Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

PREDICTION OF LOAN ELIGIBILITY OF THE CUSTOMER

Rayavarapu Jaya Lakshmi[1], Saragadam Mukesh[2], Thammu Venkata Naga Moulika[3], Mr. B N S V A Sankar[4]
[4] Assistant Professor, **Department of Computer Science & Engineering, Raghu Engineering College**, Vishakhapatnam, Andhra Pradesh
[1,2,3] Student of B-TECH, **Raghu Institute of Technology**, Vishakhapatnam, Andhra Pradesh
Email: - jayarayavarapu4@gmail.com, mukeshmickey149@gmail.com, moulikathammu@gmail.com, bnsvasankar@gmail.com

ABSTRACT
Predicting loan defaulters is a problem that is studied using a critical predictive analytics approach: Data is gathered from Kaggle for analysis and forecasting purposes. Machine learning algorithms have been used to create models, and various performance metrics have been calculated. Sensitivity and specificity, two performance metrics, are used to compare the models. The model produces diverse results, as evidenced by the end results. Therefore, by assessing their potential of loan default, the appropriate consumers can be quickly identified for loan issuing through the use of a machine learning algorithm technique. The model's conclusion is that a bank should evaluate a customer's other characteristics as well, as these factors are crucial in determining whether to give credit and in identifying potential loan defaulters. This goes beyond simply focusing on lending to wealthy clients.

keyword: - loan, machine Learning, defaulters, credit

1. INTRODUCTION
Financial institutions manage many different types of loans, such as home, vehicle, education, and personal loans. It can be found in both rural and urban settings. When a customer applies for a loan for the first time, the Finance Company confirms that the customer is eligible for approval. Applicants must complete out an application that requests information about their credit history, income, loan amount, number of dependents, gender, marital status, and education, among other things. As a result, a strong model is constructed using those particulars as input to confirm an applicant's eligibility for a loan application. Here, the applicant's "Loan Status" is the goal variable, while the other factors are predictors. Once the machine learning model has been constructed, a web application with an interface that enables users to quickly determine their loan eligibility by providing certain details will be established. This initiative has approved loans based on a set of criteria by using the data of previous customers of different banks. The machine learning model is trained on the record to yield reliable results. Ascertaining the loan eligibility of the client is our main project goal. To forecast loan eligibility, the Random Forest, Decision Tree, and Naive Bayes algorithms are used. To ensure that there are no missing values in the data set, the data are first cleansed.

2. LITERATURE SURVEY
[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 stages 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.Frencis ,JensyV.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.

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 open-ended 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.Francis Jency, 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 criteriaThe 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

[3]Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma,Namburi VimalaKumari, 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.

3. IMPLEMENTATION STUDY

Banks, Housing, Finance Companies and some NBFC (non-banking financial company) deal in various types of loans like housing loan, personal loan, business loan etc. In all over the part of countries. These companies have existence in Rural, Semi Urban and Urban areas. After applying loan by customer these companies validate the eligibility of customers to get the loan or not. This paper provides a solution to automate this process by employing machine learning algorithm. So, the customer will fill an online loan application form.

This form consists details like Sex, Marital Status, Qualification, Details of Dependents, Annual Income, Amount of Loan, Credit History of Applicant and others. To automate this process by using machine learning algorithm, First the algorithm will identify those segments of the customers who are eligible to get loan amounts so bank can focus on these customers

3.1 PROPOSED METHODOLOGY & ALOGRITHAM

The proposed model is based on our use of a machine learning algorithm known as Random Forest, decision tree, naive bayes to predict loan eligibility in this project. To train these algorithms, we used the following dataset. Since classification is the goal of the model's development, Random Forest with a sigmoid function is used to achieve it. Preprocessing is the significant region of the model where it consumes additional time and afterward Exploratory Information Examination which is trailed by Element Designing and afterward Model Determination. feeding the model the two distinct datasets and then preceding the model. To deal with the problem, we developed automatic loan prediction using machine learning techniques. We will train the machine with previous dataset. So, machine can analyse and understand the process. Then machine will check for eligible applicant and give us result. Advantages Time period for loan sanctioning will be reduced. Whole process will be automated, so human error will be avoided Eligible applicant will be sanctioned loan without any delay.

.3.2 Decision Tree Algorithm

Decision tree algorithms are commonly used for classification tasks, such as determining loan eligibility. Here's a simplified explanation of how you could use a decision tree algorithm for this purpose:

1. **Data Collection**: Gather data on past loan applicants, including features such as age, income, credit score, employment status, loan amount, etc., and whether they were approved or denied.
2. **Data Preprocessing**: Clean the data, handle missing values, and convert categorical variables into a format suitable for the algorithm.
3. **Splitting Data**: Divide the data into two subsets: one for training the model and another for testing its performance.

4. **Building the Decision Tree**: The algorithm will recursively split the data based on features that maximize the information gain or minimize impurity (such as Gini impurity or entropy). At each node of the tree, the algorithm selects the feature that best separates the data into distinct classes (e.g., approved or denied).
5. **Stopping Criteria**: The tree-building process continues until a stopping criterion is met, such as reaching a maximum depth, having a minimum number of samples in each node, or when further splits do not improve the model significantly.
6. **Prediction**: Once the tree is built, you can use it to predict the loan eligibility of new applicants by following the decision path down the tree based on their feature values.
7. **Evaluation**: Assess the performance of the model using metrics such as accuracy, precision, recall, or F1 score on the test dataset.
8. **Fine-tuning**: Adjust parameters of the decision tree algorithm, such as the maximum depth or minimum samples per leaf, to improve performance if necessary. You can also consider techniques like pruning to prevent overfitting.

3.3 Random Forest Algorithm

Certainly! Here's a step-by-step guide to detecting phishing websites using the Random Forest algorithm:
1. **Data Collection**: As with SVM, gather a dataset containing examples of both phishing and legitimate websites. Each example should have features describing various aspects of the website.
2. **Data Preprocessing**: Preprocess the dataset by cleaning the data, handling missing values, and encoding categorical variables if necessary. Ensure that all features are in a format suitable for Random Forest classification.
3. **Feature Selection/Extraction**: Select relevant features that are likely to distinguish between phishing and legitimate websites. You can use techniques like correlation analysis, feature importance from Random Forest, or domain knowledge to select the most informative features.
4. **Splitting the Dataset**: Split the dataset into training and testing sets. The training set will be used to train the Random Forest model, while the testing set will be used to evaluate its performance.
5. **Model Training**: Train the Random Forest model using the training dataset. Random Forest is an ensemble learning method that fits a number of decision tree classifiers on various sub-samples of the dataset. You can specify parameters such as the number of trees in the forest, the maximum depth of the trees, and the minimum number of samples required to split a node.
6. **Model Evaluation**: Evaluate the trained Random Forest model using the testing dataset. Use evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to assess the model's performance in distinguishing between phishing and legitimate websites.
7. **Hyperparameter Tuning**: Fine-tune the hyperparameters of the Random Forest model to improve its performance. This can be done using techniques like grid search or randomized search, where different combinations of hyperparameters are tried and evaluated using cross-validation.
8. **Model Deployment**: Once you're satisfied with the model's performance, deploy it to detect phishing websites in real-world scenarios. This may involve integrating the model into a web browser extension, an API for online scanning, or any other suitable deployment method.
9. **Monitoring and Maintenance**: Continuously monitor the model's performance in production and retrain it periodically using new data to ensure its effectiveness over time. Stay updated on emerging phishing techniques and adapt the model accordingly.

3**.4 Naive Bayes Algorithm**

Naive Bayes algorithm is another popular choice for classification tasks, including determining loan eligibility. Here's how you could use the Naive Bayes algorithm for this purpose:
1. **Data Collection**: Gather data on past loan applicants, including features such as age, income, credit score, employment status, loan amount, etc., and whether they were approved or denied.
2. **Data Preprocessing**: Clean the data, handle missing values, and convert categorical variables into a format suitable for the algorithm.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

3. **Splitting Data**: Divide the data into two subsets: one for training the model and another for testing its performance.
4. **Training the Model**: In the case of Naive Bayes, the algorithm calculates the probabilities of each feature given each class (approved or denied). It assumes that the features are conditionally independent given the class, hence the "naive" assumption. There are different variants of Naive Bayes, such as Gaussian Naive Bayes for continuous features and Multinomial Naive Bayes for discrete features.
5. **Prediction**: Once the model is trained, you can use it to predict the loan eligibility of new applicants by calculating the probability of each class given the applicant's features using Bayes' theorem. The class with the highest probability is chosen as the predicted class.
6. **Evaluation**: Assess the performance of the model using metrics such as accuracy, precision, recall, or F1 score on the test dataset.
7. **Fine-tuning**: Adjust parameters of the Naive Bayes algorithm if necessary. For example, in the case of Multinomial Naive Bayes, you might adjust smoothing parameters to handle unseen feature values.

4. RESULTS AND SCREEN SHOTS

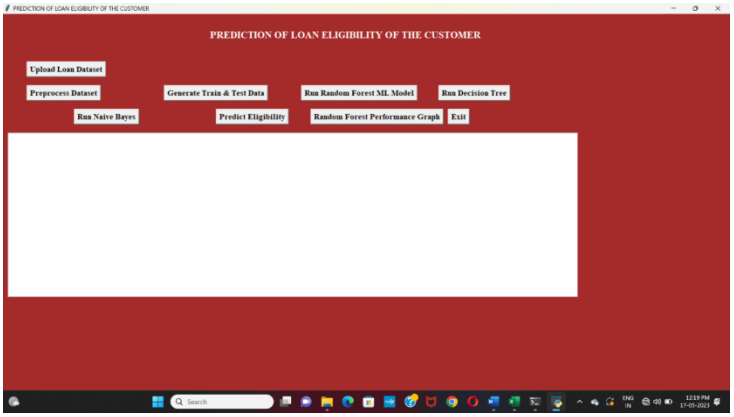

Fig 1:-In above screen is a front-end design, if we have chosen any button, you can click the button to get the result. In above screen click on 'Upload Loan Dataset' button to load dataset.

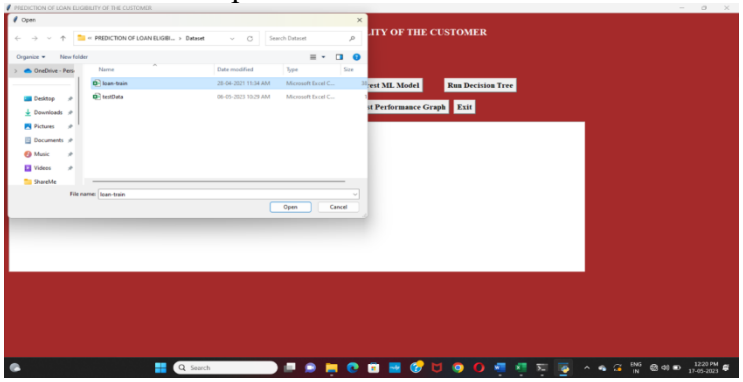

Fig 2:-In above screen selecting and uploading 'loan-train.csv' file and then click on 'Open' button to load dataset and to get the result you can observe the below screen

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

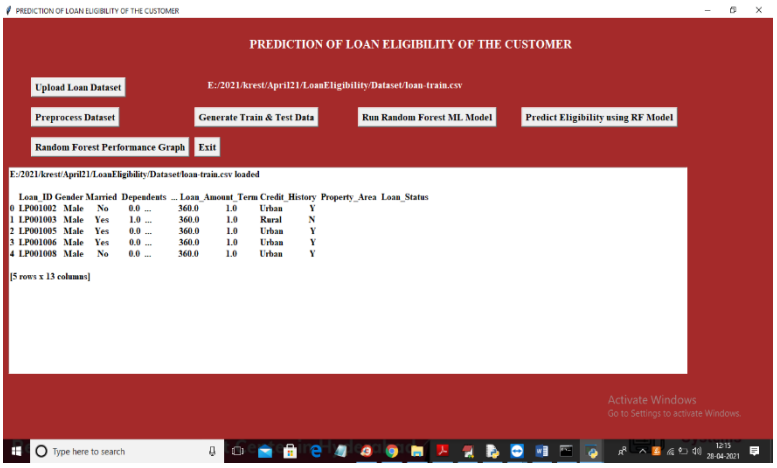

Fig 3:- In above screen dataset loaded and all columns contains non-numeric values and machine learning will not accept non-numeric values so we need to convert all those values to numeric by assigning IDs to them where MALE will replace with 0 and FEMALE will replace with 1 and below graph showing number of different values in dataset

Fig 4:- In above graph different colour lines represents counts of that column and you can see column names with colour in graph top right side. Now click on 'Preprocess Dataset' button to clean dataset

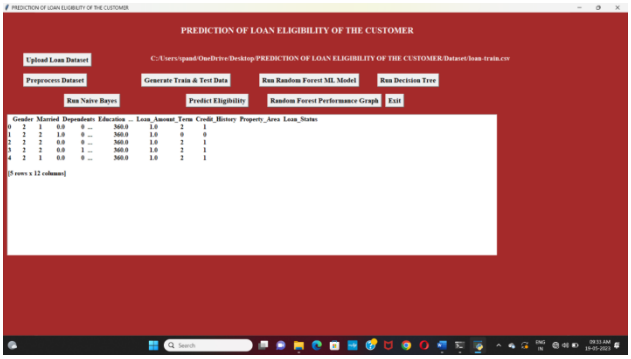

Fig5:-In above screen all non-numeric data is replace with numeric values because we don't accept the non -numeric values in machine learning.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

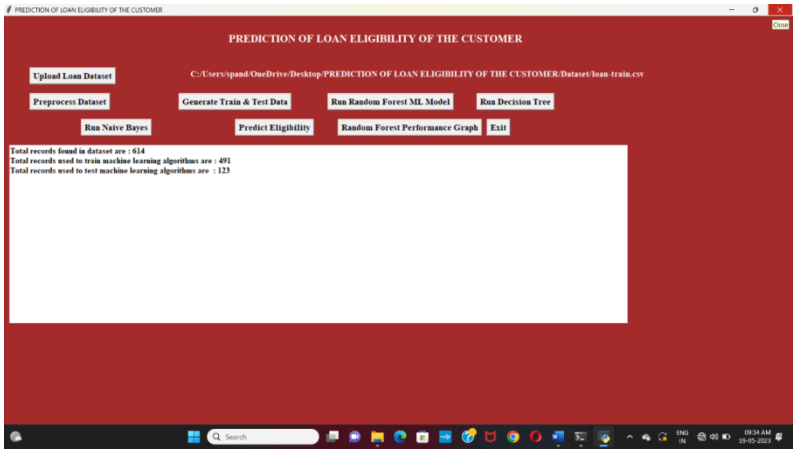

Fig 6 :-In above screen dataset contains 614 records and using 491 records to train ML and 123 records to testML accuracy. In below graph we can see importance of each attribute with other attribute by using graph correlation metric

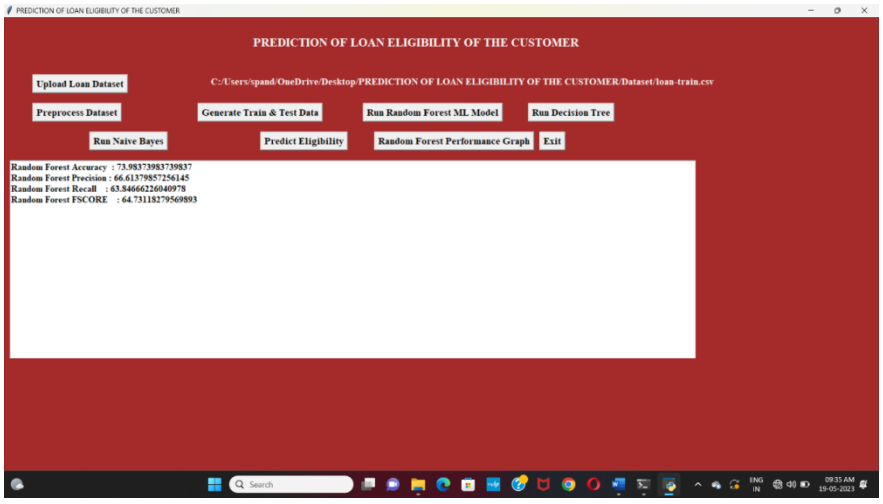

Fig 7: _In above screen random forest model generated with 73% accuracy and we can see its precision, recall and FSCORE value

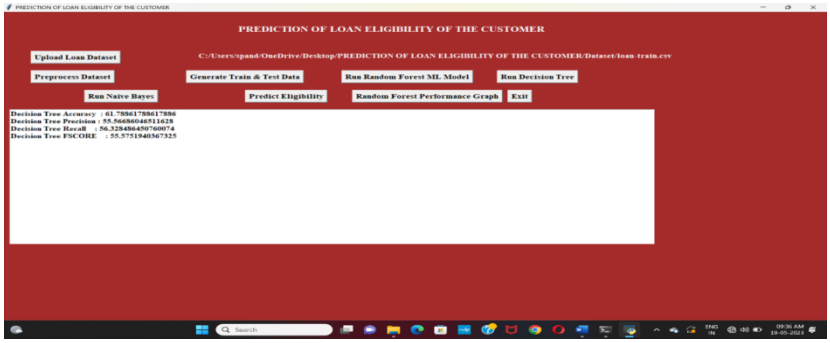

Fig 8:-In above screen Decision Tree generated with 62% accuracy, precision, recall and FSCORE value.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

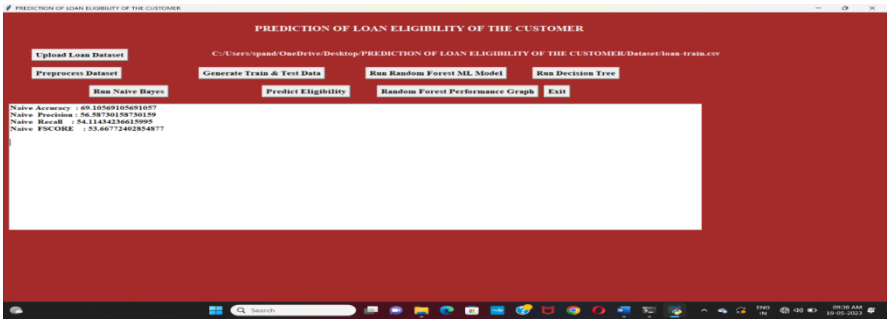

Fig 9: -In above screen Naïve bayes generated with 70% accuracy, precision, recall and FSCORE value and now click on 'Predict Eligibility using RF Model' button to upload test data and perform eligibility prediction

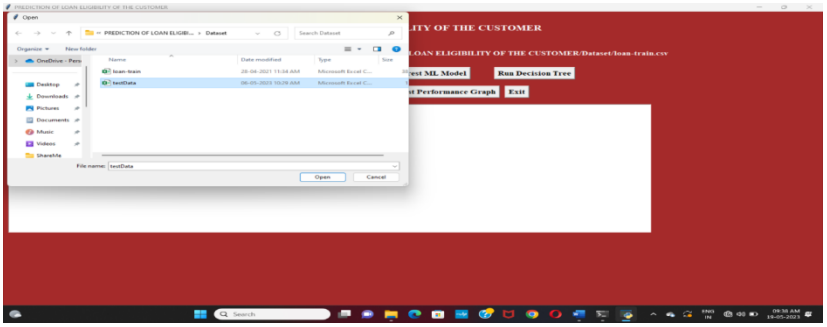

Fig 10:- In above screen selecting and uploading 'testData.csv' file and then click on 'Open' button to load test data and then will get below prediction result

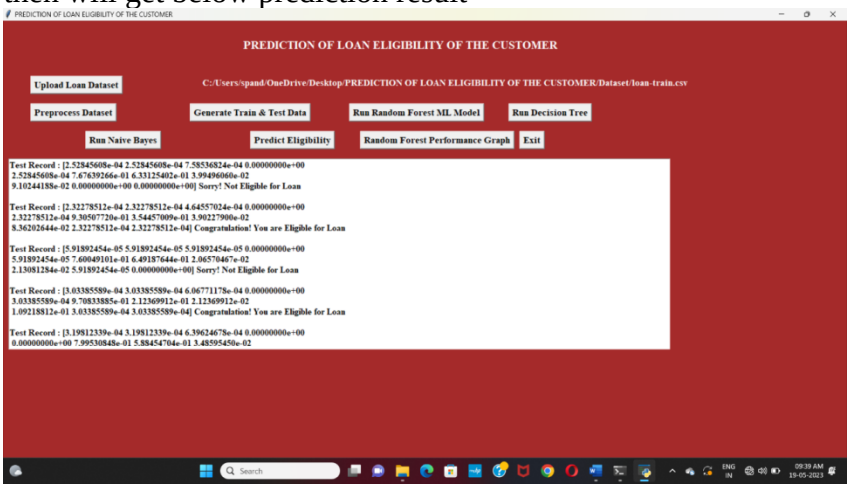

Fig 11:-In above screen in square bracket, we can see normalized test values and after square bracket we can see the prediction result as eligible or not eligible. You can scroll down above text area to view all predicted records and now click on 'Random Forest Performance Graph' button to get below graph.

Fig11:-In above graph we can see accuracy, precision, recall and FSCORE values of random forest, decision tree, naïve bayes and graph y-axis represents %value where accuracy got 80% and Precision got 65%. Each metric bar color name you can see from top right side.

5 CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

Therefore, the developed model automates the method of determining the applicant's credit worthiness. It focuses on an information containing the main points of the loan applicants. In this system random forest model is used. In Machine Learnings is one of the supervised learning algorithms, 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.

5.2 FUTURE SCOPE

In the future, these models can be used to compare various prediction models produced by machine learning algorithms, and the model with the highest accuracy will be chosen as the prediction model. In the future, this paper can be extended to a higher level. Prescient model for credits that utilizations AI calculations, where the outcomes from each diagram of the paper can be taken as individual rules for the AI calculation.

6 REFERENCES

[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] PhilHyo Jin 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, Kalpana Batra, Poonam Phogat," Manage big data using optical networks", Journal of Statistics and Management Systems "Volume 23, 2020, Issue 2, Taylors & Form