Prediction of Loan Eligibility of the Customer

C. Khyathi[1], R. Sree Vidya[2], M. Manjula Bai[3], SH. Farhath Parveen[4]

[1,2,3,4]UG Scholors, Dept. of Computer Science and Engineering, RK College of Engineering,

Vijayawada, India

khyathichennadi@gmail.com[1]: sree92072@gmail.com[2]; meghavathumanjulamanju@gmail.com[3],

parveenfarhath3@gmail.com[4]

Abstract – **With the increase in banking sector many people are applying for loans in bank. All these loans are not approvable. The main income of bank assets comes from gain earned from loans. The main objective of banks is to invest their assets in safe customers. Today many banks approve loan after many processes of verification and validation but still there is no surety that selected customer is safe or not. Therefore, it is important to apply various techniques in banking sector for selecting a customer who pays loan on time. In this report we use random forest algorithm for the classification of data. Random forests algorithm builds a model from trained dataset and this model is applied on test data and we get the required output.**
*Keywords-***Loan Prediction, Risk Management, Data Analysis**

I.INTRODUCTION

Loan Prediction is very helpful for employees of banks as well as for the applicant also. The aim of this Paper is to provide a quick, immediate and easy way to choose the deserving applicants. Various Finance Company deals in all loans. They have presence across all urban, semi urban and rural areas. Customer first applies for loan after the company or bank validates the customer eligibility for loan. Company or bank wants to automate the loan eligibility process (real time) based on customer details provided while filling the application form [1].

These details are Gender, Marital Status, Education, Number of Dependents, Applicant Income, Loan Amount, Credit History and other [2]. This project has taken the data of previous customers of various banks to whom on a set of parameters loans were approved. So the machine learning model is trained on that record to get accurate results [3]. Our main objective of this project is to predict the safety of loan.

The banking sector has seen a significant increase in loan applications, but not all loans are approvable. Banks aim to invest in safe customers to minimize risk and maximize gains. To achieve this, banks use various techniques to assess creditworthiness. This report explores the use of Random Forest algorithm for loan approval, enabling banks to make informed decisions and reduce credit risk[4]. The Random Forest algorithm is a powerful machine learning technique that can accurately classify customers as safe or risky based on their credit profiles. By analyzing historical data and identifying patterns, the algorithm can predict the likelihood of a customer repaying their loan on time[5]. This approach can help banks streamline their loan approval process, reduce defaults, and improve overall profitability.

II. LITERATURE SURVEY

With the enhancement in the banking sector, lots of people are applying for bank loans, but the bank has limited assets that it can grant to a 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 [8-9]. So, in this paper,

⊕ *United International Journal of Engineering and Sciences* ⊕
(UIJES – A Peer-Reviewed Journal); ISSN:2582-5887 | Impact Factor:8.075(SJIF)
▧ *Volume 5 | Special Issue 1 | 2025 Edition*
National Level Conference on "Advanced Trends in Engineering
Science & Technology" – Organized by RKCE

we try to reduce this risk factor behind selecting the safe person to save lots of bank efforts and assets [10-11]. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before, and based on these records/experiences, the machine was trained using the machine learning model, which give the most accurate result. The main objective of this paper is to predict whether assigning the loan to a particular person will be safe or not. This paper is divided into four sections [6].

Nowadays, there are many risks related to bank loans, for the bank and for those who get the loans. The analysis of risk in bank loans need understanding what is the meaning of risk[8-9]. In addition, the number of transactions in banking sector is rapidly growing and huge data volumes are available which represent the customers behavior and the risks around loan are increased [10]. Data Mining is one of the most motivating and vital area of research with the aim of extracting information from tremendous amount of accumulated data sets. In this paper a new model for classifying loan risk in banking sector by using data mining. The model has been built using data form banking sector to predict the status of loans. Three algorithms have been used to build the proposed model: j48, bayes Net and naïve Bayes. By using Weka application, the model has been implemented and tested [11]. The results has been discussed and a full comparison between algorithms was conducted [12]. J48 was selected as best algorithm based on accuracy [13].

In recent years, banks and creditors face lots of challenges associated with the bank loan. In addition, banking transactions are growing rapidly and large data sets that represent customer behaviour and risks around debt burdens have increased. So, it became mandatory to know the risks related to bank loans. Each consumer knows what they want and what kind of business they would like to choose. The customer can simply find another bank if the administrators are not ready to pay attention. Many data analysis technique exists that focus on loan risk [14]. These techniques are oriented toward the extraction of quantity and statistical data properties [15]. They facilitate useful data explanations and help to gain better insights in the process behind data.While there can be traditional data analysis techniques lead to intuition knowledge, it is still made by human researchers. Banks and many investment companies are pioneers using data mining. The proposed work in data mining focused on using data from banking sector to predict the status of loans. Two classification algorithms Naïve Bayes and Support Vector Machine are discussed here. The results have been proven that the speed and accuracy has been improved when combined these two techniques.

III. METHODOLOGY

3.1 Data Collection

The data is saved as a csv file as Loan data.csv and it is read and stored in the life data variable. Data Analysis verify the documents and forwards the details to loan evaluator for approval or rejection. System approves the loan if documents are cleared and reject the loan if documents are not cleared Report is delivered to the applicant according to their status.

3.2 Data Pre-processing

Missing values inside the gathered data might lead to discrepancies. Data must be normalised in order to get better results from the algorithm. We need to eliminate the outliers and transform the variables in order to have a clearer picture. We utilise the chart feature to get to the bottom of these problems.

3.3 Modelling

Now we should train the model on the training dataset and test dataset. Thus it is chosen ML Models Here We are applied Machine learning algorithm applied. Such as Random Forest. The figure shows Figure 1: Workflow of Prediction of Loan Eligibility Using Random Forest.
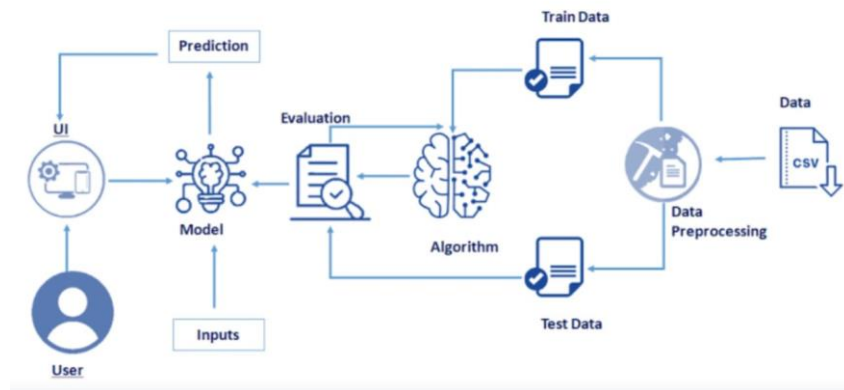
⊕ *United International Journal of Engineering and Sciences* ⊕
(UIJES – A Peer-Reviewed Journal); ISSN:2582-5887 | Impact Factor:8.075(SJIF)
▨Volume 5 | Special Issue 1 | 2025 Edition
National Level Conference on "Advanced Trends in Engineering
Science & Technology" – Organized by RKCE

Figure 1: Workflow of Prediction of Loan Eligibility Using Random Forest

3.4 Data Set Description

A loan dataset typically contains financial records and associated information used to determine loan eligibility, including factors like income, loan amount, credit history, and loan status, often used for machine learning and data analysis to predict loan approval[17].

Here's a more detailed breakdown of what you might find in a loan dataset:

Data Set Features:

- **Applicant Information:**

  - **Demographics:** Gender, marital status, education level, number of dependents, age.

  - **Employment**: Employment status (e.g., employed, self-employed), job title, years of experience.

  - **Income:** Applicant income, co-applicant income, debt-to-income ratio.

- **Loan Details:**

  - **Loan Amount:** Requested loan amount, loan term (duration), loan type (e.g., personal, business, mortgage).

  - **Credit History:** Credit score, credit history length, number of credit inquiries.

  - **Repayment:** Debt-to-income ratio, revolving balance (unpaid credit card balance).

- **Loan Status:**

  - **Outcome:** Loan approval status (approved/rejected), current loan status (current, charged-off, fully paid).

- **Repayment History:** Payment history, late payments, default status.

▢ **Other Relevant Information:**
  - **Property:** Property type (if applicable), property value.

  - **Location:** State/region, zip code.

  - **Homeownership:** Ownership status of the applicant's residence.

IV. RESULTS AND DISCUSSION

In the screen in square bracket, we can see normalised 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.
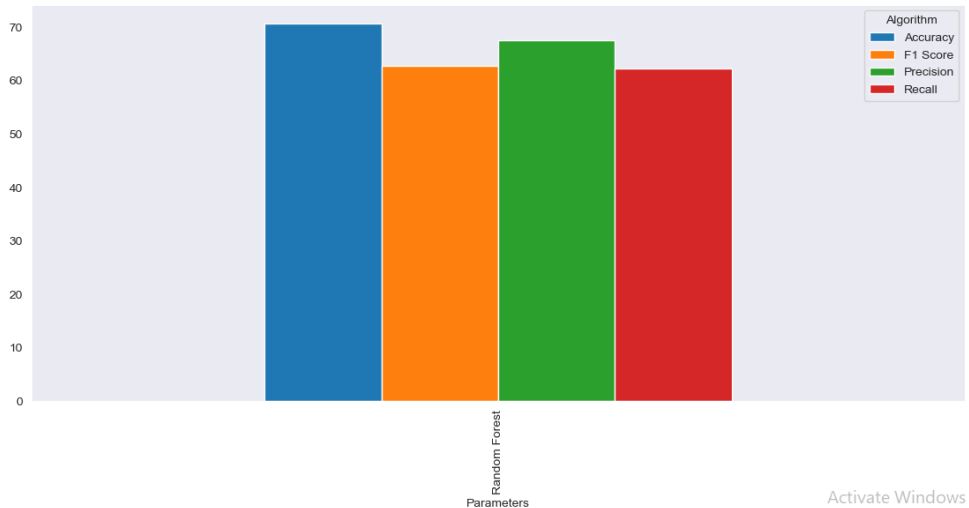
In above graph we can see accuracy, precision, recall and FSCORE values of random forest and graph y-axis represents %value where accuracy got 80% and Precision got 65%. Each metric bar colour name you can see from top right side.

V. CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

From a proper analysis of positive points and constraints on the component, it can be safely concluded that the product is a highly efficient component. This application is working properly and meeting to all Banker requirements. This component can be easily plugged in many other systems. In this we are using machine learning algorithm called Random Forest to predict loan eligibility. The prediction of loan eligibility using Random Forest has shown promising results. By analyzing customer data and credit profiles, the model can accurately classify customers as eligible or not eligible for a loan. The use of Random Forest algorithm has several benefits, including: High Accuracy: The model achieves high accuracy in predicting loan eligibility. Handling Complex Data: Random Forest can handle complex data with multiple features and interactions. Robustness: The model is robust to outliers and missing values.

5.2 Future Scope

The future scope for loan eligibility using Random Forest is vast and exciting. Here are some potential areas of exploration:

Potential Areas of Exploration

1. Integration with Other Machine Learning Algorithms: Combining Random Forest with other machine learning algorithms, such as neural networks or gradient boosting, to improve the accuracy and robustness of loan eligibility predictions.

2. Use of Alternative Data Sources: Incorporating alternative data sources, such as social media or online behavior, to improve the accuracy of loan eligibility predictions.

3. Development of More Advanced Models: Developing more advanced models, such as deep learning models, to improve the accuracy and robustness of loan eligibility predictions.

4. Explainability and Interpretability: Developing techniques to explain and interpret the predictions made by the Random Forest model, to improve transparency and trust in the loan eligibility decision-making process.

5. Real-time Processing: Developing real-time processing capabilities to enable faster and more efficient loan eligibility predictions.

Potential Applications

1. Automated Loan Approval: Using Random Forest to automate the loan approval process, reducing the need for manual review and improving the speed and efficiency of loan decisions.

2. Personalized Loan Offers: Using Random Forest to provide personalized loan offers to customers, based on their individual credit profiles and financial situations.

3. Risk Assessment: Using Random Forest to assess the risk of lending to customers, and to identify potential credit risks.

4. Portfolio Optimization: Using Random Forest to optimize loan portfolios, by identifying the most profitable and least risky loans.

Potential Benefits

1. Improved Accuracy: Random Forest can improve the accuracy of loan eligibility predictions, reducing the risk of lending to customers who may not be able to repay their loans.

2. Increased Efficiency: Random Forest can automate the loan approval process, reducing the need for manual review and improving the speed and efficiency of loan decisions.

3. Enhanced Customer Experience: Random Forest can provide personalized loan offers to customers, improving their overall experience and satisfaction with the lending process.

4. Reduced Risk: Random Forest can help lenders to identify potential credit risks, reducing the risk of lending to customers who may not be able to repay their loans.

VI.REFERENCES

[1] Kumar Arun, Garg Ishan, Kaur Sanmeet, ―Loan Approval Prediction based on Machine Learning Approach‖, IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 18, Issue 3, pp. 79-81, Ver. I (May-Jun. 2016).

[2] Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, ―Developing Prediction Model of Loan Risk in Banks using Data Mining‖, Machine Learning and Applications: An International Journal (MLAIJ), Vol.3, No.1, pp. 1-9, March 2016.

[3] S. Vimala, K.C. Sharmili, ―Prediction of Loan Risk using NB and Support Vector Machine‖, International Conference on Advancements in Computing Technologies (ICACT 2018), vol. 4, no. 2, pp. 110-113, 2018.

[4] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, kVikash,"Loan Prediction by using Machine Learning Models", International Journal of Engineering and Techniques. Volume 5 Issue 2, Mar-Apr 2019

[5] Nikhil Madane, Siddharth Nanda, "Loan Prediction using Decision tree", Journal of the Gujrat Research History, Volume 21 Issue 14s, December 2019.

[6] 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 DecisionMaking, Ministry of Education, Hefei University of Technology, Hefei 2009,

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

[8]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.

[9] Toby Segaran, "Programming Collective Intelligence: Building Smart Web 2.0 Applications." O'Reilly Media.

[10] Drew Conway and John Myles White," Machine Learning for Hackers: Case Studies and Algorithms to Get you Started," O'Reilly Media.

[11] Trevor Hastie, Robert Tibshirani, and Jerome Friedman,"The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer ,Kindle

[12] PhilHyo Jin Do ,Ho-Jin Choi, "Sentiment analysis of real-life situations using location, people and time as contextual features," International Conference on Big Data and Smart Computing (BIGCOMP), pp. 39–42. IEEE, 2015.

[13] Bing Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, May 2012.

[14] Bing Liu, "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions," Cambridge University Press, ISBN:978-1-107- 01789-4.

[15] 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.

[16] 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.

[17] 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.