Personal Loan prognosis using Machine learning Approach

SUPERVISIOR: M. Hima Bindu
Asst. Professor
Department of Computer Science and Information Technology
Institute of Aeronautical Engineering, Dundigal, Telangana
Mail: modi.himabindu@iare.ac.in

TEAM MEMBERS:E. Anvesh
Department of Computer Science and Information Technology
Institute of Aeronautical Engineering, Dundigal, Telangana
Mail: anvesh0604@gmail.com

M. Tejasvi
Department of Computer Science and Information Technology
Institute of Aeronautical Engineering, Dundigal, Telangana
Mail: madavitejasvi05@gmail.com

T. Sai Bhanu
Department of ComputerScience and Information Technology
Institute of Aeronautical Engineering, Dundigal, Telangana
Mail: teralasaibhanu19@gmail.com

Abstract: This paper talks about the important job of automating the process of checking if someone is eligible for a loan. The goal is to make the process more streamlined and efficient while using less time and resources. Using machine learning methods like Random Forest, Naive Bayes, and Voting Classifier, the study carefully looks at a number of factors, such as gender, education level, and the number of children, to see if the borrower is eligible for a loan. The training dataset is used to build and improve the models, and the testing dataset is used to see how well the models can predict who will be eligible for a loan. The Voting Classifier is the most accurate, with a score of 95.121%. It is followed by Random Forest, which got 69.105%, and Naive Bayes, which got 63.414%. This shows that ensemble learning methods, such as the Voting Classifier, are good at correctly guessing who will be eligible for a loan, which makes them a good choice for businesses and financial institutions. The paper's success in simplifying and improving the loan approval process shows that it could save a lot of time and effort in the decision-making process.

Index terms *– Random Forest, Naïve Bayes, Voting Classifier, Machine Learning, Prediction Loan.*

1. INTRODUCTION

Authorizing loans has become an important part of the banking industry and other financial companies' jobs in today's financial world. These companies get a lot of money from loans, and the interest rates they charge borrowers are meant to keep them profitable and reduce risk [1]. But to avoid possible losses, lenders must carefully check the reputation of loan seekers before agreeing to give them money. To do this, a full background check must be done, taking

into account things like gender, income, job situation, and more [2].

The creditworthiness rating is a very important part of getting a loan because it shows how confident the lender is in the borrower's ability to pay back the loan plus interest. In order to figure out how risky each loan application is, financial institutions use a structured process. One of the most important things that is looked at is the applicant's income, which shows how well they can pay back the loan [3]. Lenders also see a person's past jobs to figure out how stable they are and how much money they will have in the future [4].

Along with that, things like gender may also affect the loan pick. It is often against the law to treat people differently when it comes to getting a loan because of their gender. But lenders might still look at gender as part of a bigger plan to figure out how risky a person is. Studies have shown that men and women have different job opportunities and make different amounts of money [5]. This could have an impact on their image.

Lenders may also look at the applicant's credit records to see how they've handled money in the past, such as how they've paid back loans and used credit cards. A good credit past shows that a person is reliable and knows how to handle money well, which makes them more creditworthy [6]. On the other hand, a past of late payments or failures raises red flags and could mean that you are turned down for a loan or have to pay more in interest to make up for the risk [7].

In addition to financial factors, lenders may also look at non-financial factors, such as the borrower's general financial health and the reason for the loan. For example, people who want to borrow money to invest in things that might go up in value, like real estate or education, may be seen more positively than people who want to borrow money for personal spending [8]. Also, borrowers with a wide range of assets and a low percentage of debt to income are usually seen as less of a risk and may be able to get better loan terms [9].

Lenders use many tools and technologies, such as credit score models and risk assessment systems, to make the process of figuring out if someone is creditworthy easier. Lenders can make better choices when these advanced tools look at huge amounts of data to come up with credit scores and risk ratings [10]. It is important to make sure that these methods are fair and neutral, though, because they could unintentionally make social differences worse [11].

In conclusion, lending money is an important part of the financial sector, and lenders depend on careful trustworthiness checks to lower risk and make sure loans are paid back. Lenders can make smart choices that balance making money and being responsible with loans by looking at things like income, job status, credit history, and personal data. Using new technologies can also speed up the evaluation process and reduce any possible biases, which is good for both lenders and debtors in the long run.

2. LITERATURE SURVEY

Checking a person's reliability before approving a loan is an important job for banks to do to control risk and make sure their banking operations will last. Over the years, many methods and approaches, such

as basic statistical methods and advanced machine learning algorithms, have been used to guess how likely it is that a loan will be approved. This literature review looks at a number of studies that have used machine learning and statistical methods to help advance the area of loan prediction.

Turkson et al. [1] suggested using machine learning to guess how likely it is that a person will be able to pay back a loan. To teach their prediction model, they used a file with information about loan applicants, like their income, job situation, and credit background. The writers used algorithms like decision trees and support vector machines to sort loan applicants into groups based on whether they were creditworthy or not. Their results showed that machine learning methods can correctly predict how a loan will turn out.

Vaidya [2] looked into a statistical and predictive method for predicting loan acceptance using logistic regression. Logistic regression is a type of statistics that is often used for binary classification tasks, which means it can be used to guess whether a loan will be approved. The author used a collection of loan applicants' traits and past loan decisions to create a logistic regression model that can predict the likelihood of new applicants getting a loan. The study showed how important statistical modeling is for catching the uncertainty that comes with choices about loan acceptance.

Sheikh et al. [3] showed how to use machine learning methods to predict whether a loan will be approved. With the help of random forests and gradient boosting machines, they were able to make predictions using a set of old loan data. The authors showed that machine learning can correctly predict loan results and find factors that affect trustworthiness by looking at different aspects of loan candidates' demographics, financial situation, and work experience.

Pidikiti et al. [4] used machine learning models to study how to predict loans. They looked into how algorithms like k-nearest neighbors (KNN) and support vector machines (SVM) could be used to figure out how likely it was that a loan applicant would be approved based on their information. By looking at a dataset with information like salary, credit score, and loan amount, the authors showed how machine learning models can help lenders make smart decisions about whether to give loans.

Using K-means clustering, Kavitha [5] worked on putting loan candidates into groups based on their risk percentage. Clustering techniques are ways to learn without being watched that put together data points that are alike based on certain traits. The author of this study used K-means clustering to divide loan applications into different risk groups based on things like their income, credit score, and loan amount. Researchers learned how risk is spread among loan applicants, and the results helped lenders find high-risk candidates who needed more attention.

In conclusion, the research on loan forecast using machine learning and statistical methods shows that these methods are good at figuring out who is creditworthy and what will happen with loans. Researchers have used algorithms like decision trees, logistic regression, and clustering to look at a wide range of datasets and find useful information that can help them decide whether to approve a loan. The results of these studies show how important it is to use data-driven methods to make loans more efficient

and accurate while lowering risks for both lenders and consumers.

3. METHODOLOGY

i) Proposed Work:

Machine learning methods will be used in the suggested system to change the way loan application is checked. It will use Random Forest, Naive Bayes, and Voting Classifier to build a strong model that can look at application data like gender, school level, and children. The method will simplify the process of approving loans, which will cut down on human work and make things run more smoothly. Applicants will send in their information online, and the system will automatically figure out if they are qualified. The system will make sure that evaluations are accurate and fair by using advanced formulas. This will reduce the bias that often shows up in human evaluations. The method will also give people feedback in real time, which will make things more clear and improve the customer experience. Overall, the suggested method will improve the loan acceptance process, making it faster, more reliable, and less resource-intensive for both candidates and lenders.

ii) System Architecture:

There are seven main parts that make up the system architecture: 1) putting the loan dataset online, 2) cleaning it up, 3) making training and testing data, 4) running the Random Forest machine learning model, 5) running the Naïve Bayes machine learning model, 6) using the Voting Classifier machine learning model, and 7) figuring out who is eligible using the Voting Classifier model. It's possible that the system will also have a tool for making graphs of the data and model results. These parts work together to make it easier to guess who will be eligible for a loan based on different application and loan features.

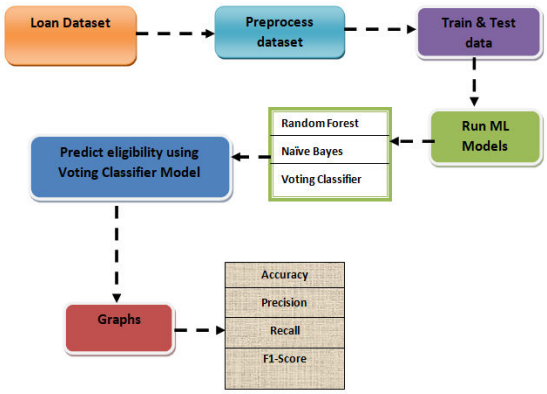

Fig 1 Proposed Architecture

iii) Modules:

1. Upload Loan Dataset: This feature lets users upload the loan dataset, which usually has details about people who are applying for loans and what they're like.

2. Preprocess Dataset: In this lesson, methods for preparing data are used to clean and get the dataset ready for analysis. Some tasks might be to deal with missing values, encode category variables, and scale number features.

3. Generate Train & Test Data: This tool divides the preprocessed dataset into training and testing sets so that the performance of machine learning models can be correctly evaluated.

4. Run Random Forest ML Model: This section uses the Random Forest[4] method to build a forecast model from the training data to figure out who is eligible for a loan.

5. Run Naïve Bayes ML Model: This module uses the Naïve Bayes[5] algorithm to make a predictive model that predicts the chances of getting a loan based on the characteristics of the dataset.

6. Run Voting Classifier ML Model: The Voting Classifier module takes results from several different models (like Random Forest and Naïve Bayes) and combines them to make a final choice about who can get a loan.

7. Predict Eligibility using Voting Classifier Model: This module guesses whether new applicants will be able to get a loan based on their characteristics using the Voting Classifier model that was learned on the dataset.

8. Graphs: This tool makes pictures, like confusion matrices or precision plots, to show how well the model worked and help you understand the results.

9. Exit: This button lets users leave the system and stop the program from running.

iv) Algorithms:

RF:Leo Breiman and Adele Cutler's random forest[4] is a prominent machine learning approach. It combines decision tree output into one result. Its ease of use and versatility make it popular for classification and regression.

NB: Note that Naive Bayes techniques are supervised learning algorithms that apply Bayes' theorem with the "naive" assumption that every pair of features is conditionally independent when the class variable is set to a given value.

Voting Classifier: A voting classifier is a type of ensemble learning that takes a vote to guess the class name from the results of several base estimators, which are machine learning models. It can be used for both regression and classification tasks.

4. EXPERIMENTAL SCREENS

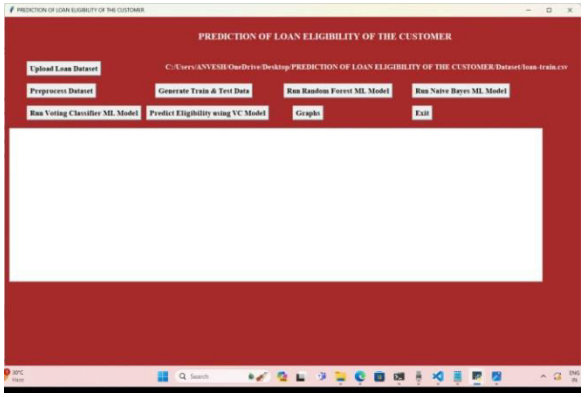

Fig 2 Home page

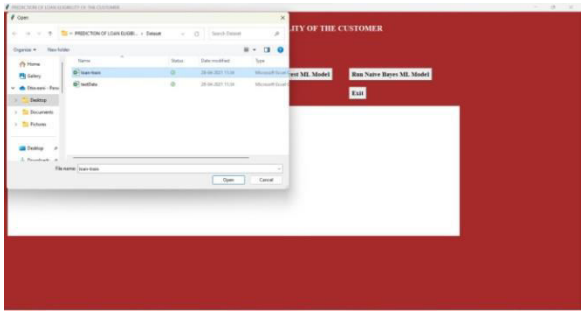

Fig 3 Upload loan dataset

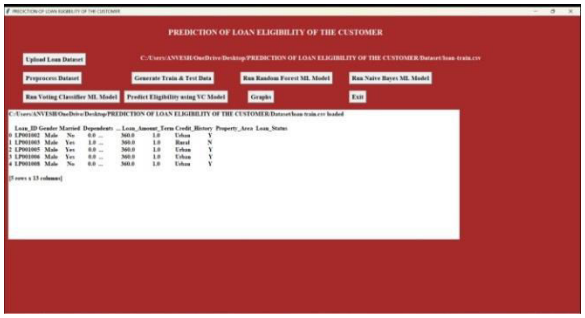

Fig 4 Preprocess dataset

Fig 5 Graph

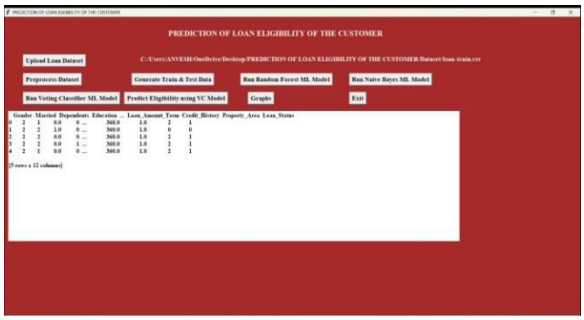

Fig 6 Dataset attributes

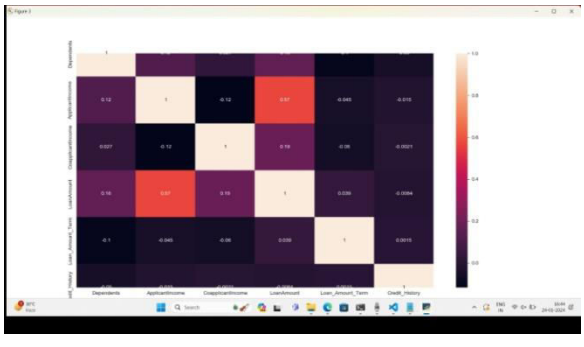

Fig 7 Correlation matrix

Fig 8 Generate train & test data

Fig 9 Random Forest

Fig 10 Naïve Bayes

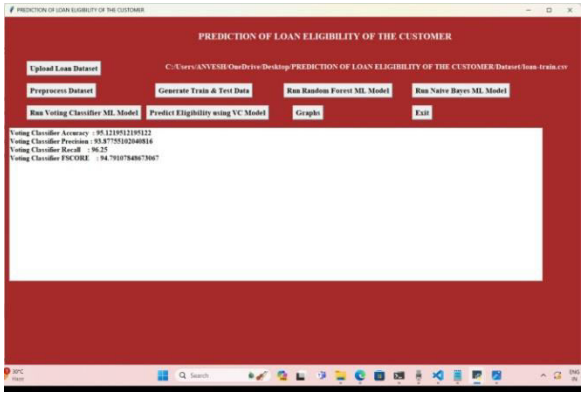

Fig 11 Voting Classifier

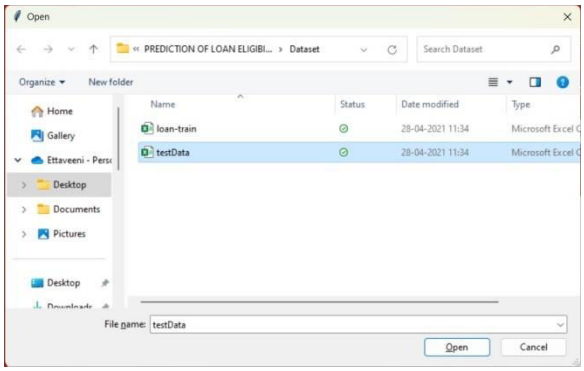

Fig 12 Output Screen

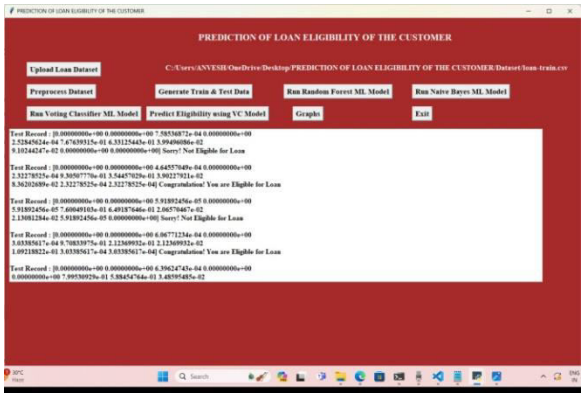

Fig 13 Test records

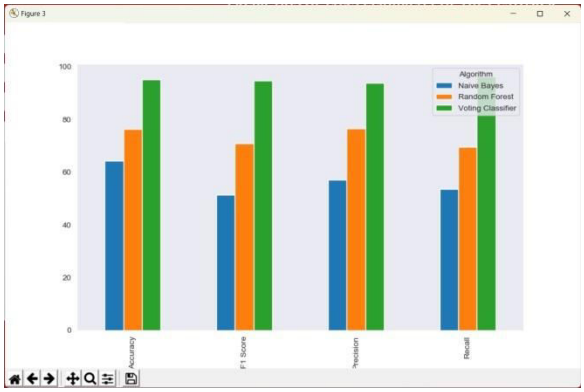

Fig 14 Graphs

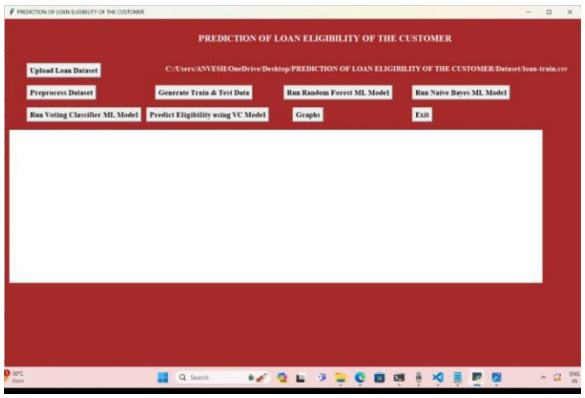

Fig 15 Exit

5. CONCLUSION

In conclusion, the suggested system is a big improvement over the current method of checking loan qualifications by hand. The system wants to change the way loans are given by using machine learning methods like Random Forest, Naive Bayes, and Voting Classifier. When the suggested and current systems are compared, the results make it clear that automation and advanced analytics are good ideas. The current system has many problems, such as subjective judgments caused by human involvement, processes that take too long, delays, and not enough information about how borrowers behave. On the other hand, there are some good things about the suggested scheme. The loan approval process is automated, which cuts down on human work and handling times and improves working efficiency. Also, using complex algorithms makes sure that evaluations are correct, reducing the flaws and errors that come with human judgment. The method also gives people comments in real time, which makes things more clear and improves the customer experience. Overall, the suggested system is a more reliable, efficient, and open way to check if someone is eligible for a loan. This makes it a valuable tool for both lenders and users in the financial world.

6. FUTURE SCOPE

To make the suggested system even better in the future, more advanced machine learning methods like gradient boosting and neural networks could be added to make predictions even more accurate. Adding other types of data, like social media accounts or transaction records, could also give you more information about how borrowers act. The system could also be improved so that it gives specific loan suggestions based on people's financial goals and risk levels. Using blockchain technology to make loan deals safe and clear could also be looked into. To make sure the system stays useful and current in today's changing financial world, it will need to be constantly watched and changed to keep up with new rules and market trends.

REFERENCES

[1]. Turkson, Regina Esi, Edward Yeallakuor Baagyere, and Gideon Evans Wenya. "A machine learning approach for predicting bank credit worthiness." 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR). IEEE, 2016.

[2]. Vaidya, Ashlesha. "Predictive and probabilistic approach using logistic regression: application to prediction of loan approval." 2017 8th International Conference on Computing, Communication, and Networking Technologies (ICCCNT). IEEE, 2017.

[3]. Sheikh, Mohammad Ahmad, Amit Kumar Goel, and Tapas Kumar. "An Approach for Prediction of Loan Approval using Machine Learning Algorithm." 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, 2020.

[4] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, K Vikas, “Loan Prediction by using Machine Learning Models”, published in International Journal of Engineering and Techniques, Available at http://www.ijetjournal.org/volume5/issue2/IJET-V5I2P28.pdf

[5] K. Kavitha, “Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques”, published in Computer Science, Business, Corpus ID: 168472569, 2016

[6] Chao-Ying Joanne Peng, Kuk Lida Lee and Gary M. Ingersoll, "An introduction to logistic Regression Analysis and Reporting", *The journal of Educational Research Indiana University-Bloomington*, vol. 96, pp. 1-13, September/October 2002.

[7] David H. Hosmer and Stanley Lemeshow, Applied Logistic Regression, John Wiley & Sons, pp. 47-135, 2004.

[8] Hyeoun-Ae, "An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain Park", *J Korean Acad Nurs.*, vol. 43, pp. 154-164, April 2013.

[9] Kevin P. Murphy, Machine Learning A Probabilistic Approach Massachusetts Institute of Technology, pp. 1-21, 2012.

[10] COKLUK Omay, "Logistic Regression: Concept and Application", *Kuram ve Uygulamada Egitim Bilimleri*, vol. 10, pp. 1397-1407, 2010.

[11] Stuart Russell and Peter Norvig, Artificial Intelligence -A mordern approach 1rst, Prentice Hall Inc., pp. 4-8, 1995.

[12] Vijayalakshmi Sampath, Andrew Flagel and Carolina Figueroa, "A Logistic Regression Model To Predict Freshmen Enrolments", *Northern Virginia Community College VA*, pp. 1-12, May 2016.

[13] "Logistic Regression in Python", *yhat- Machine Learning Data Science and Engineering blog*, March 2013.