Loan Approval Prediction by Analysing Customer Trends using Machine Learning

Pasumarthy Rambabu Asso. Prof, Saidulu Noorbhasha, Patsa Tejasai, Talluri Kalyan Satya Vara Prasad, Abbina Dileep Chowdary

Computer Science and Engineering

Department

Sasi Institute of Technology and

Engineering

Tadepalligudem, INDIA

Abstract: Loan is the amount of money sanctioned to an individual by considering an amount of interest on the principle. This can be done by considering various factors cibil score, credit purchase, income of the employee etc. In the recent years loan applications have seen a tremendous increase as many people are considering it various purposes. Financial institutions are maintaining a greater number of these applications and there should be reliable method to maintain the data in an efficient manner. Initially this loan applications approval relied on manual intervention which can be prone to human error, bias, so there is a need to provide ML decision making model that ensures fairness in lending practices. The primary objective of this study is to provide a efficient model using ML so that it can make Loan approval hassle free.

Keywords: Loan approval prediction, machine learning, customer trends

INTRODUCTION

Machine learning models is a good tool for loan approval prediction using various factors like income, credit status, since ML models can handle large datasets in a efficient way by understanding patterns between the data.AS in the banking industry loan approval is a crucial task as it involves risk approval for potential customers can create a social economical impact. The kay factors which are affecting loan approval are described below:

Demographic Factors: These factors include age, gender, martial status, if age is considered Younger applicants (people who age below 30) may have higher risk due to less credit history and lesser financial stability. The less credit score may be due not having enough time to have sufficient savings, added to that since they are early in their career income may not be stable, not having sufficient funds on an opposite way people whose age is grater than 35 – 60 are considered as stable because they are having longer credit histories, more consistent incomes. Married individuals are often considered more stable, and lenders are more interested in giving loans.

Income and Employment Status: Income is one of the primary resources for which may of loans gets approved or rejected Generally people with higher income levels trends to repay the loan in a timely manner. Even though the income is high. If the loan amount exceeds a borrower's financial capacity, the chances of approval decrease loan amount gets sanctioned accordingly., the interest associated with it also less as the risk is low Additionally if the loan getting person is an employee it ads an added advantage a fulltime job individual gets approval easily as his job is stable then risk is less on the contrary people with temporary or contract based, freelance job people have lower chances of approval due to their instability.

Credit History: It is often considered as most potential factor which decides loan, represents person's ability to repay loan in a timely manner, people who are paying timely have higher credit score, people with lower credit scores may get approval from lender but with a higher interest rates. If the person who want loan, then if the person has prolonged history of loan defaulting can severely affect their chances of approval.

Loan Characteristics: Purpose in which loan is taking for example loans for home and educational purposes are generally favoured whereas approval for risky ventures, business startups due to the uncertainty surrounding the applicant's ability to generate sufficient returns to repay the loan. While approving loan amount lenders approve by considering applicant's income and repayment ability. For instance, people who are not having higher income source even though applied for high amount of loan will be scrutinised thoroughly.

Financial Behaviour & Geographic factors: For accessing financial behaviour debt-to-income (DTI) ratio **is** considered which tells applicant's ability to manage additional debt it can be done by accessing applicants monthly debt payments (including mortgages, credit cards, and other loans) to their gross monthly income. Lower level of DTO indicates that handling can be done in a efficient manner. Geography has a major role for instance people staying in high rise apartments or residential colonies where the employment rate is high with a stable job market gets instant loan approval Conversely, applicants from areas with high unemployment rates or economic instability may face stricter evaluation criteria.

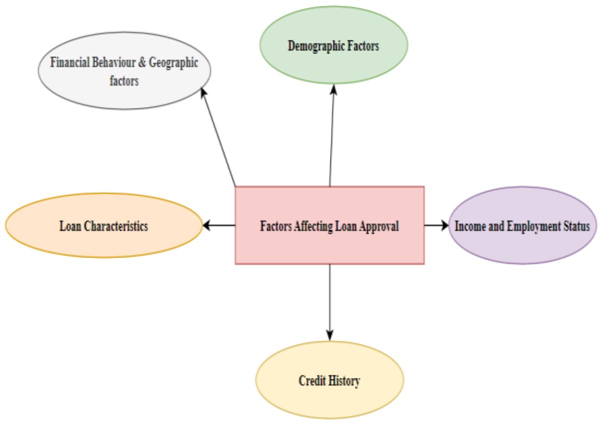


Fig.1. Various factors affecting loan approval

LITERATURE SURVEY

Viswanatha et al. [1] utilised ML model to predict loan approval an make the process much efficient way by considering various parameters like customer data, including demographic information, income details, and credit history these will help to train ML model then these were used by various ML models among them Random forests(RF) have performed better due to ensemble approach where decision of approval / not approval is taken in a good way and obtained an accuracy of 90%.

Disalwar et al. [2] proposed ML model to identify underlying trend between data parameters which makes a high efficient model by considering different ML models where probability based prediction, prediction based on the closest data points in the training set. Among them Gradient boosting achieved high accuracy due to combining multiple models and correcting error at each stage and obtained an accuracy of 92%.

Shaik et al. [3] made a comparative analysis using ML and neural networks to enhance efficiency in loan prediction by enhancing processing time where various factors such as age, gender, employment status, credit score, income, and the requested loan amount. among them the interrelation is identified by using feature extraction technique and ANN obtained an accuracy of 87%.

Athreyas et al. [4] made a comparative analysis of various ML models to predict genuine loan approval by rejecting defaulters who not met the eligibility criterial by considering various parameters particularly when dealing with structured financial data it is observed that Among all the compared model RF out performed due ensemble approach and obtained an accuracy of 93%.

Sharmila et al. [5] used various ML model to optimise loan approval process for fine tuning Decision Trees were considered as they ensemble approach which makes decisions best suitable for numerical data, also added to that feature selection where author used

Correlation Analysis to identify key features that have the most significant impact on loan approval and obtained an accuracy of 92%.

Lakhani et al. [6] used ML model to mitigate credit risk by predicting defaults, where parameters like credit score, age, gender, employability status, individual income of the individual as they are corelated the key features were extracted using feature extraction by using suitable feature selection techniques and obtained an accuracy of 90%.

Kumar et al. [7] utilised ML models to enhance loan approval prediction as of that oversampling is done in order to handle imbalance in the dataset where one class is significantly more represented than the other n loan approval datasets, for example, the number of approved loans (the majority class) may be much higher than the number of rejected loans (the minority class). This imbalance leads to reducing the model's ability to accurately predict the minority class outcomes. So this oversampling method helps in addressing this by representing with minority class so that both classes have similar proportions and obtained an accuracy of 94%.

Kumar et al. [8] utilised ML model and integrated them with chatbot to make automated loan approval system without human intervention making it more efficient and user-friendly. where parameters like credit scores, income, employment history, and existing debt were considered then in order to understand human queries the chatbot system gets trained with language model and obtained an accuracy of 90%.

AC et al. [9] made a comparative analysis of several ML models by considering data like income, credit score, loan amount, marital status, and employment history. then initially data normalisation to convert them into a similar scale after that models gets trained on the existing key features Among them RF obtained way better than other ML models of 94%.

Prasanth et al. [10] utile ml and proposed a RF algorithm which is a collection of multiple decision trees where the decision of previous tree is taken into consideration and gets corrected according as this is highly imbalance data better feature extraction is necessary to identify how features are corelated and obtained an accuracy of 91%.

Sam et al. [11] identified a relation bank customer churn with loan approval and then defined a ml model by considering various factors like age, income, credit score, loan history, and account activity. then initially handled missing values as they can affect the model prediction After that the categorical features are converted into numerical representations. then significant features were extracted by performing correlational analysis how they are interrelated can be drawn out by using these model gets trained, and RF obtained an accuracy of 90%.

Murenzi et al. [12] utilised data mining by considering various factors like demographics, financial status, and behaviours of loan applicants. then tried to identify which variable have good impact in making better prediction to make these categorical variables suitable for ML model they were encoded, and suitable features were extracted thi helped in reducing dimensionality among implanted models' logistic regression obtained an accuracy of 90%.

Lusinga et al. [13] utilised ML and developed a statis tical and predictive model by considering various factors like age, income, education), credit history, loan amounts requested, and repayment history since by using these feature identifying defaulters made easy then suitable data cleaning methods were considered after that by using correlation matrix important features were identified and given to model to train and evaluate and among them SVM obtained higher accuracy of 90% due to it support vectors which able distinguish defaulter vs good approval.

Deepa et al. [14] using ml based regression model developed a prediction system where both demographics data (Age, gender, marital status, education level), financial data(Annual income, credit score, existing debts), employment status(employed, unemployed, no of years) make the model diversified enough while evaluating and obtained an curacy of90%.

Meenakumari et al. [15] identified that may of the loan approval organisation are depending on data s=driven models so a accurate prediction model is necessary so that various factors were considered and feature selection techniques were used to extract key relevant features and model gets finetunes by optimising hyperparameters and obtained an accuracy of 96%.

METHODOLOGY

A. Dataset details

Loan-Approval-Prediction-Dataset: This dataset consists various collection of financial records where parameters like cibil score, income, employment status, loan term, loan amount, assets value, and loan status considered which is very suitable ML models to get trained on and to identify corelation between them.

B. Proposed Model

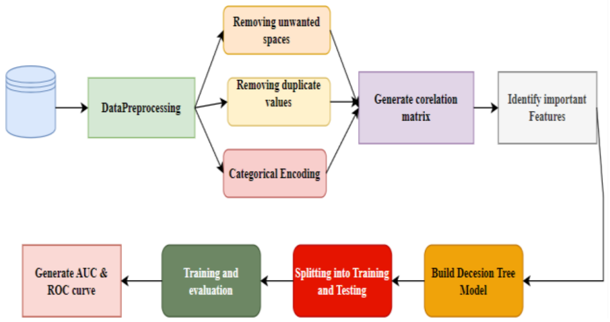


Fig.2. Propose model architecture

Here is a step by step implementation of the above algorithm

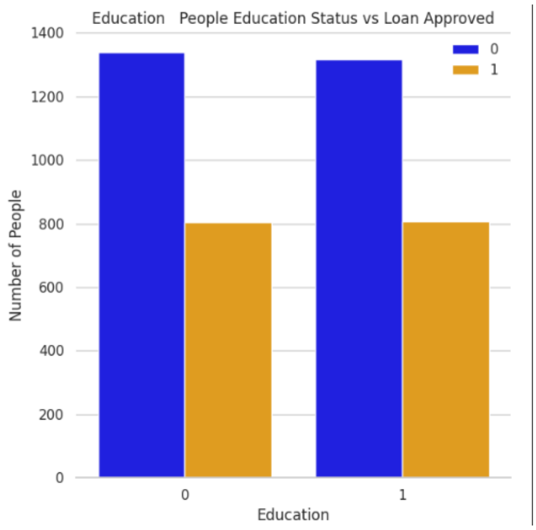


Fig.3. Education vs loan approval

The below Fig.4 represents amount of loan sanctioned with respect to applicant income people with higher income and good transactional history are often attracted by potential lenders with lower interest rates due to stable income and the risk associated with it is less while the

Step 1: It is necessary to install necessary libraries like NumPy, pandas, matplotlib, seaborn for visualisation purposes.

Step 2: Then as part of data preparation identified the columns and removed if there is any unwanted spaces then categorical encoding is used that converts categorical data into numerical values so that machine learning algorithms can use it

Step 3: A correlation matrix is generated which helps in visualizing what features are intercorrelated with each other through which relation can we drawn out which helps in better prediction.

Step 4: A pretrained decision tree model is built where number of nodes and trees are defined then the data is fed into this model.

Step 5: Then the model is trained on the dataset on a 80-20 scale split ratio then evaluation is made to predict.

Step 6: Finally AUC & ROC curve which measure the performance made by binary classification model.

low income people also gets loan sanctioned but the amount may vary also it is with higher interest rates

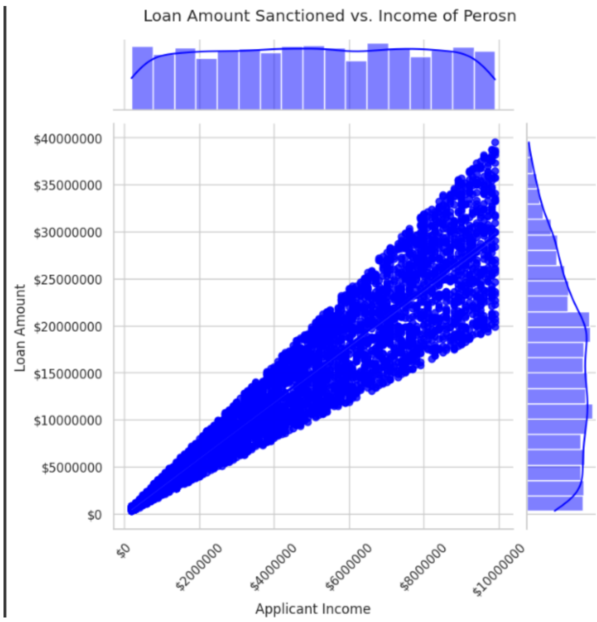


Fig.4. Loan amount sanctioned vs income of the person

C. Challenges Identified & Mitigating strategies

TABLE 1. A OVERVIEW OF CHALLENGES IDENTIFIED AND MITIGATING STRATEGIES

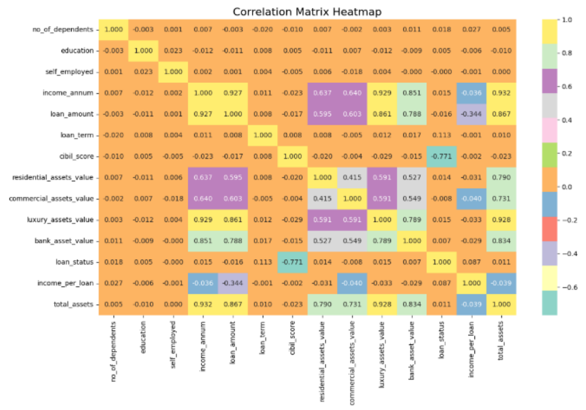
S.No	Author Name	Challenges Identified	Mitigating Strategies
1.	Sharma et al. (2022)	High dimensionality and non-linearity in data.	Utilized feed-forward neural networks and SGD optimization to improve predictions.
2.	Mousaeirad (2020)	Inadequate customer segmentation in the banking sector.	Implemented intelligent vector-based segmentation to enhance customer analysis.
3.	Costa (2022)	Data scarcity and complexity in credit risk evaluation.	Utilised ML odel to improve credit evaluation
4.	Neifaltas (2023)	Lack of effective prediction methods for credit rating.	Analyzed statistical and machine learning techniques for enhanced individual credit rating predictions.
5.	Nazareth et al. (2023)	Limited understanding of machine learning applications in finance.	Conducted a literature review on machine learning applications in financial contexts.
6.	Lee et al. (2006)	Complexity in customer credit assessment processes.	Used classification and regression tree models for improved credit mining.
7.	Yohanes (2024)	Challenges in customer segmentation in banking.	Applied machine learning techniques for effective customer segmentation.
8.	Moghe & Johri	Inefficient credit scoring methodologies.	Reviewed methodologies for credit scoring implementation in banking.
9.	Du & Guruprasad (2023)	Balancing data protection and model accuracy.	Investigated protection methods impacting machine learning model performance.
10.	Thirumagal et al. (2024)	Inefficiencies in loan decision-making processes.	Developed IoT-driven credit scoring models to enhance decision-making.

EXPERIMENTAL SETUP

This experiment requires python version of 3.8 with all the necessary libraries like numpy, pandas, matplotlib, Grayscale or binary images of different resolutions and complexities were used for testing. This will ensure a good environment makes this suitable for Loan approval prediction

RESULTS DISCUSSION

he results obtained demonstrate that machine learning algorithm have performed better especially decision trees due to their ensemble architecture which helped to correct mistake made at previous step in the current step It has seen good performance by understanding factors like credit score, transactional history, income of an individual, employability status, by analysing these below are the results obtained



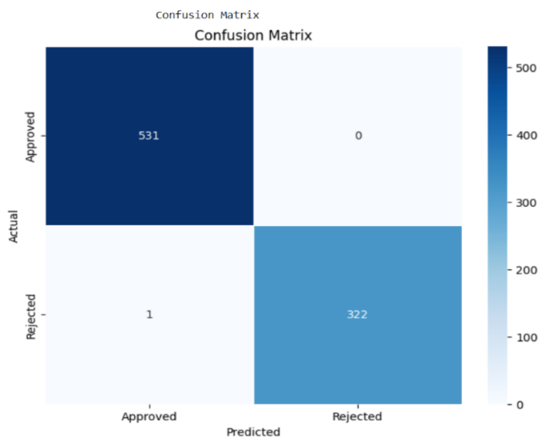


Fig.5. correlation matrix & Confusion matrix for loan approval

CONCLUSION & FUTURE SCOPE

Loan approval prediction is an important aspect which need to be addressed as the demand increases there is a proper system which need to access genuine loan approvals by rejecting loan defaulter profiles the propose model which is ML based decision tree utilise various preprocessing techniques like removing unwanted space, applying categorical encoding helped to get the model best fine tuned data available various visualisation wee illustrated in this paper by considering various parameters like education status vs loan approval, income of an individual vs the amount of loan sanctioned which helps to analyse what type of factors contribute to better loan approval with lesser interest rates also it is identified that credit score which is amount of transaction customer able to repat in a certain amount of time is considered as key factor. In the future there is a need to consider much diversified data where finetuning need to be done in a extensive manner

REFERENCES

[1] Viswanatha, V., Ramachandra, A. C., Vishwas, K. N., & Adithya, G. (2023). Prediction of loan approval in banks using machine learning approach. International Journal of Engineering and Management Research, 13(4), 7-19.

[2] Dosalwar, S., Kinkar, K., Sannat, R., & Pise, N. (2021). Analysis of loan availability using machine learning techniques. International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), 9(1), 15-20.

[3] Shaik, A., Asritha, K. S., Lahre, N., Joshua, B., & Harsha, V. S. (2022). Customer loan eligibility prediction using machine learning. Journal of Algebraic Statistics, 13(3), 2053-2062.

[4] Athreyas, S. S., Thanmai, B. K., Bairy, S. S., & Harish, K. (2022). A Comparative Study of Machine Learning Algorithms for Predicting Loan Default and Eligibility. Perspectives in Communication, Embedded-systems and Signal-processing-PiCES, 5(12), 116-118.

[5] Sharmila, S., Sandhya, P. V. S., Kousar, P. S., Anuradha, P., & Deekshitha, S. (2024, February). Bank Loan Approval using Machine Learning. In 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-5). IEEE.

[6] Lakhani, M., Dhotre, B., & Giri, S. (2019). Prediction of credit risks in lending bank loans using machine learning. SAARJ Journal on Banking & Insurance Research, 8(1), 55-61.

[7] Kumar, R., Prasad, P. D., Sangeetha, P., & Vinay, P. L. (2022). An Approach Based on a Machine Learning Algorithm for Predicting of Loan Approval. Dr. JMS VRaviKumar, P. DurgaPrasad, P. Sangeetha and P. LakshmanVinay. An Approach Based on a Machine Learning Algorithm for Predicting of Loan Approval. International Journal for Modern Trends in Science and Technology, 8, S03.

[8] Kumar, M. D. R., Maddikunta, R., Chella, A., Myadaraboina, R., & Kanchanapalli, S. MACHINE LEARNING APPROACHES FOR PREDICTING LOAN APPROVAL USING CHAT BOTS.

[9] AC, R., & KN, V. (2023). Prediction of Loan Approval in Banks Using Machine Learning Approach.

[10] Prasanth, C., Kumar, R. P., Rangesh, A., Sasmitha, N., & Dhiyanesh, B. (2023, March). Intelligent loan eligibility and approval system based on random forest algorithm using machine learning. In 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA) (pp. 84-88). IEEE.

[11] Sam, D., Suresh, K. C., Kanya, N., Tamilselvi, C., & Tejasria, M. V. S. L. (2021). An Improved Bank Customer Churn and Loan Prediction Model using Supervised Machine Learning Approach. Turkish Journal of Physiotherapy and Rehabilitation, 32(3).

[12] Murenzi, J. (2020). Predictive model for loan using data mining techniques case study: business development fund (Doctoral dissertation, University of Rwanda).

[13] Lusinga, M., Mokoena, T., Modupe, A., & Mariate, V. (2021, September). Investigating statistical and machine learning techniques to improve the credit approval process in developing countries. In 2021 IEEE AFRICON (pp. 1-6). IEEE.

[14] Deepa, M., Pal, S., & Ghusey, V. P. (2024, February). Monetary Loan Eligibility Prediction using Logistic Regression Algorithm. In 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE) (pp. 1-4). IEEE.

[15] Meenaakumari, M., Jayasuriya, P., Dhanraj, N., Sharma, S., Manoharan, G., & Tiwari, M. (2022, December). Loan eligibility prediction using machine learning based on personal information. In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 1383-1387). IEEE.

[16] Sharma, A., & Kumar, V. (2022, December). Predicting Loan Grant using Feed Forward Neural Nets and SGD optimizer. In 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 1568-1573). IEEE.

[17] Mousaeirad, S. (2020). Intelligent vector-based customer segmentation in the banking industry. arXiv preprint arXiv:2012.11876.

[18] Costa, P. A. G. (2022). Machine learning: challenges and opportunities on credit risk. PQDT-Global.

[19] Neifaltas, A. (2023). Analysis of statistical and machine learning techniques for predicting individual credit rating based on lithuanian data (Doctoral dissertation, Vilniaus universitetas.).

[20] Nazareth, N., & Reddy, Y. V. R. (2023). Financial applications of machine learning: A literature review. Expert Systems with Applications, 219, 119640.

[21] Lee, T. S., Chiu, C. C., Chou, Y. C., & Lu, C. J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. Computational Statistics & Data Analysis, 50(4), 1113-1130.

[22] Yohanes, Y. (2024). CBEBirr Customer Segmentation Using Machine Learning in Commercial Bank of Ethiopia (Doctoral dissertation, St. Mary's University).

[23] Moghe, M. S., & Johri, S. The Role of Credit Scoring in Modern Banking–An Overview of Methodology & Implementation.

[24] Du, J., & Guruprasad, S. (2023). Balancing Data Protection and Model Accuracy: An Investigation of Protection Methods on Machine Learning Model Performance for a Bank Marketing Dataset (Master's thesis).

[25] Thirumagal, P. G., Das, T., Das, S., Sikka, C. V., Kumar, C. A., & Kusuma, T. (2024, April). IoT-Driven Credit Scoring Models: Improving Loan Decision Making in Banking. In 2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST) (pp. 193-197). IEEE.