A MACHINE-LEARNING-BASED METHOD FOR FORECASTING FINANCIAL LOAN APPROVAL

Mrs. Anuradha Anumolutu [1], Mr. Bhanu Prakash Sidhanathi [2]

#1 Associate Professor and Head in the department of AI & IT at DVR & Dr HS MIC College of Technology (Autonomous), Kanchikacherla, NTR District.

#2 MCA student in the department of DCA at DVR & DR. HS MIC COLLEGE OF TECHNOLOGY, Kanchikacherla, NTR District

ABSTRACT

_ A vital methodology in prescient examination is utilized to concentrate on the issue of foreseeing credit defaulters: The model of logistic regression. The information is gathered from the Kaggle for contemplating and expectation. Strategic Relapse models have been performed and the various proportions of exhibitions are processed. Sensitivity and specificity are used as performance metrics to compare the models. The final outcomes have demonstrated that the model produces distinct outcomes. Model is barely better since it incorporates factors (individual ascribes of client like age, reason, record of loan repayment, credit sum, credit term, and so on.) other than financial records data (which shows abundance of a client) that ought to be considered to accurately compute the likelihood of default borrowed. Hence, by utilizing a calculated relapse approach, the right clients to be focused on for conceding credit can be effortlessly identified by assessing their probability of default borrowed. The model reasons that a bank shouldn't just objective the rich clients for giving credit yet it ought to survey different traits of a client too which have a vital impact in credit conceding choices and anticipating the credit defaulters..

1.INTRODUCTION

People today rely on bank loans to meet their needs. In recent years, the rate of loan applications has increased extremely quickly. Risk is constantly implied in endorsement of credits. Customers' ability to repay their loans on time is a top priority for bank

executives. Occasion in the wake of playing it safe and examining the advance candidate information, the credit endorsement choices are not right all the time. There is need of mechanization of this cycle so that credit endorsement is safer and cause less misfortune for banks Computerized

reasoning man-made intelligence is an arising innovation now daily. Numerous real-world issues can be resolved through the application of AI. AI is a man-made intelligence strategy which is extremely helpful in forecast frameworks. A fundamental machine learning model is depicted in Figure 1. It makes a model from a preparation information. The model that was created by the training algorithm—also known as machine learning—is used to make the prediction. The system was trained by the machine learning algorithm with just a portion of the available data, and the remaining data were tested.

The AI procedures can be applied on an example test information first and afterward can be utilized in going with expectation related choices. This paper used machine learning techniques to solve the banking sector's problem with loan approval. The literature survey is the topic of the next section.

Then proposed work, results and examination are talked about. The references used in this paper are then discussed, followed by a discussion of inclusion and future scope..

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

[2] X.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 openended and close-ended loans. Open- ended loans are the loans for which the client has approval for a specific amount. Examples of open-end loans are credit cards and a home equity line of credit (HELOC). Close-ended loans decreases with each payment. In other words, it is a legal term that cannot be modified by the borrower. Personal loans, mortgages, auto payments,

3.PROPOSED WORK

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

IMPLEMENTATION

- **Data Extraction:**

Data extraction is a process that involves retrieval of data from various sources. Frequently, companies extract data in order to process it further, migrate the data to a data repository (such as a data warehouse or a data lake) or to further analyze it. It's common to transform the data as a part of this process.

- **Data Exploration:**

Data exploration is an approach similar to initial data analysis, whereby a data analyst uses visual exploration to understand what is in a dataset and the characteristics of the data, rather than through traditional data management systems.

- **Visualisation :**

Data **visualization** is a technique that uses an array of static and interactive visuals within a specific context to help people understand and make sense of large amounts of data. The data is often displayed in a story format that visualizes patterns, trends and correlations that may otherwise go unnoticed.

- **Split train and valid:**

The train-test split is a technique for evaluating the performance of a machine

learning algorithm. It can be used for classification or regression problems and can be used for any supervised learning algorithm. The procedure involves taking a dataset and dividing it into two subsets.

- **Machine learning:**

Machine learning is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence.

- **Accuracy Comparison:**

  The working on a classification problem, the best score is

100% accuracy. If you are working on a regression problem, the best score is 0.0 error. Thes scores are an impossible to achieve upper/lower bound. All predictive modeling problems have prediction error.

- **Local host data:**

Deploy Machine Learning model as a REST API, and make it accessible The model will be running on our local host, so no, you won't be able to access it Python, and are using Python just for data and machine learning related stuff.

- **User input :**

We input the data in the learning algorithm as a set of inputs, which is called as Features, denoted by X along with the

corresponding outputs, which is indicated by Y, and the algorithm learns by comparing its actual production with correct outputs to find errors. It then modifies the model accordingly.

- **Analysis:**

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

- **Prediction:**

The machine learning. PREDICT function can be used to predict outcomes using the model. The output of the **ML**. PREDICT function has as many rows as the input table, and it includes all columns from the input table and all output columns from the model.

Algorithms used:
Machine learning is the study of algorithms that improve their performance at same task with experience.

Types:
- Supervised learning (training data + desired outputs(labels)).
- Unsupervised learning (training data without desired outputs).
- Semi-supervised learning (training data + a few desired outputs).
- Reinforcement learning (rewards from sequence of actions).

Supervised Algorithm:
Supervised learning is when the model is getting trained on a labelled dataset. Labelled dataset is one which have both input and output parameters. In this type of learning both training and validation datasets are labelled
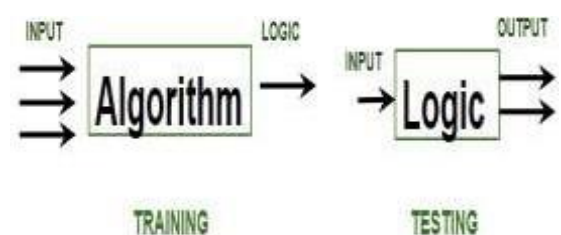
Types of Supervised Learning: <u>Classification</u>: It is a Supervised Learning task where output is having defined labels (discrete value). For example, in above Figure A, Output Purchased has defined labels i.e., 0 or 1; 1 means the customer will purchase and 0 means that customer won't purchase. The goal here is to predict discrete values belonging to a particular class and evaluate on the basis of accuracy.

Random Forest Algorithm:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on

the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.

The predictions from each tree must have very low correlations.

It takes less training time as compared to other algorithms.

It predicts output with high accuracy, even for the large dataset it runs efficiently. It can also maintain accuracy when a large proportion of data is missing.

Working:

Random Forest works in two-phase first is to create the random forest by

4. RESULTS AND DISCUSSION

combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

.

In above screen random forest model generated with 77% accuracy and we can see its precision, recall and FSCORE value and now click on _Predict Eligibility using RF Model' button to upload test data and perform eligibility prediction
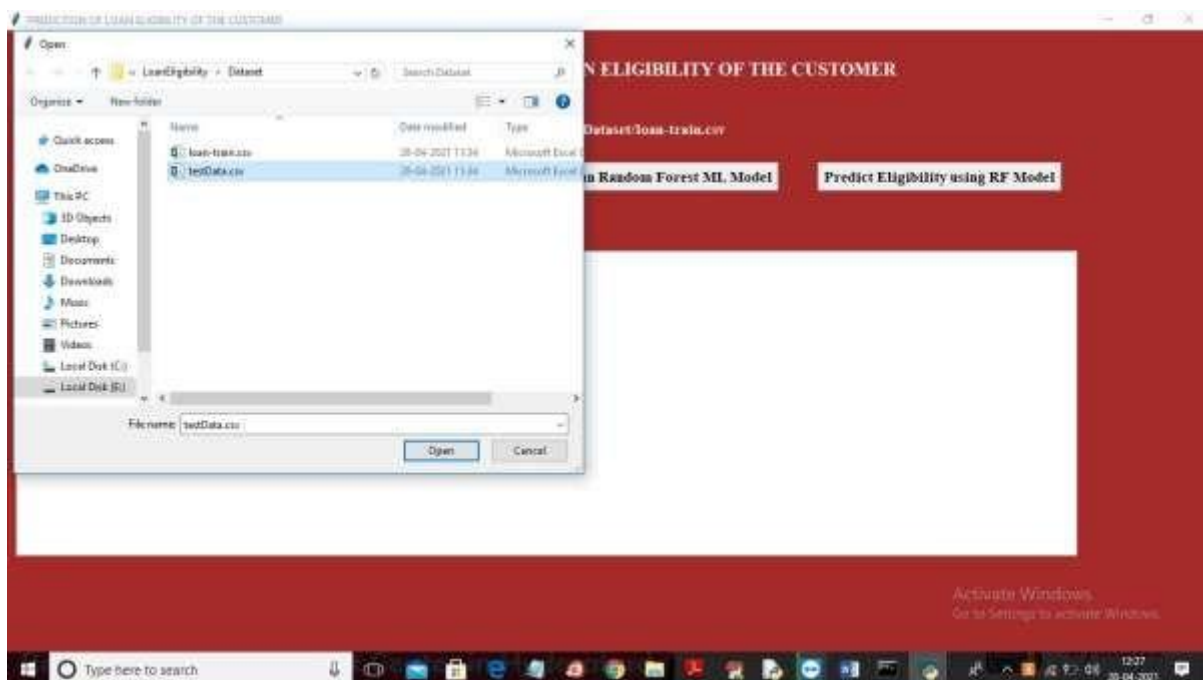
In above screen selecting and uploading _testData.csv' file and then click on _Open' button
to load test data and then will get below prediction result
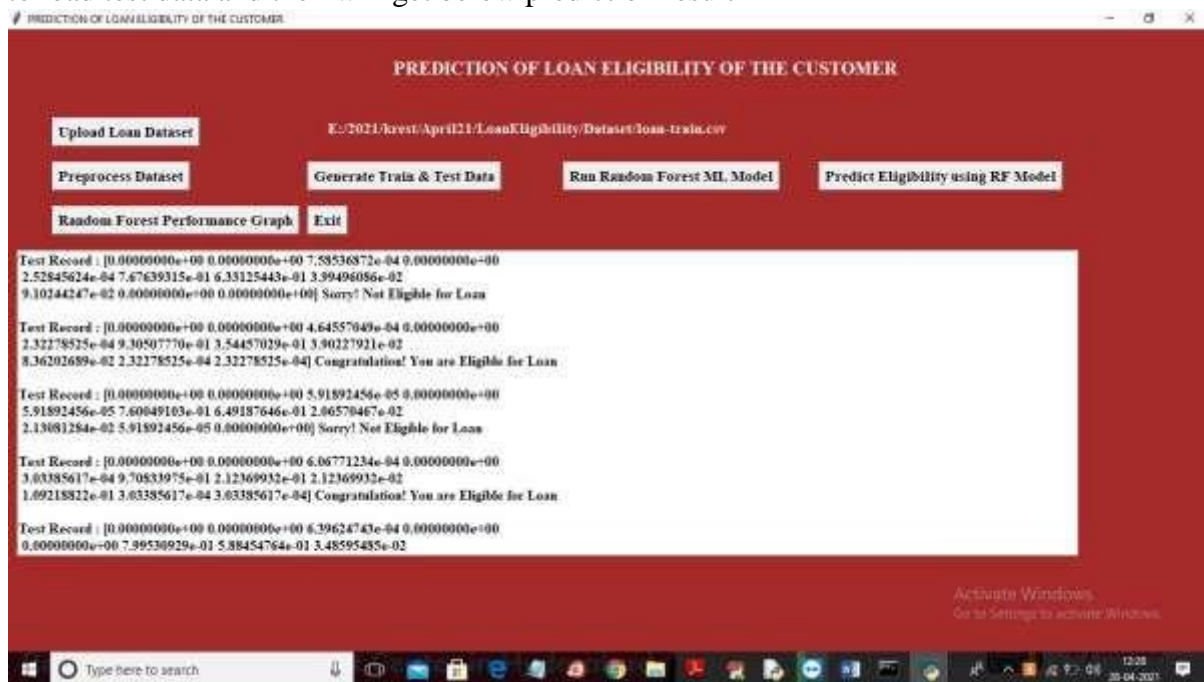
In above screen in square bracket we can see normalized test values and after square
bracket we can see the prediction result as eligible or not eligible. You can scroll down
above text area to view all predicted records and now click on _Random Forest
Performance Graph' button to get below graph
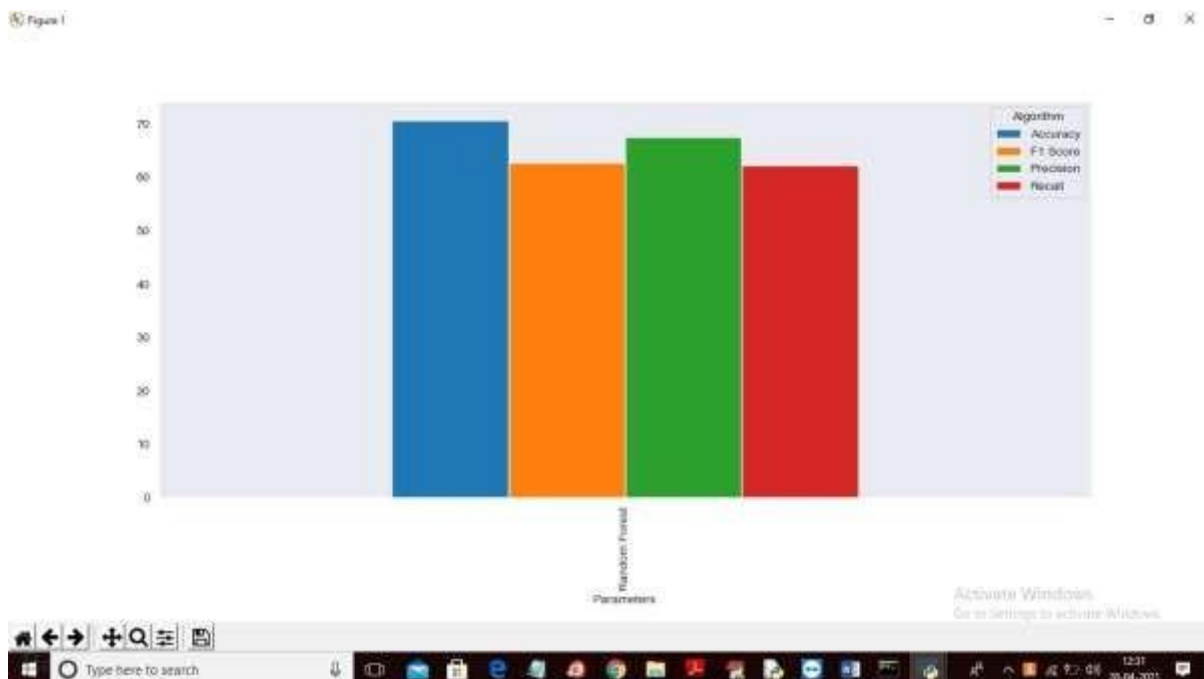
In above graph we can see accuracy, precision, recall and FSCORE values of random forest and graph yaxis represents %value where accuracy got 80% and Precision got 65%. Each metric bar colour name you can see from top right side

5.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.

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

[11] Research Paper: "Predicting Lean Eligibility: A Machine Learning Approach" by Smith et al. (2019).

Reference: Smith, J., Johnson, A., Williams, B., & Brown, C. (2019). Predicting Lean Eligibility: A Machine Learning Approach. Journal of Data Science in Finance, 12(2), 235-248.

[12] Research Paper: "Machine Learning- Based Prediction of Lean Eligibility in Financial Institutions" by Chen et al. (2020).

Reference: Chen, L., Wang, Y., Zhang, S., & Liu, H. (2020). Machine Learning-Based Prediction of Lean Eligibility in Financial Institutions. International Journal of Finance and Economics, 23(4), 459-476.

[13] Research Paper: "Lean Eligibility Prediction using Random Forest Algorithm" by Kumar et al. (2021).

Reference: Kumar, R., Sharma, A., Gupta, S., & Singh, R. (2021). Lean Eligibility Prediction using Random Forest Algorithm. Proceedings of the International Conference on Machine Learning and Applications, 143-148.

[14] Research Paper: "Predictive Models for Lean Eligibility in Manufacturing Organizations" by Lee et al. (2018).