A REVIEW ON BANK LOAN APPROVAL PREDICTION USING DATA MINING

Sreedhar Ambala[1*], Naresh Bandam[2], Dr. A. Sai Hareesh[3], Dr. M Bal Raju[4]

Abstract:

With the upgrade in the financial area loads of individuals are applying for bank advances however the bank has its restricted resources which it needs to allow to restricted individuals just, so discovering to whom the credit can be conceded which will be a more secure choice for the bank is a commonplace interaction. So in this task we attempt to decrease this danger factor behind choosing the protected individual in order to save bunches of bank endeavors and resources. This is finished by mining the Data of the past records of individuals to whom the advance was conceded previously and based on these records/encounters the machine was prepared utilizing the AI model which give the most precise outcome. The principle objective of this paper is to anticipate whether relegating the advance to specific individual will be protected or not. This paper is separated into four areas (i)Data Collection (ii) Comparison of AI models on gathered information (iii) Training of framework on most encouraging model (iv) Testing.

Key Words: Support Vector Machine (SVM), Machine learning, Loan Prediction

[1*,2]Assistant Professor, Dept. of Computer Science Engineering, Pallavi Engineering College, Hayathnagar_Khalsa, Hyderabad, Telangana 501505

[3]Associate Professor, Dept. of Computer Science Engineering, Pallavi Engineering College, Hayathnagar_Khalsa, Hyderabad, Telangana 501505

[4]Professor Dept. of Computer Science Engineering, Pallavi Engineering College, Hayathnagar_Khalsa, Hyderabad, Telangana 501505

*Corresponding Author: Sreedhar Ambala

*Assistant Professor, Dept. of Computer Science Engineering, Pallavi Engineering College, Hayathnagar_Khalsa, Hyderabad, Telangana 501505

1. INTRODUCTION

This task has taken the information of past clients of different banks to whom on a bunch of boundaries credit were approved. So the AI model is prepared on that record to get precise outcomes. Our fundamental target of this exploration is to anticipate the security of credit . To foresee advance well being, the strategic relapse calculation is utilised. First the information is cleaned in order to stay away from the missing qualities in the informational collection. To prepare our model informational collection of 1500 cases and 10 mathematical and 8 downright characteristics has been taken. To credit an advance to client different boundaries like CIBIL Score (Credit History), Business Value, Assets of Customer and so forth has been thought of.

The financial sector, particularly the banking industry, thrives on effective risk management. Loan approvals are a crucial aspect of banking operations, and accurately assessing an applicant's creditworthiness is paramount. Traditionally, loan decisions have relied on manual evaluation of financial statements, credit history, and other factors. However, with the ever-increasing volume of loan applications and the growing complexity of financial data, data mining techniques offer a powerful approach to automate and enhance loan approval prediction.

1.2 What is Data Mining?

Data mining refers to the process of extracting hidden patterns and insights from large datasets. In the context of bank loan approval prediction, data mining involves analyzing historical loan data to identify key factors that influence loan repayment behavior. These factors, also known as features, can include:

- **Applicant Demographics:** Age, gender, marital status, dependents
- **Financial Information:** Income, employment status, debt-to-income ratio, credit score, loan amount, loan purpose
- **Credit History:** Delinquency history, defaults, bankruptcies

    By applying data mining algorithms to historical data, banks can build predictive models that estimate the probability of an applicant repaying a loan. These models can then be used to:

- **Automate Initial Screening:** Data mining models can efficiently evaluate a large number of loan applications based on predefined criteria, freeing up loan officers' time to focus on complex cases.
- **Improve Decision-Making:** Data mining provides objective insights that can supplement loan officers' subjective evaluations, leading to more informed decisions.
- **Reduce Risk:** By identifying loan applicants with a higher risk of default, banks can minimize potential losses and maintain a healthy loan portfolio.
- **Optimize Loan Terms:** Data mining can help tailor loan terms (e.g., interest rate, loan duration) based on an applicant's risk profile, leading to more competitive loan offerings.

1.3 Benefits of Data Mining for Loan Approval Prediction

- **Increased Efficiency:** Automating initial loan screening saves time and resources for loan officers.
- **Improved Accuracy:** Data mining models can identify complex relationships between variables that might be missed by traditional methods.
- **Reduced Risk:** Identifying high-risk applicants allows banks to make informed lending decisions.
- **Enhanced Customer Experience:** Faster loan processing and tailored loan offerings improve customer satisfaction.
- **Competitive Advantage:** Banks leveraging data mining can offer more competitive loan products and pricing.

1.4 Data Mining Techniques for Loan Approval Prediction

Several data mining techniques are well-suited for loan approval prediction. Here are some commonly used methods:

- **Logistic Regression:** This statistical technique estimates the probability of a binary outcome (loan approval or rejection) based on a set of independent variables (applicant features).
- **Decision Trees:** These tree-like structures classify loan applications based on a series of sequential decision rules derived from the data.
- **Support Vector Machines (SVMs):** SVMs create hyperplanes in a high-dimensional space to separate data points belonging to different classes (approved and rejected loans).
- **Neural Networks:** Inspired by the human brain, these interconnected nodes can learn complex non-linear relationships between features and loan approval.

The choice of data mining technique depends on the specific characteristics of the loan data and the desired outcome of the prediction model.

1.5 Challenges and Considerations

While data mining offers significant benefits for loan approval prediction, there are challenges to consider:

- **Data Quality:** The accuracy of the model heavily relies on the quality and completeness of historical loan data. Missing or inaccurate data can lead to unreliable predictions.
- **Model Bias:** Data mining models can inadvertently inherit biases present in the historical data. Mitigating bias is crucial to ensure fair and ethical lending practices.
- **Model Explainability:** Complex models like neural networks may be difficult to interpret, making it challenging to understand the rationale behind loan approval decisions. Explainable AI techniques can address this concern.
- **Regulatory Compliance:** Banks must comply with data privacy regulations when collecting, storing, and using loan applicant data.

Objectives of the Study:

The three main objectives of using data mining for bank loan approval prediction are:

1. **Reduce Risk:** This is the primary objective. By identifying applicants with a higher risk of defaulting on their loans, banks can minimize potential losses and maintain a healthy financial portfolio. Data mining helps unearth hidden patterns in historical data that traditional methods might miss, allowing banks to make more informed loan decisions.
2. **Improve Efficiency:** Automating the initial loan screening process with data mining models frees up valuable time and resources for loan officers. These models can quickly evaluate a large number of applications based on predefined criteria, allowing loan officers to focus on complex cases requiring more nuanced analysis.
3. **Enhance Customer Experience:** Data mining can contribute to a faster loan processing experience for customers. Additionally, the ability to tailor loan terms based on an applicant's risk profile allows banks to offer more competitive loan products. This combination of faster processing and personalized loan options leads to an improved customer experience.

2. LITERATURE REVIEW

An Approach for Prediction of Loan Approval using Machine Learning Algorithm Author: Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar

In our banking system, banks have many products to sell but main source of income of any banks is on its credit line. So, they can earn from interest of those loans which they credit. A bank's profit or a loss depends to a large extent on loans i.e., whether the customers are paying back the loan or defaulting. By predicting the loan defaulters, the bank can reduce its Non-Performing Assets. This makes the study of this phenomenon very important. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default. But as the right predictions are very important for the maximization of profits, it is essential to study the nature of the different methods and their comparison. A very important approach in predictive analytic is used to study the problem of predicting loan defaulters: The Logistic regression model. The data is collected from the Kaggle for studying and prediction. Logistic Regression models have been performed and the different measures of performances are computed. The models are compared on the basis of the performance measures such as sensitivity and specificity. The final results have shown that the model produce different results. Model is marginally better because it includes variables (personal attributes of customer like age, purpose, credit history, credit amount, credit duration, etc.) other than checking account information (which shows wealth of a customer) that should be taken into account to calculate the probability of default on loan correctly. Therefore, by using a logistic regression approach, the right customers to be targeted for granting loan can be easily detected by evaluating their likelihood of default on loan. The model concludes that a bank should not only target the rich customers for granting loan but it should assess the other attributes of a customer as well which play a very important part in credit granting decisions and predicting the loan defaulters.

Loan Default Forecasting using Data Mining Author: Bhoomi Patel1, Harshal Patil2, Jovita Hembram3, ShreeJaswal

Estimation or assessment of default on a debt is a crucial process that should be carried out by banks to help them to assess if a loan applicant can be a defaulter at a later phase so that they process the application and decide whether to approve the loan or not. The conclusion derived from such

assessments helps banks and other financial institutions to lessen their losses and eventually increase the number of credits. Hence, it becomes vital to construct a model that will take into account the different aspects of an applicant and derive a result regarding the concerned applicant. All available means to loan the money from their illicit activities are used for criminal activities in today's technology-based realm. The increasing number of bad debts resulting from commercial banks'loans reflects the growing problem of distraught banks within the economic system. We have used data mining algorithms to predict the likely defaulters from a datasets that contains information about home loan applications, thereby helping the banks for making better decisions in the future.

Prediction of Loan Status in Commercial Bank using Machine Learning Classifier Author: G. Arutjothi,Dr. C. Senthamarai

Banking Industry always needs a more accurate predictive modeling system for many issues. Predicting credit defaulters is a difficult task for the banking industry. The loan status is one of the quality indicators of the loan. It doesn't show everything immediately, but it is a first step of the loan lending process. The loan status is used for creating a credit scoring model. The credit scoring model is used for accurate analysis of credit data to find defaulters and valid customers. The objective of this paper is to create a credit scoring model for credit data. Various machine learning techniques are used to develop the financial credit scoring model. In this paper, we propose a machine learning classifier-based analysis model for credit data. We use the combination of Min-Max normalization and K Nearest Neighbor (K-NN) classifier. The objective is implemented using the software package R tool. This proposed model provides the important information with the highest accuracy. It is used to predict the loan status in commercial banks using machine learning classifier.

Overdue Prediction of Bank Loans Based on LSTM-SVM Author: : Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li

In the aspect of bank loans, the accuracy of traditional user loan risk prediction models, such as KNN, Bayesian,DNN, are not benefit from the data growth. This article is based on the work of Overdue Prediction of Bank Loans Based on Deep Neural Network. And we propose to analyze the dynamic behavior of users by LSTM algorithm, and use the SVM algorithm to analyze the user's static data to solve the current prediction

problems. This article uses users basic information, bank records, user browsing behavior, credit card billing records, and loan time information to evaluate whether users are delinquent. These static data are the basic input for SVM. For LSTM model, we extract user's recent transaction type from browsing behavior as input to LSTM, to predict the probability of users' overdue behavior Finally, we calculate the average of the two algorithms as the final result. From the experimental results, this LSTM-SVM model shows a great improvement than traditional algorithms.

Prediction Defaults for Networked-guarantee Loans Author: Dawei Cheng, ZhibinNiu†, Yi Tu and Liqing Zhang

Networked-guarantee loans may cause the systemic risk related concern of the government and banks in China. The prediction of default of enterprise loans is a typical extremely imbalanced prediction problem, and the networked-guarantee make this problem more difficult to solve. Since the guaranteed loan is a debt obligation promise, if one enterprise in the guarantee network falls into a financial crisis, the debt risk may spread like a virus across the guarantee network, even lead to a systemic financial crisis. In this paper, we propose an imbalanced network risk diffusion model to forecast the enterprise default risk in a short future. Positive weighted k-nearest neighbors (pwkNN) algorithm is developed for the stand-alone case – when there is no default contagious; then a data-driven default diffusion model is integrated to further improve the prediction accuracy. We perform the empirical study on a real-world three years loan record from a major commercial bank. The results show that our proposed method outperforms conventional credit risk methods in terms of AUC. In summary, our quantitative risk evaluation model shows promising prediction performance on real-world data, which could be useful to both regulators and stakeholders

Personal Credit Rating Using Artificial Intelligence Technology for the National Student Loans Author: Jian HU, Zibo, China

National student loans have the general features of commercial loans, and are a financial credit services provided by commercial banks. But the general personal credit rating assessment system of commercial bank can not make the correct credit rating because the lender, college students, have no credit history. To avoid the credit risk, a rational credit assessment system must to be established for college Students. With the self-

learning, self-organizing, adaptive and nonlinear dynamic handling characteristics of Artificial Neural Network, a Back Propagation neural network was developed to evaluate the credit rating about a college student. Several samples, which were provided by a bank, were used for network training and testing by MATLAB. The maximum value of the error between the prediction value of the network and actual value is only 2.92that the algorithm developed is fairly efficient for the assessment about the college student's personal credit situation

3. SYSTEM ARCHITECTURE

At first the client login's and transfer the information.. Furthermore, that information is utilized as preparing datasets. From that point onward, information pre-preparing is performed on that datasets. Further this pre-processed information goes for feature extraction where the recognizable proof of significant qualities of information is done and chosen. At that point we characterize that
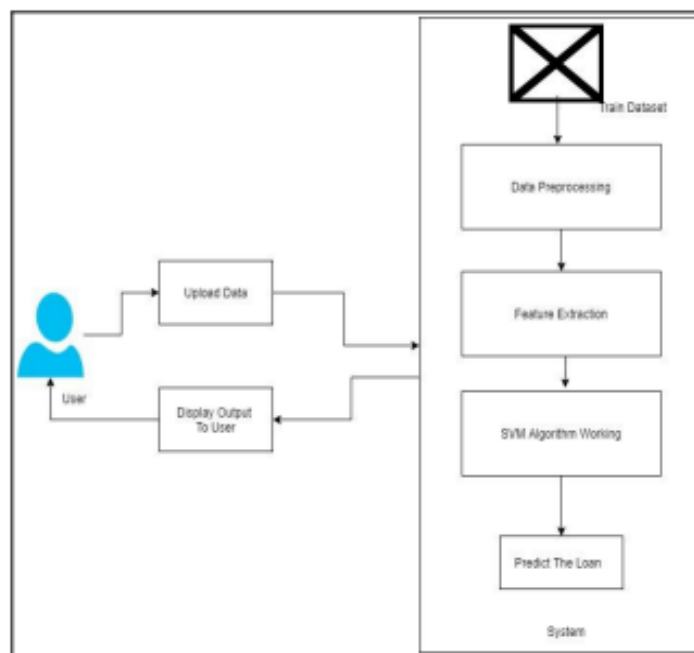

Fig: Data Mining System

CONCLUSIONS

The interaction of expectation begins from cleaning and handling of information, attribution of missing qualities, exploratory investigation of informational collection and afterward model structure to assessment of model and testing on test information. On Data set, the best-case precision acquired on the first informational collection is 0.811. The accompanying ends are reached after examination that those candidates whose credit score rating was most noticeably awful will neglect to get advance endorsement, because of a higher likelihood of not repaying the credit sum. More often than not, those candidates who have top level salary and requests for lower measure of advance are bound to get affirmed which bodes well, bound to take care of their credits. Some other trademark like gender and conjugal status appears to be not to be mulled over by the organization.

Data mining offers a powerful approach to predict loan approvals and improve loan decision-making in the banking industry. By leveraging data-driven insights, banks can achieve operational efficiency, mitigate risk, and ultimately, enhance customer satisfaction. However, addressing data quality, bias, explainability, and regulatory compliance is essential for responsible and effective implementation of data mining in loan approval prediction.

The future of data mining in loan approval prediction is promising. As data collection and processing capabilities continue to evolve, we can expect even more sophisticated models that incorporate alternative data sources (e.g., social media data) and leverage advanced techniques like deep learning. Additionally, continuous research efforts are focused on developing more interpretable models and ensuring fairness and compliance within the loan approval process. By embracing these advancements, banks can further optimize their lending practices and create a win-win situation for both the institution and its borrowers.

REFERENCES

1. Phil Hyo Jin Do, Ho-Jin Choi, "Sentiment analysis of reallife situations using loca- tion, people and time as contextual features," International Conference on Big Data and Smart Computing (BIGCOMP), pp. 39–42. IEEE, 2015.

2. Bing Liu, "Sentiment Analysis and Opinion Mining," Morgan Claypool Publishers, May 2012. [3] Bing Liu, "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions," Cambridge University Press, ISBN:978-1-107-01789-4.

3. 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.

4. 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.

5. 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.

6. 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