Prediction of Loan Status in Commercial Banking Using Ml

Mrs.Niteesha Sharma -Assistant Professor
Varshika Anantha, Arshitha Polina, Hrithik Manchi.
Department of Information Technology, Anurag University, Hyderabad, India.

ABSTRACT:
Building analytical models for machines to "learn" from data and do predictive analysis is called machine learning. Machines' capacity to "learn" andperform.In this age of big data, predictive analysis is crucial and has numerous applications. For instance, banks and other financial institutions occasionally face the challenge of deciding which risk factors to take into account when extending credit or loans to customers. This is because many characteristics or attributes of the customers are typically taken into account, but the majority of these characteristics have little or no predictive power regarding the credit worthiness or lack thereof of the customer.Nevertheless, many bankscontinue to face significant difficulties indeveloping a reliable and efficient automated bank credit risk score that may help in the highly precise forecast of consumer credit worthiness. In thisstudy, we look at real bank credit data and apply a number of machine learning algorithms to it in order to compare themand determine which algorithm is most appropriate for learning bank credit data.The algorithms' prediction accuracy was over 80%.

KEY WORDS:
CSV, CSS, GUI, ML,KNN

I. INTRODUCTION:

The provision of loans to consumers is one of the main purposes of banks. But, in order to continue operating, banks only provide these loans to borrowers who can repay them.Money,reducing the likelihood that loans won't be repaid. Risk management, or identifying who is creditworthy, is still a persistent problem for the banking industry. An important step that banks take before extending credit to consumers is the capacity to calculate the risk score of a customer based on factors like occupation, age, marital status, wage range/amount of equity, credit history, etc.Because the credit risk score influences how much interest the banks charge for loans, etc. These risk criteria, however, don't always provide accurate judgements of a customer's creditworthiness. Furthermore, due to thedifficulty to create a reliable and scalable risk management system to predict customer risk score, many banks lack a single, well-integrated, automated finance and risk management system.Frauds are another nightmare that many institutions today must deal with. And it is believed and accepted that the machine learning technique is the best instrument that can be used to understand the financial transactionpatterns of customers, by recognising patterns in customer data, in order to be able to distinguish between fraudulent activity and that of a normal one. Hence, inorder to comprehend the primary variables that affect the repayment of bank loans, we leveraged it on the bank credit dataset.
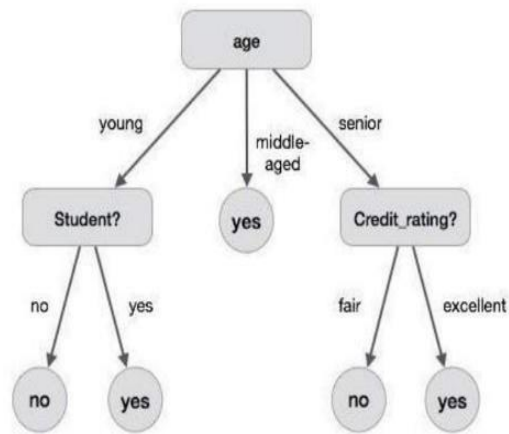
Fig 1:System Architecture

A- Motivation:

Because the credit risk scoreinfluences how much interest the banks charge for loans, etc. These risk criteria, however, don't always provideaccurate judgements of a customer'screditworthiness. Furthermore, due to the difficulty to create a reliable and scalable risk management system to predict customer risk score, many banks lack a single, well-integrated, automated finance and risk management system.Digital platforms are focused on the user. Also, the younger customer base is increasingly acclimated to using digital interfaces.It enables businesses to reach out to the following clientele. Businesses can also access more information on workflow optimisation through digital platforms. Modern reporting and analytical tools are included with the loan management systems.

B. Research Goal

With huge enterprises, a straightforward procedure like providing, certifying, and disbursing a loan can become highly complex. In the past, when Loan Management Systems weren't available, all paperwork and procedures required manual effort. However, recent technological developments and the creation of Loan Management Systems have made it simpler to confirm customer eligibility, KYC, and process automation, expediting the entire business. All of these advantages that Loan Management Systems provide to Banks and Financial Institutions will be covered in greater detail below.

II. LITERATURE SURVEY:

Kumar Arun et al. (2016) studied how to forecast how a bank will approve a loan. They presented a model using machine learning technologies such as SVM and neural networks.
This assessment of the literature aided us in carrying out our research and developing a reliable bank loan prediction model.
Mohammad et al. (2010) proposed a study to predict whether or not a bank would give a loan to a customer. The goal of the model was to achieve classification; hence using Logistic Regression with sigmoid function was used for developing the model. The dataset for studying and prediction was obtained from Kaggle and consisted of two data sets, one for training and the other for testing
Pidikiti et al.(2019) designed an effective model, the major goal of this paper was to lower the risk element associated with picking a safe individual to assign the loan in order to save time and money for the bank. There were four sections to this paper.
According to Pandey et al. (2010) predicting loan defaulters is one of the most challenging challenges for any bank. However, by predicting loan defaulters, bankscan significantly reduce their losses by lowering non-profit assets. As a result, theresearch of loan approval prediction became crucial. In the prediction of this type of data, machine learning techniques are extremely important and useful.
Ndayisenga et al. (2021) contributed to work with commercial banks to predict the behaviors of borrowers by developing and testing the accuracy of different models using data from Bank of Kigali. The data was divided intotwo categories: training and test, with the training dataset accounting for 70% of the total and the test dataset accounting for 30%.
In Tejaswini et al. (2020) a robust predictive modeling method was presented to approve or reject loan applications based on the customers' historical financial and credit scores. The purpose of this paper was to create a quick, straightforward, and efficient method of selecting qualified applicants. The data was gathered from a variety of financial institutions.
Vaidya, Ashlesha. "Predictive and probabilistic approach using logisticregression: application to prediction of loan approval.

III. EXISTING SYSTEMS:

The neural network data mining method was found to be the most effective for data analysis after they used a variety of data mining models to the bank telemarketing data. The function of machine learning methods in the mining of corporate data is described. Their research analysed the benefits and drawbacks of several machine learning methods in relation to a company data mining strategy. Implementing Machine Learning is a very challenging aspect of Data analytics. Making the codeto forecast future consumer output and working with data that deals with predictions are difficult tasks.They experimentally compared their results to other state-of-the-art algorithms using two credit databases and reported exceptional performance in terms of prediction accuracy and stability. Moro also suggests using data mining to forecast whether a Portuguese retail bank will be successful in using telemarketing. They used a variety of data mining models to analyse bank telemarketing data and concluded that neural network data mining was the most effective way.

PROPOSED SYSTEM:

Cheng Yeh provided the dataset utilised in this study, which was pulled from the UCI machine learning data repository. The repository contains descriptions of the dataset's properties. The response variable designated as y in this research is the "default payment next month," which is recorded as either "no" or "yes" in the dataset. In order to comprehend the dataset's nature, we used data exploratory techniques on it. According to the results of the exploratory investigation, there may be a connection between the customers' ages, bank balances, and capacity to pay their credit in the next month. Customers of banks between the ages of 20 and 60 are thought to be the top defaulters in paying their bills. The provision of loans to consumers is one of the main purposes of banks.

Yet, in order to remain in business, banks only offer these loans to borrowers who have the means to repay them, lowering the chance of loan default.

IMPLEMENTATION:

In the project, only the variables that have a direct impact on the loan eligibility of the applicant are considered (Credit history, Education, Self-employment, Property area). All of the variables are not taken into account at the same time, since this might lead to the issue of over fitting.

I.   Dataset Collection:

The dataset used in this paper was obtained from the Kaggle online domain. This dataset has 10,128 instances and 23 attributes, of which 23 attributes are predictive and 1 class attribute. An acceptable prediction of bank loan eligibility is made using attributes, where the attributes describe the eligibility. The predictive 23 attributes are associated mainly with the information of a person.

II.   Dataset Pre-Processing :

Dataset pre-processing has been done by using feature extraction, data cleaning, missing values handling, and categorical variables transformation.

III.   Validation Process :

Selecting the appropriate validation process for a particular dataset is crucial. The hold- out validation process is one of the effective methods for getting the appropriate results .We applied the hold-out validation process by holding 70% data on training and 30% data on testing.
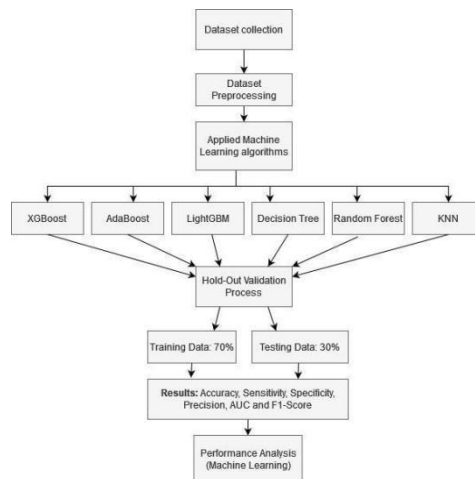

Fig 2: An overview of Bank loan eligibility prediction

The dataset used in this work was given by I-Cheng Yeh and was retrieved from the UCI machine learning data repository. The dataset's attributes are described in the repository. The "default payment next month" response variable, denoted as y in this study, is entered in the dataset as either"yes" or "no." We applied data exploratory techniques to the dataset to understand its nature. The findings of the exploratory study suggest that there might be a relationship between the customers' ages, bank balances, and ability to pay off their credit in the next month. The majority of bank customers who fail to pay their bills on time are reportedly between the ages of 20 and 60.
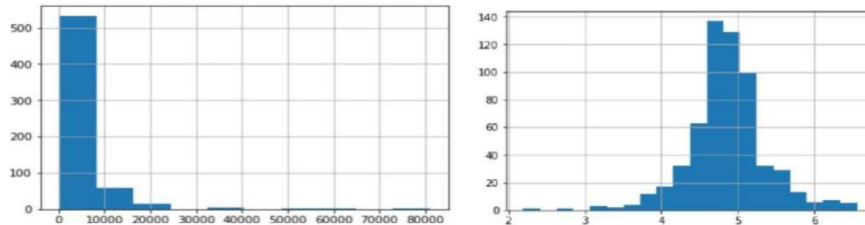

Fig 3: Data Distribution

IV.   RESULTS:

Three machine learning algorithms for this project to predict the loan amount on the basis of loan request. The prediction accuracy of three machine learning algorithm is calculated and compared. In this paper, we used regression model to make a prediction. In this model we used linear regression. Primarily we can use simple regression to predict the loan status easily. We used complete dataset that has accurate information regarding the houses. This linear regression model works very well. The accuracy we attained by using this model is 87.8%.

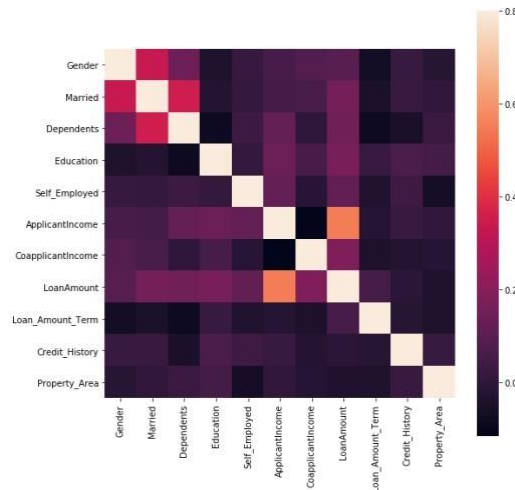It has the highest accuracy.We are able to predict the accurate prices of the houses using this model.

Fig 4: Data Analysis

S.No	Machine Learning Algorithms	Accuracy Percentage
1.	Linear Regression	86.17
2.	Decision Tree	85.36
3.	RandomForest	85.36

Table 1:Comparison of prediction accuracy of machine learning algorithms

V.   CONCLUSIONS:

To enhance the predictions made by our models, there are still a lot of things that can be explored. We develop and includenew variables, test various models with various subsets of features and/or rows, etc.Here are a few of the suggestions.Grid search can be used to train the XGBoost model so that its hyperparameters are optimised and its accuracy is increased.As was stated inthe EDA section, we can join applicants with one, two,

three, or more dependents to create a new feature.We may also visualise independent vs independentvariables to find further trends.Byutilising a better formula, which might additionally take interest rates into account. We can also determine the EMI. Ensemble modelling is a possibility (acombination of different models). Read more.

FUTURE ENHANCEMENTS:

In order to remove human involvement and increase job opportunities in the rapidly expanding IT sector of today, new technology must be developed and existing technology must be updated.productivity. Anyone looking to apply for a loan or the banking system will use this model. It is obvious from the analysis of the data that it lessens all frauds perpetrated during the loan approval process. Everyone values their time, and by doing this, the bank and the applicant's wait time will both be shortened.The processes in the prediction process include data cleaning andprocessing, imputation of missing values, experimental analysis of the data set, modelcreation, and testing on test data. The original data set's best-case accuracy is 0.9189 on Data set. The results of the data analysis showed that applicants with the lowest credit ratings will not be approved for loans since they are more likely to default on them.

REFERENCES:

[1]. Arun Kumar, Ishan Garg, and Sanmeer Kaur, \"Loan Approval Prediction Using Machine Learning Approach,\" 2018.

[2]. K. Hanumantha Rao, G. Srinivas, A. Damodhar, and M. Vikas Krishna at International Journal of Computer Science and Telecommunications published an article titled \"Implementation of Anomaly Detection Technique Using Machine Learning Algorithms\" (Volume2, Issue3,June 2011).

[3]. G. Arutjothi and C. Senthamarai,

[4]. \"Prediction of loan status in commercial banks using machine learning classifier,\" International Conference on Intelligent Sustainable Systems (ICISS), 2017.

[5]. \"AzureML based analysis and prediction of loan applicantscreditworthy,\" by Alshouiliy K, Alghamdi A, and Agrawal D P I n 2020, Third International conference on informationand computer technologies.

[6]. \"Developing prediction model of loan risk in banks using data mining Machine Learning and Applications,\" Hamid A J and Ahmed T M, 2016.

[7]. M. Li, A. Mickel, and S. Taylor \"Should this loan be approved or denied?\" published a paper in the Journal of Statistics Education in 2018. **[7]**A.Vinayagamoorthy, M. Somasundaram, and C. Sankar, \"Impact of Personal Loans Offered by Banks and Non-BankingFinancial Companies in Coimbatore City,\" 2012.

[8]. M. Cary Collins, Ph.D., and Frank M.Guess, Ph.D., MIT\'s Information Quality Conference, 2000, \"Improving information quality in loan approval processes for fairlending and fair pricing.\"

[9]. Arun Kumar, Ishan Garg, and Sanmeet Kaur, \"Loan approval prediction based on machine learning approach,\" National Conference on Recent Trends in Computer Science and Information Technology, 2016.

[10]. Sivasree M S and Rekha Sunny T, \"Loan Credibility Prediction System Using Decision Tree Algorithm,\" International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue 09,September-2015.

[11]. Jaroslav Belás, Zuzana Vincúrová, "Model of the loan process in the context ofunrealized income and loss prevention", Journal of International Studies, Vol. 8, No 1, 2015, pp. 91-106.

[12]. DOI: 10.14254/2071-8330.2015.

[13]. Gupta, Anshika, et al. "Bank Loan Prediction System using Machine Learning." 2020 9th International Conference System Modeling andAdvancement in Research Trends (SMART). IEEE, 2020.

[14]. Kumar, Arun, Garg Ishan, and Kaur Sanmeet. "Loan approval prediction based on machine learning approach." IOSR J.Comput. Eng 18.3, 18-21, 2016.

[15]. M. A. Sheikh, A. K. Goel and T. Kumar Machine Learning Algorithm," 2020 International Conference on Electronics and Sustainable Communication Systems(ICESC), pp. 490-494, 2020.

[16]. Supriya, P. Usha et al. "LoanPrediction by using Machine LearningModels." ,2019. Loan Approval Predictionusing Machine Learning Algorithms Approach. 2021 [Ebook].
Retrieved fromhttps://ijirt.org/master/publishedpaper/IJIR T151769_PAPER.pdf

[17]. Ndayisenga, Theoneste. Bank Loan Approval Prediction Using MachineLearning Techniques. Diss. 2021.

[18]. Tejaswini, J., et al. "Accurate loan approval prediction based on machine learning approach." Journal of Engineering Science vol. 11, no.4, pp. 523-532. 2020.

[19]. KUMAR, SOURAV. "LOANPREDICTION SYSTEM." 2021.

[20]. Nikhil Madane, Siddharth Nanda-LoanPrediction using Decision tree, Journal of the Gujrat Research History, Volume 21Issue 14s, December, 2019

[21]. Shrishti Srivastava, Ayush Garg, ArpitSehgal, Ashok kumar – Analysis and comparison of Loan Sanction Prediction Model using Python, International journal of computer science engineering.