to find the most relevant attributes, i.e. The factors that affect the prediction result the most. Using different type of Machine Learning algorithm and Predicting accuracy result and Plotting a graph

2.3.1 ADVANTAGES OF THE PROPOSED SYSTEM

Using this system helps the employees to complete this process very quickly. This system helps in minimizing the internal corruption that happens when loans are approved manually by humans.

2.4 SYSTEM REQUIREMENTS SPECIFICATION

2.4.1 FUNCTIONAL REQUIREMENTS

- Input data to train the model.

- Data processing to clean the raw input data.

- Accurate algorithm for training model.

- Post-processing to manage the output.

2.4.2 NON-FUNCTIONAL REQUIREMENTS

- **Accuracy:**Accuracy is the major part as we require accurate results.

- **Reliability:**Reliability is needed so that model is not disturbed.

- **Adaptability**:The model should adapt automatically according to user input.

- **Security:**Security has to be provided to the data model in case of sensitive data.

- **Feasibility Study:**The model to be trained on data has to be feasible.

2.5 HARDWARE & SOFTWARE REQUIREMENTS

2.5.1 SOFTWARE REQUIREMENTS

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements,

- Operating system   :   Windows 8, 10
- Languages   :   Python
- Back end   :   Machine Learning
- IDE   :   Jupyter

2.5.2 HARDWARE REQUIREMENTS

Hardware interfaces specify the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements.

- Processor   :   Intel Dual Core@ CPU 2.90GHz.
- Hard disk   :   16GB and Above.
- RAM   :   4GB and Above.
- Monitor   :   5 inches or above.

3. ARCHITECTURE

3.ARCHITECTURE

3.1 PROJECT ARCHITECTURE

This project architecture shows the procedure for loan approval prediction using machine learning Algorithms, starting from input to final prediction.
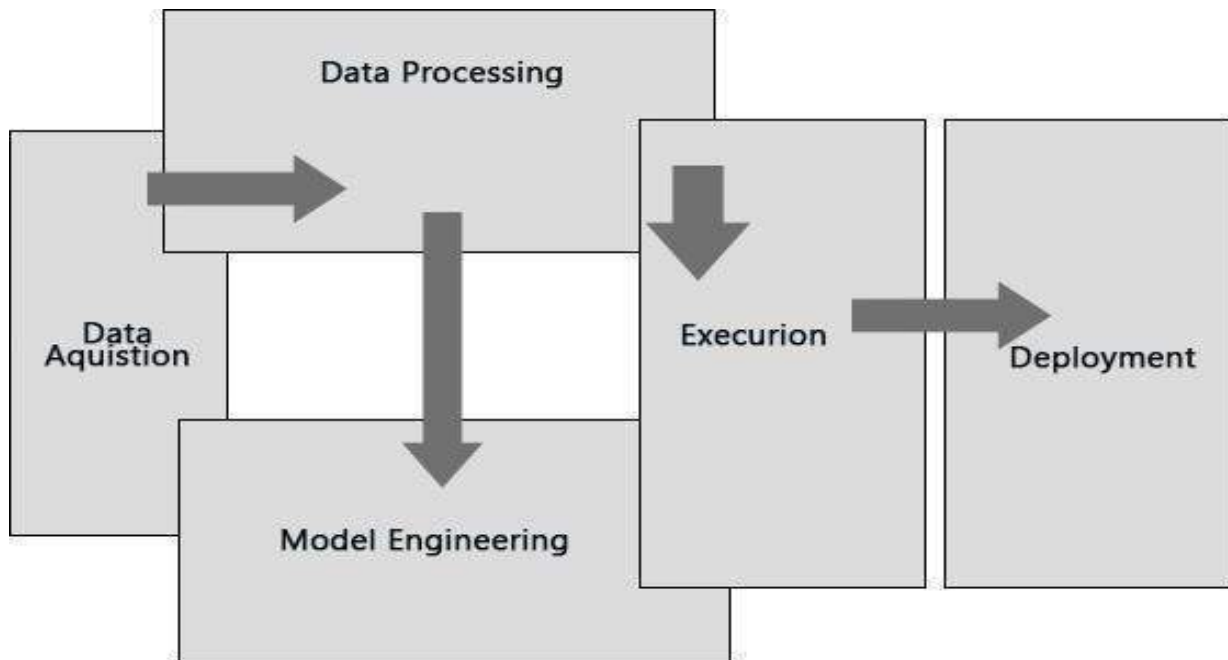
Figure 3.1: Project Architecture of Machine Learning process

3.2 DESCRIPTION

Data Acquisition

As machine learning is based on available data for the system to make a decision hence the first step defined in the architecture is data acquisition. This involves data collection, preparing and segregating the case scenarios based on certain features involved with the decision making cycle and forwarding the data to the processing unit for carrying out further categorization.

This stage is sometimes called the data processing stage. The data model expects reliable, fast and elastic data which may be discrete or continuous in nature. The data is then passed into stream processing systems (for continuous data) and stored in batch data warehouses (for discrete data) before being passed on to data modeling or processing stages.

Data Processing

The received data in the data acquisition layer is then sent forward to the data processing layer where it is subjected to advanced integration and processing and involves normalization of the data, data cleaning, transformation, and encoding. The data processing is also dependent on the type of learning being used. For e.g., if supervised learning is being used the data shall be segregated into multiple steps of sample data required for training of the system and the data thus created is called training sample data or simply training data. Also, the data processing is dependent upon the kind of processing required and may involve choices ranging from action upon continuous data which will involve the use of specific function based architecture, for example, lambda architecture, Also it might involve action upon discrete data which may require memory bound processing. The data processing layer defines if the memory processing shall be done to data in transit or in rest.

Data Modeling

This layer of the architecture involves the selection of different algorithms that might adapt the system to address the problem for which the learning is being devised. These algorithms are being evolved or being inherited from a set of libraries. The algorithms are used to model the data accordingly, this makes the system ready for the execution step.

Execution

This stage in machine learning is where the experimentation is done, testing is involved and turnings are performed. The general goal behind being to optimize the algorithm in order to extract the required machine outcome and maximize the system performance, The output of the step is a refined solution capable of providing the required data for the machine to make decisions.

Deployment

Like any other software output, ML outputs need to be operationalized or be forwarded for further exploratory processing. The output can be considered as a non-deterministic query which needs to be further deployed into the decision-making system. It is advised to seamlessly move the ML output directly to production where it will enable the machine to directly make decisions based on the output and reduce the dependency on the further exploratory steps.

3.3   MODULE DESCRIPTION

- Loan Dataset.
- Data Analysis.
- Training of Systems.
- Testing.
- Result as a graph and Accuracy.

3.3.1   LOAN DATA SET

Here we first take the data set from the bank which has various details of the accountant such as applicant personal details, income, co-applicant income, credit score etc..

3.3.2 DATA ANALYSIS

At the start, the data set was cleaned. Then exploratory data analysis and feature engineering were performed. Then a model was created which predicted whether the applicant would repay the loan or not. Whenever the bank makes a decision to give a 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 us to do an EDA on the given data set and thus analyzing the nature of the customer. The data set that uses EDA undergoes the process of normalization, missing value treatment, choosing essential columns using filtering, deriving new columns, identifying the target variables and visualizing the data in the graphical format. Python is used for easy and efficient processing of data. We used the pandas library available in Python to process and extract information from the given data set. The processed data is converted into appropriate graphs for better visualization of the results and for better understanding. For obtaining the graph Mat plot library is used.
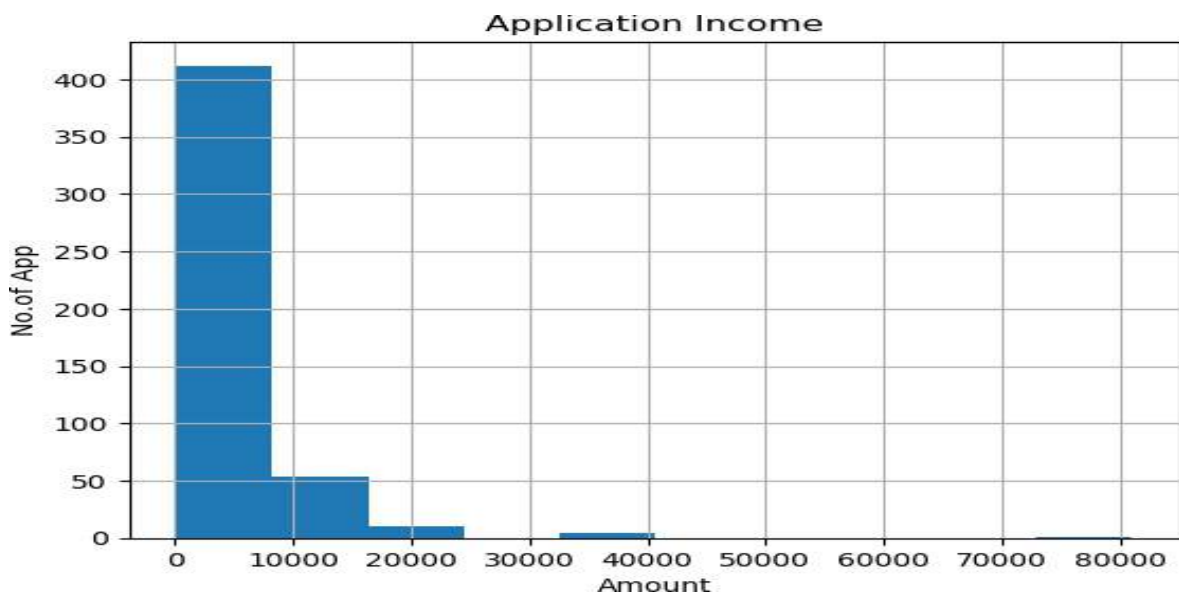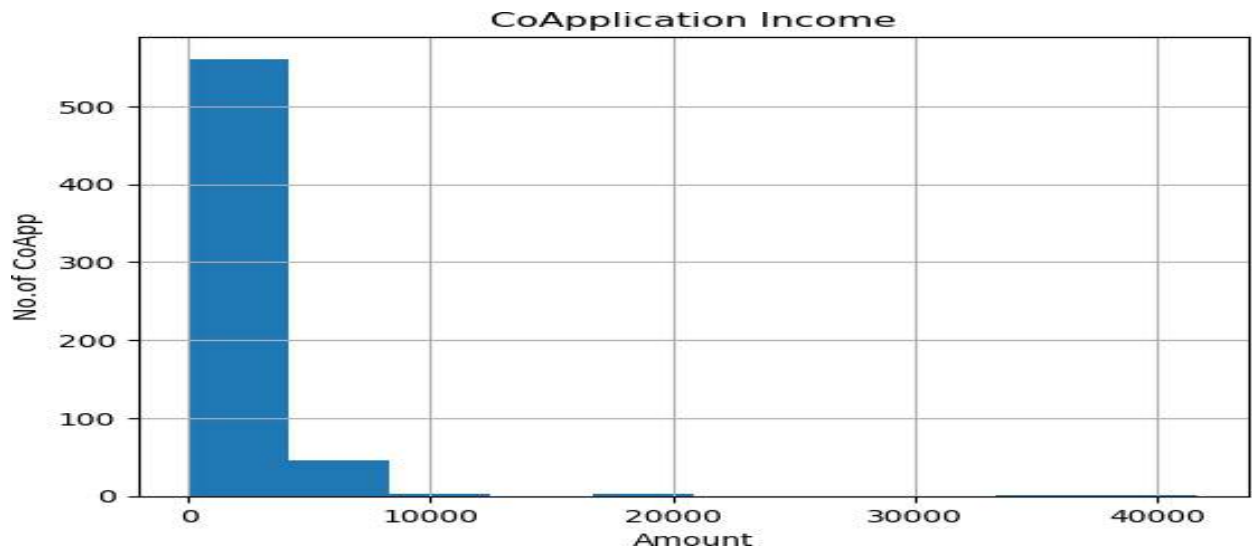
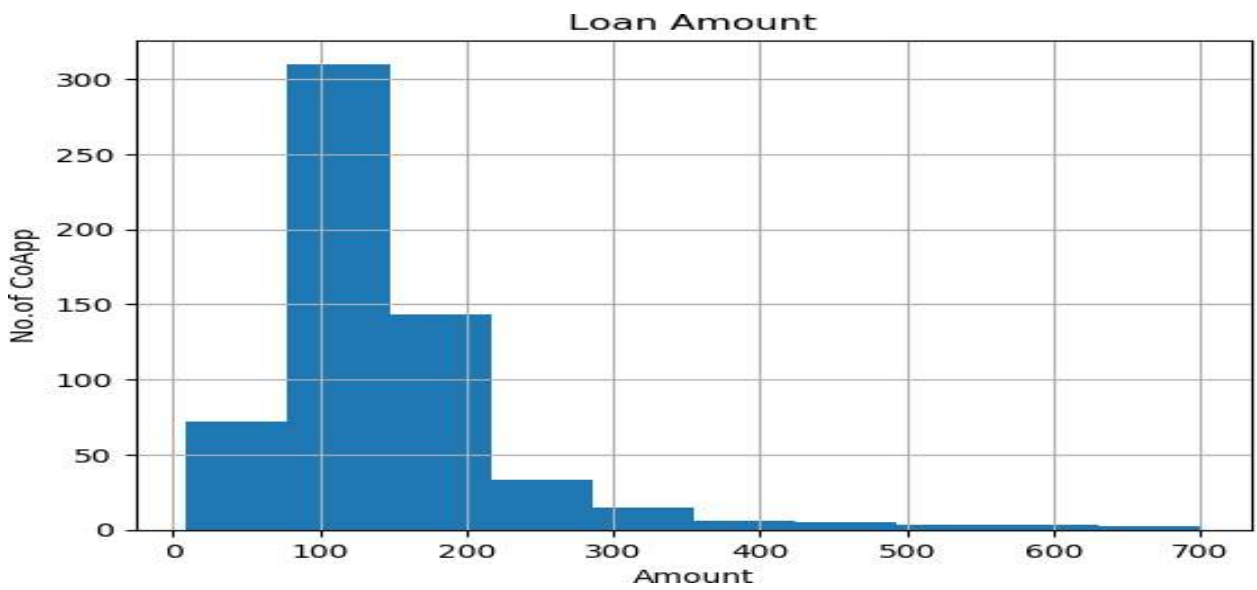Figure 3.4: Co-Applicant Income

Figure 3.5: Loan Amount

3.3.3   TRAINING OF THE SYSTEM

In supervised learning, a machine learning algorithm builds a model by examining many examples and attempting to find a model that minimizes loss; this process is called empirical risk minimization. Loss is the penalty for a bad prediction. That is, loss is a number indicating how bad the model's prediction was on a single example. If the model's prediction is perfect, the loss is zero; otherwise, the loss is greater. The goal of training a model is to find a set of weights and biases that have low loss, on average, across all examples. Here we have used four algorithms -

- Support Vector Machine.
- Logistic Regression.
- Decision Tree.
- Naive Bayes.

Training the Algorithams

1.SVM-Support Vector Machine

```
2]: from sklearn import svm
    from sklearn import metrics
```

```
3]: model1=svm.SVC()
```

```
4]: model1.fit(x_train,y_train)
```

Figure 3.6: Training machine using SVM

2.LogisticRegression

```
In [38]: from sklearn.linear_model import LogisticRegression
         model2=LogisticRegression()
```

```
In [39]: model2.fit(x_train,y_train)
```

C:\Users\ML\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver
'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ML\Anaconda3\lib\site-packages\sklearn\utils\validation.py:761: DataConversionWarning: A column-ve
en a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
  y = column_or_1d(y, warn=True)

```
Out[39]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
              intercept_scaling=1, max_iter=100, multi_class='warn',
              n_jobs=None, penalty='l2', random_state=None, solver='warn',
              tol=0.0001, verbose=0, warm_start=False)
```

Figure 3.7: Training machine using Logistic Regression

3.Decision Tree algorithm

```
In [42]: from sklearn.tree import DecisionTreeClassifier
```

```
In [43]: model3=DecisionTreeClassifier()
```

```
In [44]: model3.fit(x_train,y_train)
```

```
Out[44]: DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
              max_features=None, max_leaf_nodes=None,
              min_impurity_decrease=0.0, min_impurity_split=None,
              min_samples_leaf=1, min_samples_split=2,
              min_weight_fraction_leaf=0.0, presort=False, random_state=None,
              splitter='best')
```

Figure 3.8: Training machine using Decision Tree

Naive Bayes

```
In [47]: from sklearn.naive_bayes import GaussianNB

In [48]: model4=GaussianNB()

In [49]: model4.fit(x_train,y_train)

C:\Users\ML\Anaconda3\lib\site-packages\sklearn\utils\validation.py:761: DataConversionWarning: A column-vec
en a 1d array was expected. Please change the shape of y to (n samples, ), for example using ravel().
```

Figure 3.9: Training machine using Naive Bayes

3.3.4 TESTING

```
In [35]: xpredict=model1.predict(x_test)

In [36]: svm=metrics.accuracy_score(y_test,xpredict)
         print(svm)

         0.7291666666666666

In [37]: svm=model1.score(x_test,y_test)
         svm

Out[37]: 0.7291666666666666
```

```
In [45]:  xpredict=model3.predict(x_test)
```

```
In [46]:  dt=metrics.accuracy_score(y_test,xpredict)
          print(dt)

          0.7430555555555556
```

Figure 3.11: Testing machine using Logistic Regression algorithm

```
In [40]:  xpredict=model2.predict(x_test)
```

```
In [41]:  lr=metrics.accuracy_score(y_test,xpredict)
          print(lr)

          0.7986111111111112
```

Figure 3.12: Testing machine using Decision Tree Algorithm

```
In [50]:  xpredict=model4.predict(x_test)
```

```
In [51]:  nv=metrics.accuracy_score(y_test,xpredict)
          print(dt)

          0.7430555555555556
```

3.3.5   RESULT AS A GRAPH
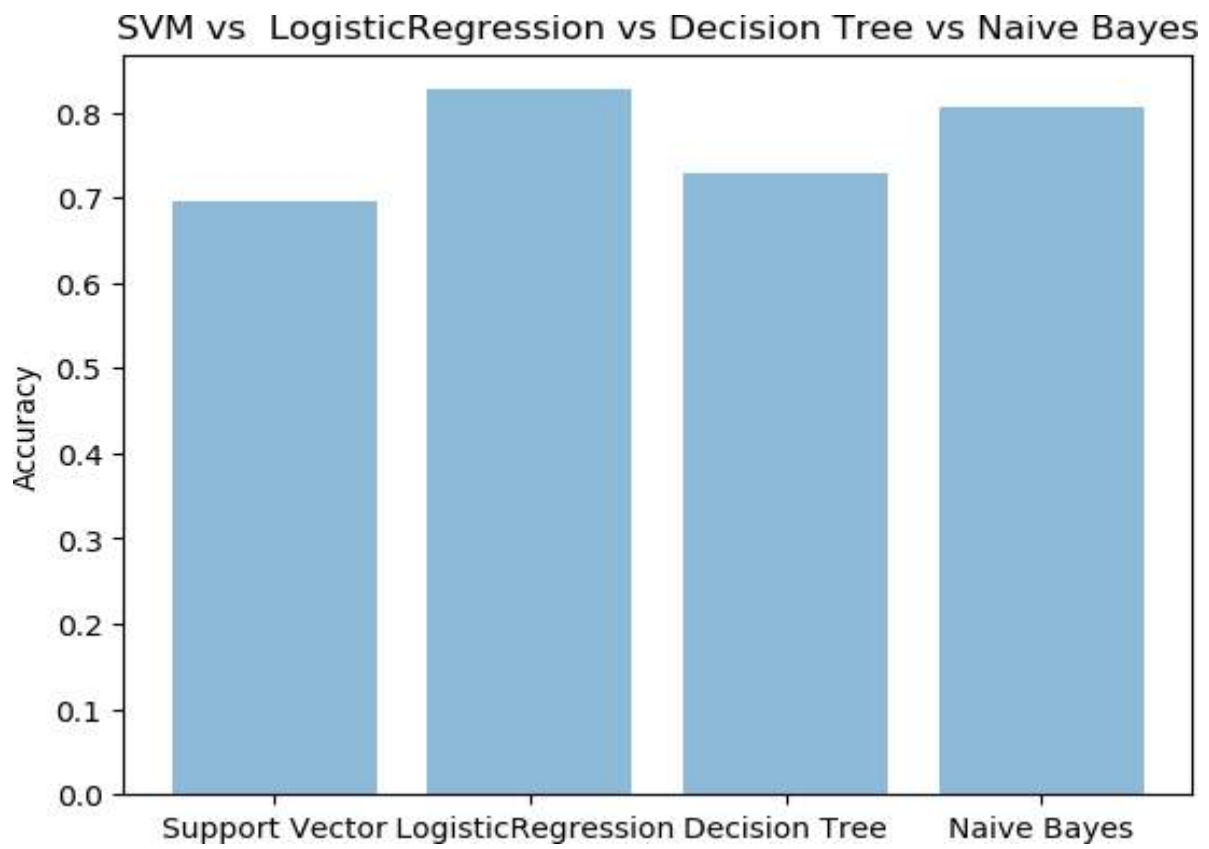
Figure 3.14: Predicting the Result

3.4 USE CASE DIAGRAM

Use case diagrams represent the functionality of the system from a user point of view. A Use case describes a function provided by the system that yields a visible result for an actor. An actor describes any entity that interacts with the system.

The actors in loan prediction are:

- User
- System
- Data set Collection
- Pre processing
- Features Extraction
- Classification
- Apply Algorithm in datasets
- Accuracy of result
- Visualization of graph model
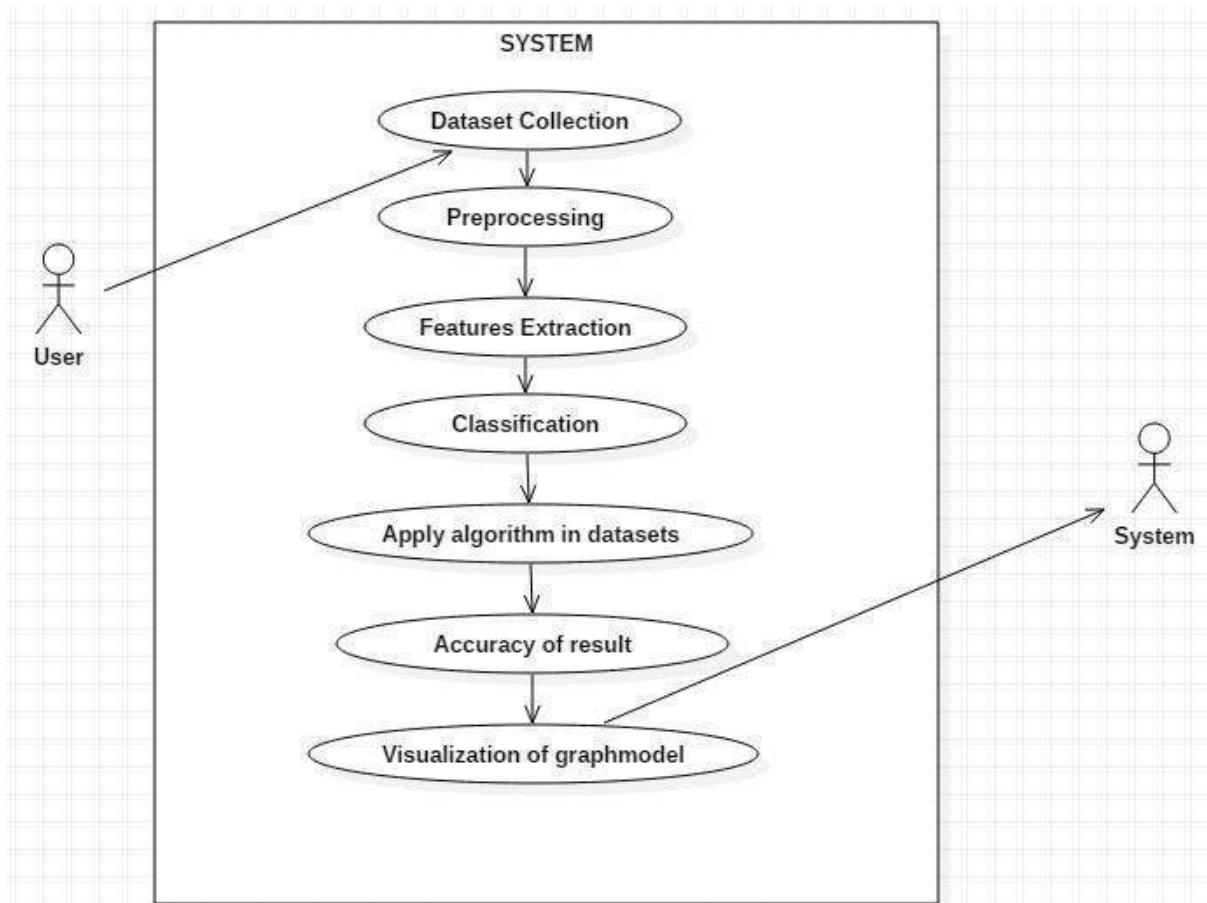
Figure 3.15: Use Case Diagram for user for Loan Prediction System.

3.5 CLASS DIAGRAM

Class Diagram is a collection of classes and objects.The class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing and documenting different aspects of a system but also for constructing executable code of the software application.The class diagram shows a collection of classes, interfaces, associations, collaborations and constraints.

Classes:

● User

● System

● Data set
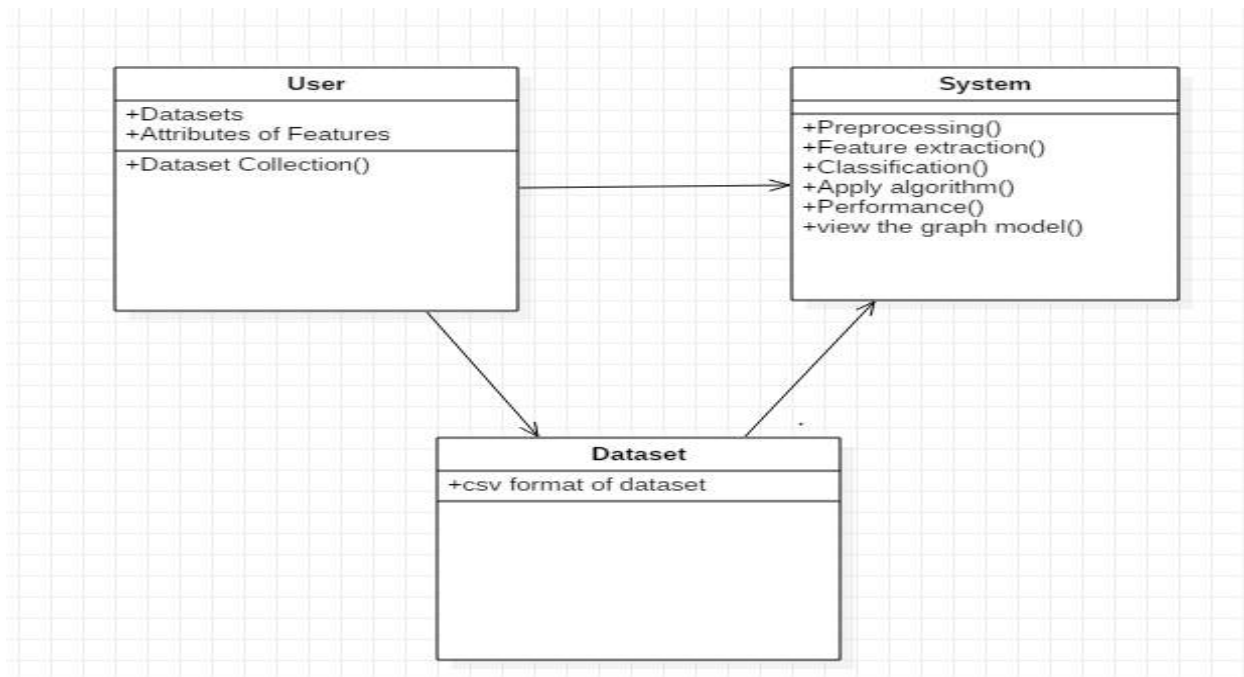
Figure 3.1.6: Class Diagram for Loan Prediction

3.6 SEQUENCE DIAGRAM

A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario.

The sequence of object interactions are:

- Data set collection
- CSV format of data set
- Pre processing
- Feature extraction
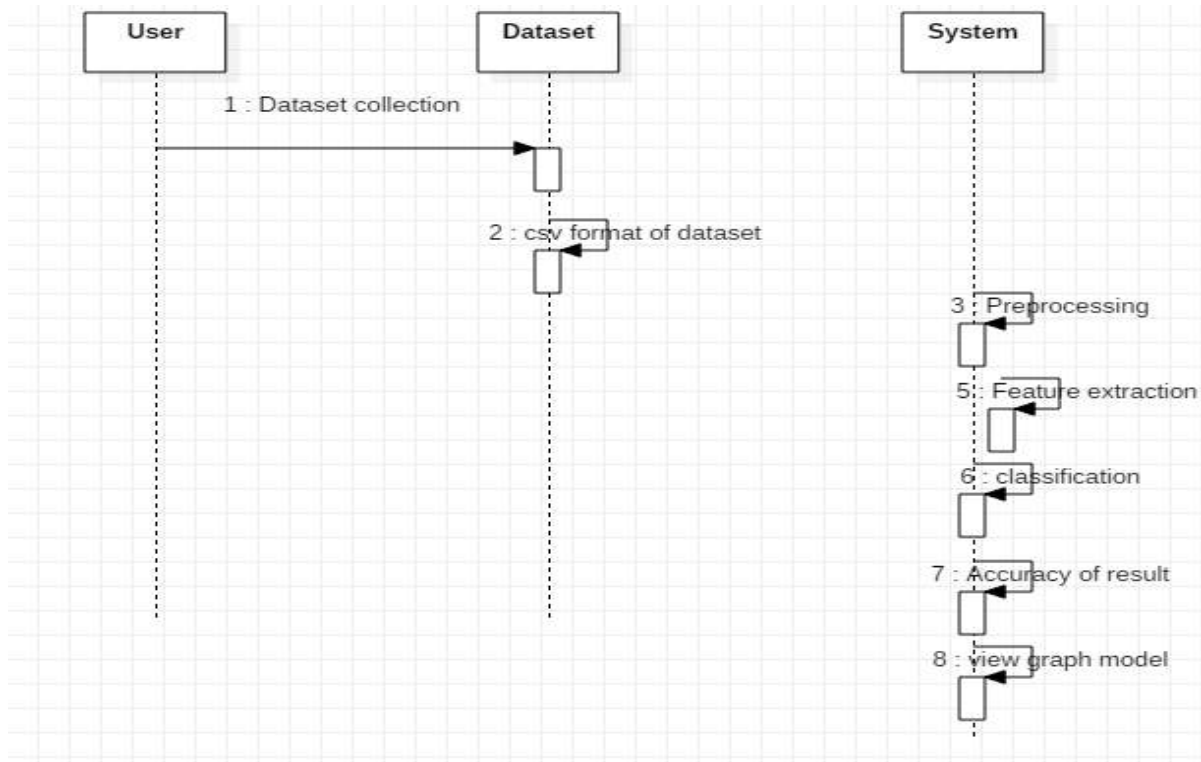- Classification
- Accuracy of result
- View graph model

Figure 3.1.7: Sequence Diagram for loan Prediction

3.7 ACTIVITY DIAGRAM

It describes the flow of activity states.A collaboration diagram is a type of visual presentation that shows how various software objects interact with each other within an overall IT architecture and 30 how users can benefit from this collaboration. A collaboration diagram often comes in the form of a visual chart that resembles a flow chart. The various objects in chat board are:

- User
- Data set collection
- Pre processing
- Trained data set
- Testing data set
- Feature extraction
- Classification
- Accuracy of result.
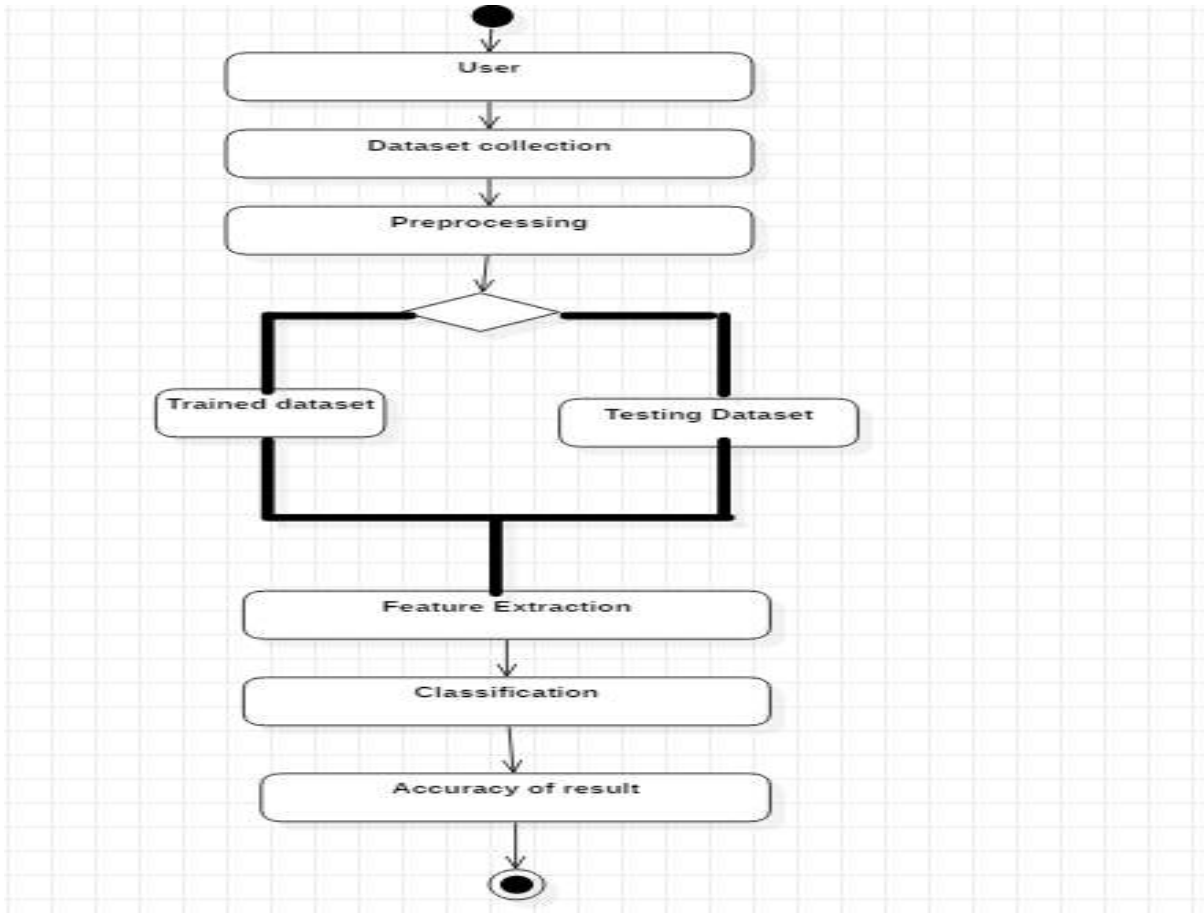
Figure 3.1.8: Activity Diagram for loan Prediction

4.IMPLEMENTATION

4. IMPLEMENTATION

SAMPLE CODE

```python
import pickle
from sklearn.linear_model import LogisticRegression


app=Flask(_name_)

@app.route('/')

def index1():
    return render_template('index1.html')

@app.route('/predict',methods=['POST'])

def predict():
    if request.method=='POST': Dependents=request.form['Dependents']
        ApplicantIncome=request.form['ApplicantIncome']
        CoapplicantIncome=request.form['CoapplicantIncome']
        LoanAmount=request.form['LoanAmount']
        Loan_Amount_Term=request.form['Loan_Amount_Term'
        ] Credit_History=request.form['Credit_History']
        Gender=request.form['Gender']
        Married=request.form['Married']
        Self_Employed=request.form['Self_Employed']
        Education=request.form['Education']
        Property_Area=request.form['Property_Area']


data=[[float(Dependents),float(ApplicantIncome),float(CoapplicantIncome),float(Loan
Amount),float(Loan_Amount_Term),float(Credit_History),float(Gender),float(Married),float
(Self_ Employed),float(Education),float(Property_Area)]]

    lr=pickle.load(open('project.pkl','rb'))
    prediction=lr.predict(data)[0]
  return render_template('index1.html',prediction=prediction)
if_name =='' main ':
    app.run()
```

5. SCREENSHOTS

5.1 LOAN PREDICATION FORM

Figure 5.1 Loan predication form

5.2 LOAN APPROVAL

Loan Prediction form

Figure 5.2 Loan Approval

5.3 LOAN REJECTED

Loan Prediction form

Basic Information

First Name Ram

Contact Number 12345

Applicant Personal Information

Dependents 2

Marrieds yes ▾

Gender Male ▾

Applicant Details

ApplicantIncome(in 100's) 3510

CoapplicantIncome(in 100's) 0

LoanAmount(in 1000's) 66

Loan_Amount_Term 360

Credit_History 0 ▾

Education Not Graduate ▾

Self_Employed No ▾

Property_Area Rural ▾

Predict

Status 0:Loan Rejected

Figure 5.3 Loan Rejected

6. TESTING

6. TESTING

6.1 INTRODUCTION TO TESTING

Testing is the process of finding differences between the expected behavior specified by system models and the observed behavior Testing is the process of executing a program with the intent of finding any errors. Testing is vital to the success of the system. Without proper testing hide errors will surface after sometime of use and perhaps irreversible damage has been done to valuable data. A series of tests like responsiveness, its value, stress and security are performed before the system is ready for user acceptance testing. System testing follows the logical conclusion that all parts of the system are tested and found to be working properly under all kinds of situations, and then the system is achieving its goal of processing the data perfectly according to user rules and requirements.

6.2   TESTING ACTIVITIES

Different levels of testing are used in the testing process, each level of testing aims to test different aspects of the system. The basic levels

- Unit testing
- Integration testing
- System testing
- Acceptance testing

6.2.1   UNIT TESTING

Unit testing focuses on the building blocks of the software system, that is, objects and sub system. There are three motivations behind focusing on components. First, unit testing reduces the complexity of the overall tests activities, allowing us to focus on smaller units of the system. Second, unit testing makes it easier to pinpoint and correct faults given that few components are involved in the test. Third, Unit testing allows parallelism in the testing activities in which each component can be tested.

6.2.2 INTEGRATION TESTING

In the integration testing, many test modules are combined into sub systems, which are then tested. The goal here is to see if the modules can be integrated properly, the emphasis being on testing module interaction. After structural testing and functional testing get error free modules . These modules are to be integrated to get the required results of the system. After checking a module, another module is tested and is integrated with the previous module. After the integration, the test cases are 54 generated and the results are tested.

6.2.3 SYSTEM TESTING

In system testing the entire software is tested. The reference document for the process is the requirement document and the goal is to see whether the software meets its requirements. The system was tested for various test cases with various inputs.

6.2.4 ACCEPTANCE TESTING

Acceptance testing is sometimes performed with realistic data of the client to demonstrate that the software is working satisfactory. Testing here focuses on the external behavior .

6.3 TYPES OF TESTING

- ❖ Black box or functional testing
- ❖ White box testing or structural testing

6.3.1 BLACK BOX TESTING

The method is used when knowledge of the specified function that a product has been designed to perform is known. The concept of black box is used to represent a system whose inside workings are not available to inspection. In a black box the test item is a "Black" , since its logic is unknown , all that is known is what goes in and what comes out or the input and output.

6.3.2 WHITE BOX TESTING

White box testing is concerned with testing the implementation of the program. The intent of structural is not to exercise all the inputs or outputs but to exercise the different programming and data structure used in the program. Thus structural testing aims to achieve test cases that will force the desired coverage of different structures. Two types of path testing are statement testing coverage and branch testing coverage.

Test Plan:

Testing process starts with a test plan. The plan identifies all the testing related activities that must be performed and specifies the schedules, allocates the resources, and specified guidelines for testing. During the testing of the unit the specified test cases are executed and the actual result compared with expected output. The final output of the testing phase is the test report and the error report.

Test Data:

Here all test cases that are used for the system testing are specified. The goal is to test the 55 different functional requirements specified in Software Requirements Specifications (SRS) document.

Test Report:

The module is working properly provided the user has to enter information. All data entry forms have been tested with specified test cases and all data entry forms are working properly.

6.4 TEST CASES

We have used four algorithms and trained it so that we can choose the best algorithm for Loan approval prediction.
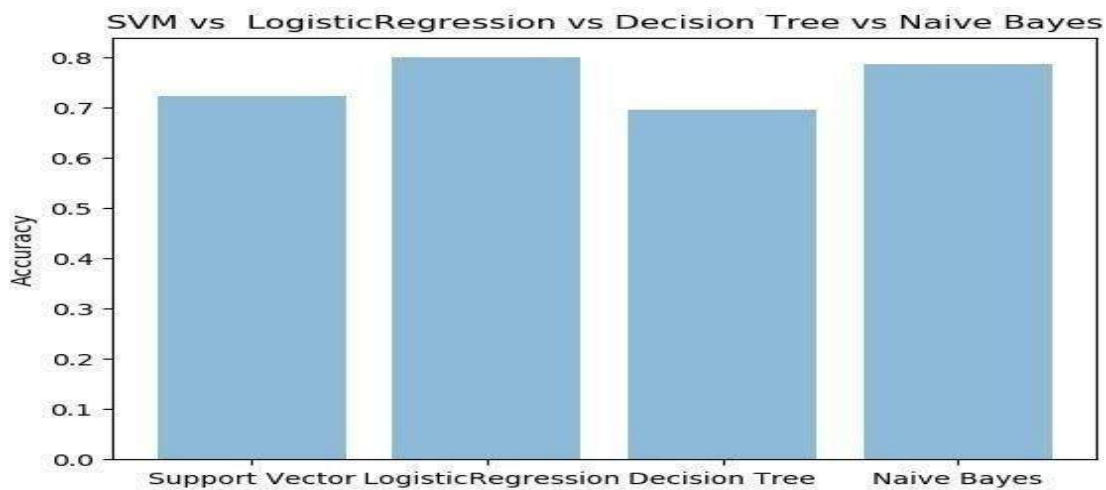
6.4.1.TESTING AT FIRST RUN

Figure 6.4.1: Testing at first run

6.4.2 TESTING AT SECOND RUN
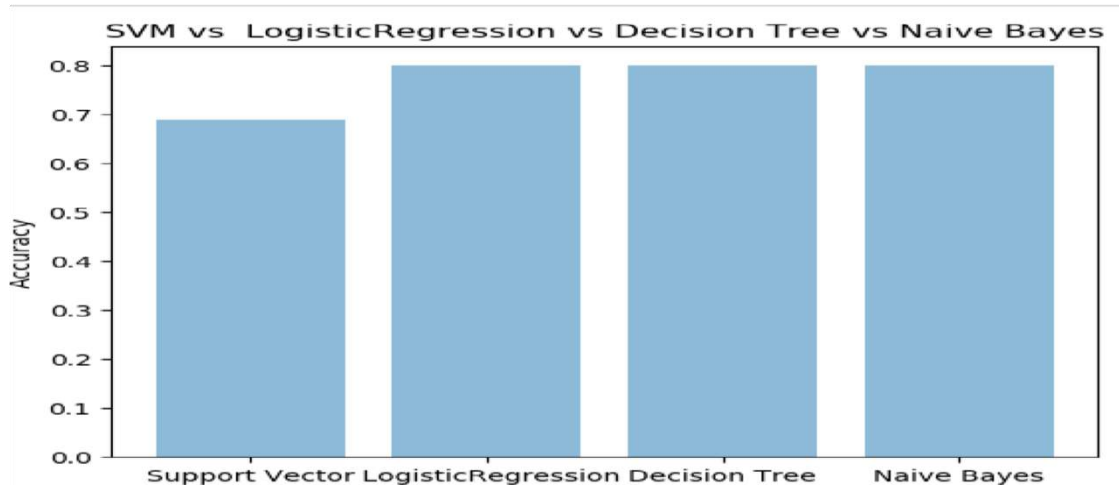

Figure 6.4.2: Testing at second run
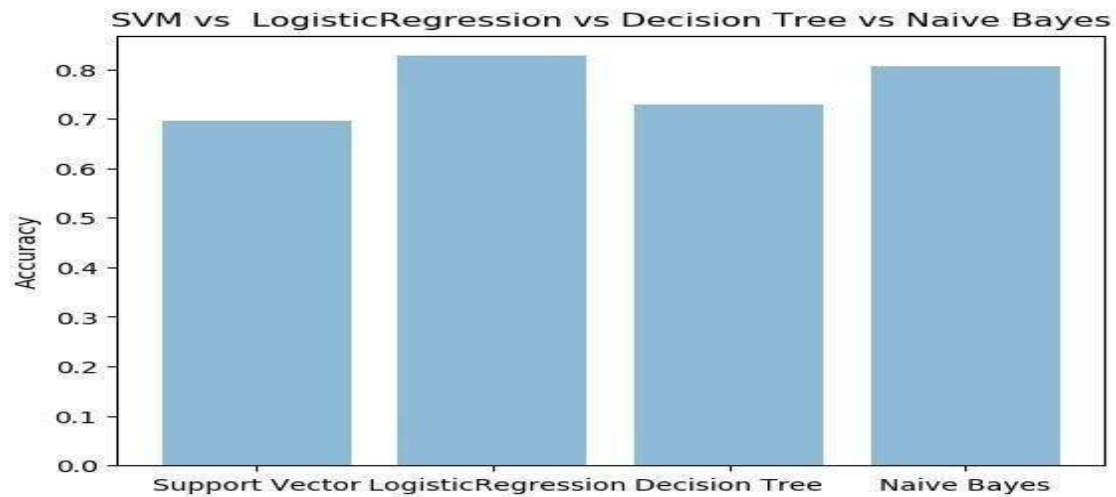
6.4.3 TESTING AT THIRD RUN

Figure 6.4.3: Testing at third run
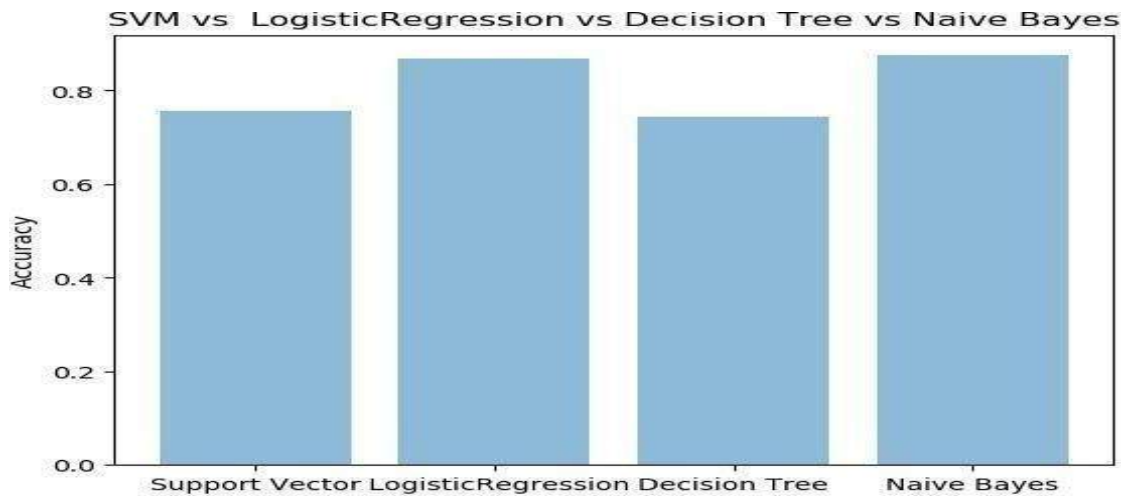
6.4.4 TESTING AT FOURTH RUN

Figure 6.4.4: Testing at fourth run

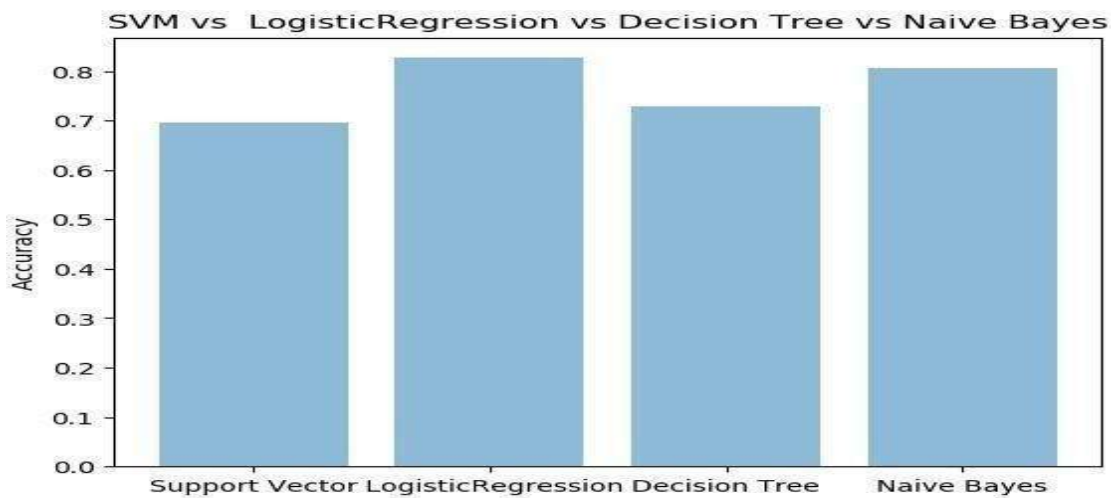### 6.4.5 ACCURACIES OF EACH ALGORITHM IN DIFFERENT RUNS

Figure 6.4.5: Accuracies of Each algorithm in different runs

By comparing the accuracy of all the four algorithms by running and testing them various times, we can say that Logistic Regression has the better accuracy than other algorithms in all test cases.

7. CONCLUSION

.

CONCLUSION

7.1 PROJECT CONCLUSION

The machine learning field is continuously evolving. And along with evolution comes a rise in demand and importance. There is one crucial reason why data scientists need machine learning, and that is: 'High-value predictions that can guide better decisions and smart actions in real-time without human intervention.'

Machine learning as technology helps analyze large chunks of data, easing the tasks of data scientists in an automated process and is gaining a lot of prominence and recognition. Machine learning has changed the way data extraction and interpretation works by involving automatic sets of generic methods that have replaced traditional statistical techniques.

From a proper analysis of positive points and constraints on the component, it can be safely concluded that the product is a highly efficient component. This application is working properly and meeting all Banker requirements. This component can be easily plugged in many other systems. There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system, So in the near future the so – called software could be made more secure, reliable and dynamic weight adjustment .In near future this module of prediction can be integrate with the module of automated processing system. Since we are using big data, analyzing such huge amount of data will not be a problem and the system gets trained according to the data given even if the amount of data is very large.

By this project, we can make it easy for the bank employers to predict to whom the loan can be sanctioned and it also minimizes time and the risk factor involved in approving the loan.

7.2 FUTURE ENHANCEMENTS

AI and machine learning are revolutionizing all sectors in the global economy with banking and financial institutions being no exception to this trend. They are adopting the latest software and bots to change the face and image of their industry and offer extraordinary services to their existing and future customers.Banking and financial sectors have been using some form of machine learning to keep track of data but it is usually tedious and manual in nature. With high volume of data, accurate historical records and the quantitative nature of financial institutions, this sector is particularly suited for artificial intelligence. Here are some scope of usage:

Improved customer support

How many times have you been irritated by the lack of real solutions when calling any bank helpline numbers? Machine learning is expected to change this picture with its intelligent solutions. Customer support will be aided with previous accounts details to help the system provide a quick and helpful response.

Enhanced data quality

Machine learning and AI will certainly enhance the quality of data. Right from collection to sorting and filtering complex and cumbersome calculations, a lot can be managed by modern day software and bots. It saves a large amount of manpower and yields higher accuracy in data.

Fraud prevention

Artificial intelligence plays an important role in data security which is one of the greatest responsibilities and concerns of banking and financial services. Preventing financial frauds and data robberies are two key areas where AI has been a game changer for financial institutions.

Digital financial assistants

Allo, Siri and Crotona are pioneers in the space of digital assistants. For financial institutions, these assistants are going to be the new big thing and AI and machine learning are the key propellers of this huge change. Assistants will help the banks to maintain a personal relationship with each employee without having to spare their human resources.

8. BIBLIOGRAPHY

..

8. BIBLIOGRAPHY

8.1 REFERENCES

1. https://www.datascience.com
2. https://www.machinelearningexpert.com
3. https://www.wikipedia.com
4. https://www.linkedin.com
5. https://www.pythonanywhere.com
6. https://www.anaconda.com
7. https://www.flask.com
8. https://www.github.com
9. https://www.towardsdatascience.com

Prediction of Loan using Machine learning Algorithm

M Chalapathi rao[1], Rachana Arelli[2], Nikhitha Maduri[3], Rachana Pallikonda[4]

[1]Associate Professor, CMR Technical Campus,
[2,3,4]CMR Technical Campus

---****************---

ABSTRACT

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 data of the previous records of the people to whom the loan was granted before. So the idea of this project is to gather loan data from multiple data sources and train machine learning algorithms on this data and to deploy this model we use flask framework which redirects it to a web browser that predicts loan approval/rejection. This model can be used by the organizations in making the right decision to approve or reject the applicant's request.

Keywords: Support vector machine, Logistic regression, Decision Tree, Naive Bayes

INTRODUCTION

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The primary aim is to allow the computers to learn automatically without human intervention or assistance and adjust actions accordingly.

Machine learning in conjunction with big data can not only collect information but also find specific patterns. The main portion of the bank's assets comes directly from the profit earned from the loans distributed by the banks. The prime objective in the banking environment is to invest their assets in safe hands where it is.

Today many banks/financial companies approve loans after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique. The aim of this system is to provide a quick, immediate and easy way to choose the deserving applicants.

This system is exclusively for the managing authority of a Bank/finance company, the whole process of prediction is done privately and no stakeholders would be able to alter the processing. Results against particular Loan Id can be sent to various departments of banks so that they can take appropriate action on application. This helps all others departments to carry out other formalities.

LITERATURE SURVEY

Data mining is the process of analyzing data from different perspectives and extracting useful knowledge from it. It is the core of the knowledge discovery process. The various steps involved in extracting knowledge from raw data. 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 techniques. This approach frequently employs Decision tree based classification Algorithms. 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.

Proposed System

The primary goal of this project is to extract patterns from a common loan approved dataset, and then build a model based on these extracted patterns, in order to predict the likely loan defaulters by using classification data mining algorithms. The historical data of the customers like their age, income, loan amount, employment length etc. will be used in order to do the anal ysis. Later on, some analysis will also be done to find the most relevant attributes, i.e. The factors that affect the prediction result the most. Using a different type of Machine Learning algorithm and Predicting accuracy result and Plotting a graph.

METHODOLOGY

System Architecture

This project architecture shows the procedure for loan approval prediction using machine learning Algorithms, starting from input to final prediction.
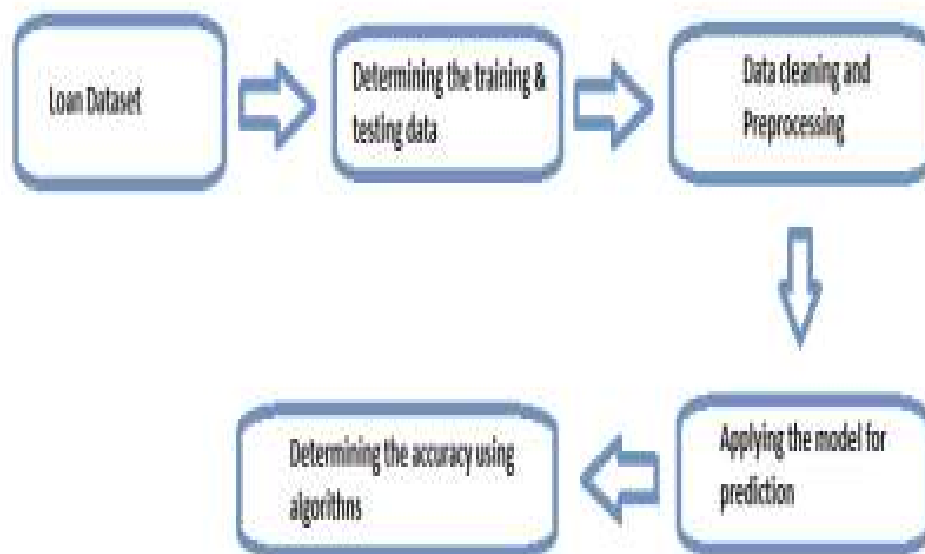
Machine Learning Classifiers

Support Vector Machine:

SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

Logistic Regression:

Logistic Regression is a Machine Learning algorithm which is used for classification problems, it is a predictive analysis algorithm based on the concept of probability.

Decision Tree:

Decision tree is one of the predictive modelling approaches used in statistics , data mining and machine learning . Decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions
.

Naive Bayes:

A Naive Bayes Classifier is a supervised machine-learning algorithm that uses the Bayes Theorem, which assumes that features are statistically independent. Regardless of this assumption, it has proven itself to be a classifier with good results.

Loan Approved

Loan Rejected

CONCLUSION

The machine learning field is continuously evolving. And along with evolution comes a rise in demand and importance. There is one crucial reason why data scientists need machine learning, and that is: 'High-value predictions that can guide better decisions and smart actions in real-time without human intervention.'

Machine learning as technology helps analyze large chunks of data, easing the tasks of data scientists in an automated process and is gaining a lot of prominence and recognition. Machine learning has changed the way data extraction and interpretation works by involving automatic sets of generic methods that have replaced traditional statistical techniques.
From a proper analysis of positive points and constraints on the component, it can be safely concluded that the product is a highly efficient component. This application is working properly and meeting all Banker requirements. This component can be easily plugged in many other systems. There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system, So in the near future the so – called software could be made more secure, reliable and dynamic weight adjustment .In near future this module of prediction can be integrate with the module of automated processing system. Since we are using big data, analyzing such a huge amount of data will not be a problem and the system gets trained according to the data given even if the amount of data is very large.

By this project, we can make it easy for the bank employers to predict to whom the loan can be sanctioned and it also minimizes time and the risk factor involved in approving the loan.

REFERENCES

[1]   https://www.datascience.com
[2]   https://www.machinelearningexpert.com
[3]   https://www.wikipedia.com
[4]   https://www.linkedin.com
[5]   https://www.pythonanywhere.com
[6]   https://www.anaconda.com
[7]   https://www.flask.com
[8]   https://www.github.com
[9]   https://www.towardsdatascience.com
[10]   https://en.wikipedia.org/wiki/Linear_regression#:~:text=In%20statistics%2C%20linear%20regression%20is,as%20dependent%20and%20independent%20variables).&text=Such%20models%20 are%20called%20linear%20models.
[11]   https://en.wikipedia.org/wiki/NumPy
[12]   https://en.wikipedia.org/wiki/Decision_tree#:~:text=A%20decision%20tree%20is%20a,only%20contains%20conditional%20control%20statements.
[13]   https://en.wikipedia.org/wiki/Support-vector_machine
[14]   https://en.wikipedia.org/wiki/Machine_learning

IJARESM

Certificate of Publication

M Chalapathi Rao

Associate Professor, CMR Technical Campus

TITLE OF PAPER

Prediction of Loan using Machine learning Algorithm

has been published in

IJARESM, Impact Factor: 7.429, Volume 9 Issue 6, June- 2021

Paper Id: IJARESM/June21

Date: 22-06-2021

Authorized Signatory

IJARESM

Certificate of Publication

Rachana Arelli

CMR Technical Campus

TITLE OF PAPER

Prediction of Loan using Machine learning Algorithm

has been published in

IJARESM, Impact Factor: 7.429, Volume 9 Issue 6, June- 2021

Paper Id: IJARESM/June21

Date: 22-06-2021

Website: www.ijaresm.com
Email: editor.ijaresm@gmail.com

Authorized Signatory

IJARESM

Certificate of Publication

Nikhitha Maduri

CMR Technical Campus

TITLE OF PAPER

Prediction of Loan using Machine learning Algorithm

has been published in

IJARESM, Impact Factor: 7.429, Volume 9 Issue 6, June- 2021

Paper Id: IJARESM/June21

Date: 22-06-2021

Website: www.ijaresm.com
Email: editor.ijaresm@gmail.com

Authorized Signatory

IJARESM

Certificate of Publication

Rachana Pallikonda

CMR Technical Campus

TITLE OF PAPER

Prediction of Loan using Machine learning Algorithm

has been published in

IJARESM, Impact Factor: 7.429, Volume 9 Issue 6, June- 2021

Paper Id: IJARESM/June21

Date: 22-06-2021

Authorized Signatory