IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005238|

Loan Prediction Using Machine Learning

B. Kalyan Chakravarthy[1] , Peddireddy Narendra Reddy[2], Miriyala Surendra Nath[3], Korlakunta Pavan Kumar[4], Muddana Harsha Vardhan[5]

Assistant Professor, Department of IT, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India[1]

UG Student, Department of IT, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India[2]

UG Student, Department of IT, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India[3]

UG Student, Department of IT, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India[4]

UG Student, Department of IT, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India[5]

ABSTRACT: In our banking system, banks have many products to sell but the main source of income of any bank is on its credit line. So they can earn from the interest of those loans which they credit. A bank's profit or a loss depends to a large extent on loans i.e. whether the customers are paying back the loan or defaulting. By predicting the loan defaulters, the bank can reduce its Non-performing Assets. This makes the study of this phenomenon very important. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default. But as the right predictions are very important for the maximization of profits, it is essential to study the nature of the different methods and their comparison. A very important approach in predictive analytics is used to study the problem of predicting loan defaulters (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting.

KEYWORDS: - Key word1: Big Data, Key word2: Machine Learning, Key word3: SVM, Key word4: Navies Bayes, Key word5: Prediction

I. INTRODUCTION

This Problem is completed by mining the large Data of the previous records of the people to whom the loan was granted before and on the basis of these accounts/experiences the machine was qualified using the machine learning model which gives the further accurate result.

The main objective of this paper is to predict whether assigning the loan to a specific person will be safe or not. We have implemented this loan prediction problem using the 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 techniques on a dataset which has 1000 cases and seven numerical and 6 categorical attributes. The credibility of a customer for sanctioning loan depends on several parameters, such as credit history, Installment etc.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005238|

Loan amount, costumer's history governs his credibility for receiving loan. The problem is to classify the borrower as defaulter or non-defaulter. However, developing such a model may be a very challenging task thanks to snowballing in demands for loans. A sample of the model is described in the paper which can be used by the organizations for making the correct or right decision to approve or reject the request for loan of the customers. This work includes the development of an ensemble model by combining three different machine learning models.

The aim of this Paper is to provide a quick, immediate and easy way to choose the worthy applicants. It can provide special compensations to the bank. The Loan Prediction System can automatically calculate the weight of individual features taking part in loan processing and on new test data the same features are processed with respect to their associated weight. A time edge can be set for the applicant to check whether his/her loan can be sanctioned or not.

MOTIVATION:

Loan approval is a very important process for banking organizations. The system approved or reject the loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. Using Machine learning we predict the loan approval.

II. LITERATURE SURVEY

1." Loan Approval Prediction based on Machine Learning Approach" Author- Kumar Arun, Garg Ishan, Kaur Sanmeet Year- 2018The main objective of this paper 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

2."Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process" Author E. Chandra Blessie, R. Rekha - Year- 2019 Extending credits to corporates and individuals for the smooth functioning of growing economies like India is inevitable. As increasing number of customers apply for loans in the banks and non- banking financial companies (NBFC), it is really challenging for banks and NBFCs with limited capital to device a standard resolution and safe procedure to lend money to its borrowers for their financial needs. In addition, in recent times NBFC inventories have suffered a significant downfall in terms of the stock price. It has contributed to a contagion that has also spread to other financial stocks, adversely affecting the benchmark in recent times. In this paper, an attempt is made to condense the risk involved in selecting the suitable person who could repay the loan on time thereby keeping the bank's nonperforming assets (NPA) on the hold. This is achieved by feeding the past records of the customer who acquired loans from the bank into a trained machine learning model which could yield an accurate result. The prime focus of the paper is to determine whether or not it will be safe to allocate the loan to a particular person. This paper has the following sections

(i) Collection of Data,

(ii) Data Cleaning and

(iii) Performance Evaluation.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005238|

Experimental tests found that the Naive Bayes model has better performance Evaluation.
Experimental tests found that the Naive Bayes model has better performance than other models in terms of loan forecasting.

3. "Loan Prediction using machine learning model" Year2019whether or not it will be safe to allocate the loan to a particular person. This paper has the following sections

(I) Collection of Data,

(ii) Data Cleaning and

(iii) Performance Evaluation.

Experimental tests found that the Naive Bayes model has better performance than other models in terms of loan fore casting. 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."Loan Prediction using Decision Tree and Random Forest"Author- Kshitiz Gautam, Arun Pratap Singh, Keshav Tyagi, Mr. Suresh Kumar Year-2020. In India the number of people or organization applying for loan gets increased every year. The bank have to put in a lot of work to analyze or predict whether the customer can pay back the loan amount or not (defaulter or non-defaulter) in the given time. The aim of this paper is to find the nature or background or credibility of client that is applying for the loan. We use exploratory data analysis technique to deal with problem of approving or rejecting the loan request or in short loan prediction. The main focus of this paper is to determine whether the loan given to a particular person or an organization shall be approved or not.

II. EXISTING SYSTEM

Bank employees check the details of applicants manually and give the loan to eligible applicants. Checking the details of all applicants takes a lot of time. The method in which two or more classifiers are combined together to produce an assembly model for the better prediction. They used the bagging and boosting techniques and then used random forest techniques. The process of classifiers is to improve the performance of the data and it gives better efficiency. Finally they concluded that the ensemble based algorithm improves the results for the training data set.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005238|

III. PROPOSED MODEL

In a Simple Term, Company wants to make automate the Loan Eligibility Process in a real time scenario related to customer's detail provided while applying application for home loan forms.
You will use the training set to build your model, and the test set to validate it.

This system predict whether the loan is approve or reject. This System refers the following things or ways.

Data Collection

Data Pre-processing (Data Cleaning)

Model Selection

Model Evaluation

Classification Result (output)

Libraries for Data Analysis
The models are implemented using Python 3.7 with listed libraries:

Pandas
Pandas is a Python package to work with structured and time series data. The data from various file formats such as csv, json, sql etc can be imported using Pandas. It is a powerful open source tool used for data analysis and data manipulation operations such as data cleaning, merging, selecting as well wrangling.

Sklearn
This python library is helpful for building machine learning and statistical models such as clustering, classification, regression etc. Though it can be used for reading, manipulating and summarizing the data as well, better libraries are there to perform these functions.

IV. METHODOLOGY

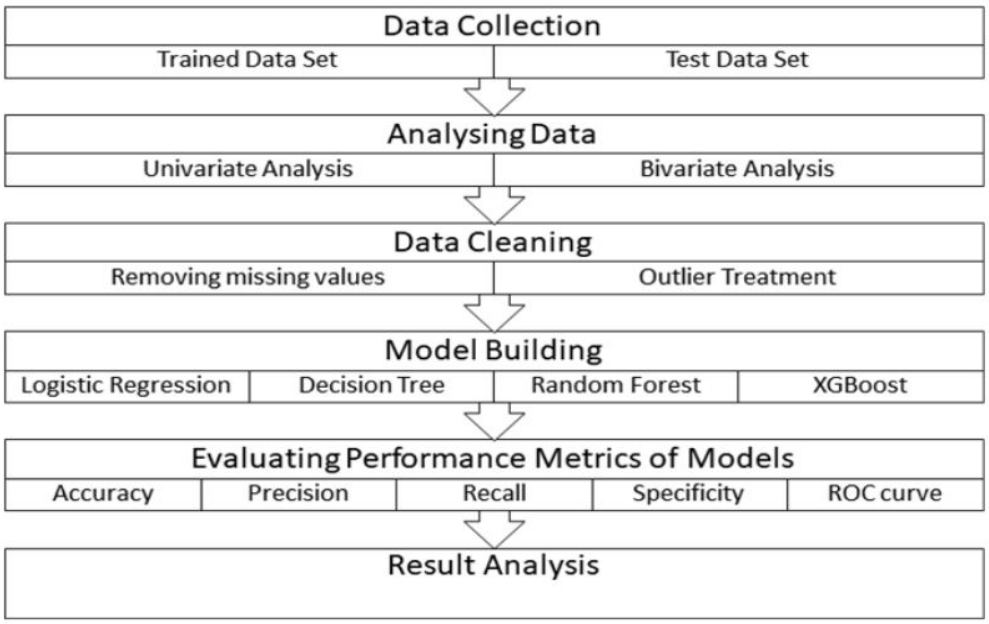

fig-1

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005238|

System Architecture

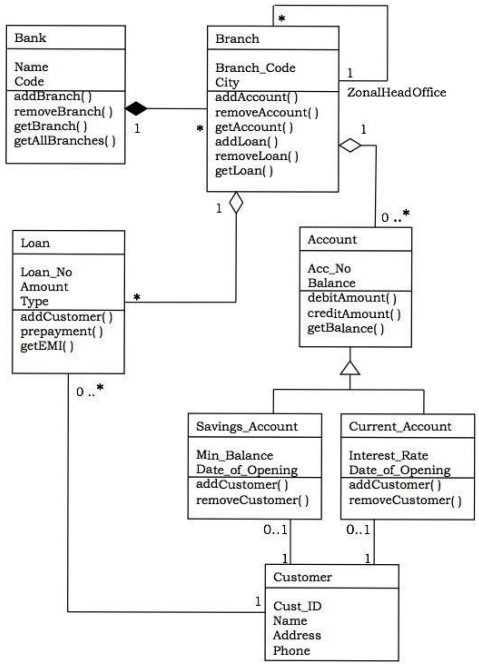

fig-2 Class Diagram

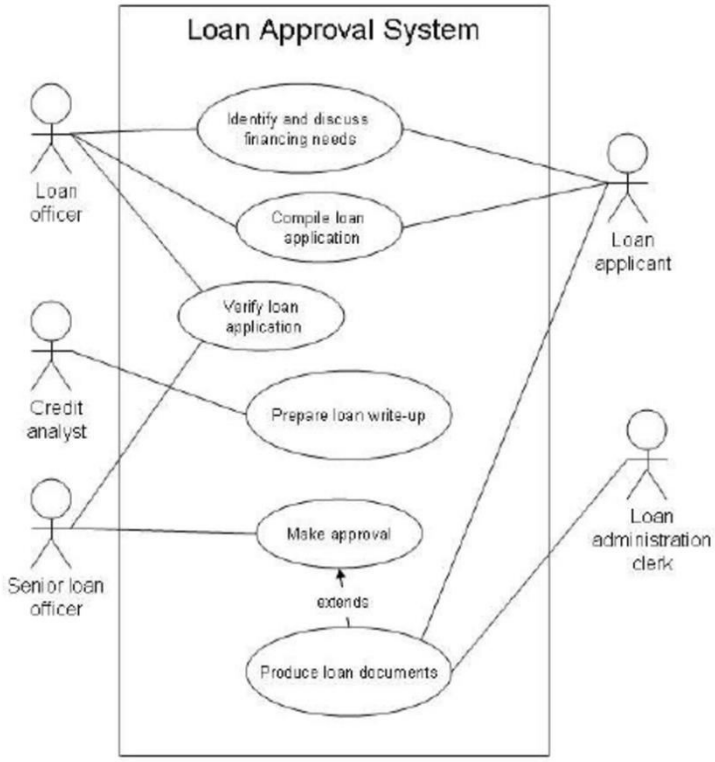

fig-3 Use Case Diagram

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005238|

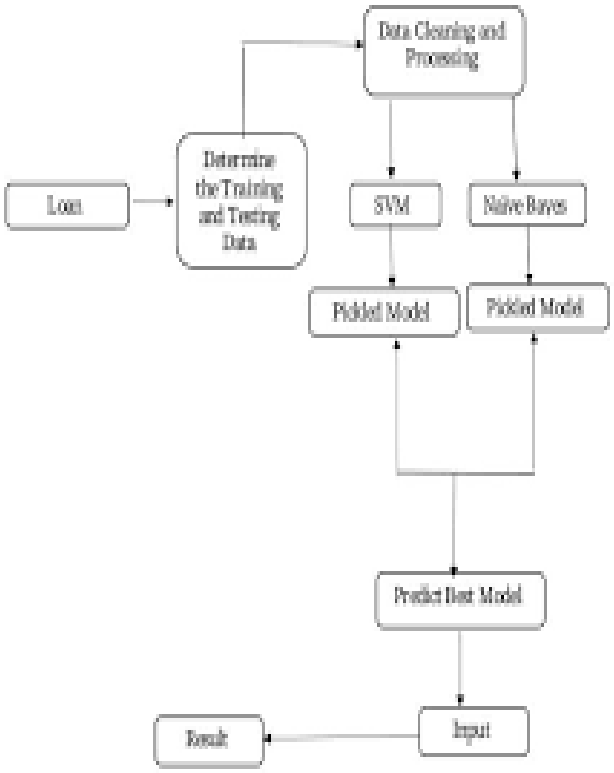

fig- 4

V. RESULTS & DISCUSSION

jupyter project (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
611    1
612    1
613    0
Name: Loan_Status, Length: 480, dtype: int64

In [24]: X_train, X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.1,stratify=Y,random_state=2)

In [25]: print(X.shape, X_train.shape, X_test.shape)
         (480, 11) (432, 11) (48, 11)

In [26]: classifier = svm.SVC(kernel='linear')

In [27]: classifier.fit(X_train,Y_train)
Out[27]: SVC(kernel='linear')

In [28]: # accuracy score on training data
         X_train_prediction = classifier.predict(X_train)
         training_data_accuray = accuracy_score(X_train_prediction,Y_train)

In [29]: print('Accuracy on training data : ', training_data_accuray)
         Accuracy on training data :  0.7986111111111112

In [30]: # accuracy score on training data
         X_test_prediction = classifier.predict(X_test)
         test_data_accuray = accuracy_score(X_test_prediction,Y_test)

In [31]: print('Accuracy on test data : ', test_data_accuray)
         Accuracy on test data :  0.8333333333333334
```

table – 1

	A	B	C
1		Loan_ID	Loan_Status
2	0	LP001015	Y
3	1	LP001022	Y
4	2	LP001031	Y
5	3	LP001035	N
6	4	LP001051	Y
7	5	LP001054	N
8	6	LP001055	Y
9	7	LP001056	N
10	8	LP001059	Y
11	9	LP001067	Y
12	10	LP001078	N
13	11	LP001082	Y
14	12	LP001083	Y
15	13	LP001094	N
16	14	LP001096	Y
17	15	LP001099	Y
18	16	LP001105	N
19	17	LP001107	Y
20	18	LP001108	Y
21	19	LP001115	N
22	20	LP001121	Y
23	21	LP001124	Y
24	22	LP001128	Y
25	23	LP001135	Y
26	24	LP001149	N

table-2

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005238|

Based on the accurate data the loan will be predicted and shows whether the applicant is eligible or not.

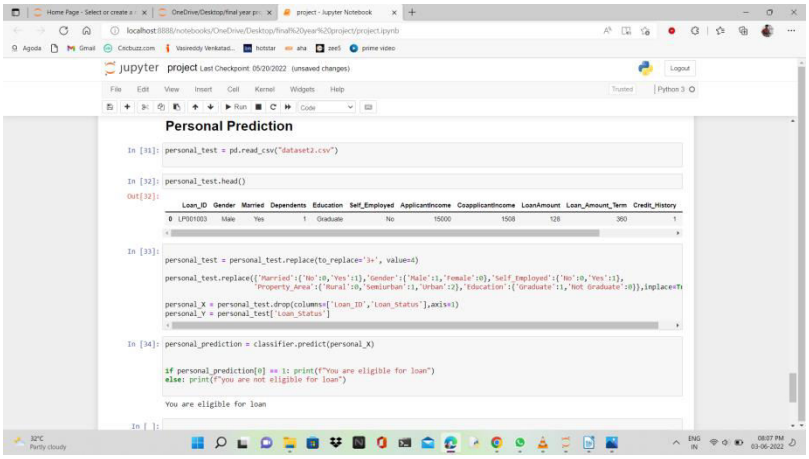

table - 3

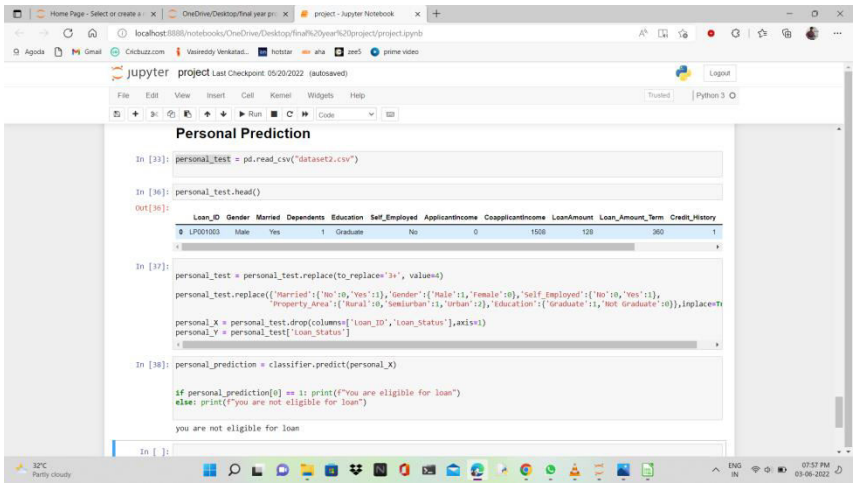

table – 4

VI. CONCLUSION

From a proper analysis of positive points and constraints on the member, it can be safely concluded that the product is a considerably productive member. This use is working duly and meeting to all Banker requisites. This member can be freely plugged in numerous other systems. There have been mathematics cases of computer glitches, violations in content and most important weight of features is fixed in automated prophecy system, so in the near future the so – called software could be made more secure, trustworthy and dynamic weight conformation. In near future this module of prophecy can be integrated with the module of automated processing system. The system is trained on old training dataset in future software can be made resembling that new testing date should also take part in training data after some fix time.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005238|

REFERENCES

1. *Kumar Arun, Garg Ishan, Kaur Sanmeet, May-Jun. 2016. Loan Approval Prediction based on Machine Learning Approach, IOSR Journal of Computer Engineering (IOSR-JCE).*

2. Wei Li, Shuai Ding, Yi Chen, and Shanlin Yang, Heterogeneous Ensemble for Default Prediction of Peer-to-Peer Lending in China, Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei University of Technology, Hefei 2009, China

3. Short-term prediction of Mortgage default using ensembled machine learning models, Jesse C.Sealand on july 20, 2018

4. Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques, Dr. K. Kavitha, International Journal of Advanced Research in Computer Science and Software Engineering

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details