

LITERATURE SURVEY-SMART LENDER-APPLICANT CREDIBILITY PREDICTION FOR LOAN APPROVAL

INTRODUCTION:

With the enhancement in the financial sector, a number of individuals are applying for loans. However the bank has only limited resources and assets, so discovering to whom the credit can be conceded would definitely be a more secure choice for the bank thus preventing the bank from incurring losses. So in this task we attempt to decrease this danger factor behind by choosing the protected individual in order to save bunches of bank endeavors and resources. Here we implement ML models to automate the process of finding out if an individual is reliable or not.

LITERATURE REVIEW:

This section discusses in brief about some of the work that has already been done on creating ML and DL models using various algorithms to improve the loan prediction process and help the banking authorities and financial firms select an eligible candidate with very low credit risk. In recent years, banking authorities have been focused on applying machine learning algorithms and neural networks for credit scoring and risk assessment. Many noteworthy conclusions have been drawn in this regard which serve as stepping-stones for researches and studies.

1) **Paper Name:** Loan Prediction Using Ensemble Technique[1]

Authors: Anchal Goyal, Ranpreet Kaur

Description: Eleven machine learning models with nine properties are built in the proposed work to predict the credit risk of customers who have applied for a loan. This paper presented an ensemble model for loan predictions using several parameters such as Accuracy, Gini, AUC, Roc, and others to compare different training algorithms. The main goal of this paper is to evaluate the accuracy of models and to create an ensemble model that combines the outputs of three different models to predict customer loan amounts. The feature importance is calculated using Real Coded Genetic Algorithms. These features aid in predicting a customer's credit risk. The K-fold validation method is used to determine the predictive model's robustness.

Results: Maximum accuracy achieved: 81.25%

Model: Tree model for genetic algorithm

2) **Paper Name:** "Overdue Prediction of Bank Loans Based on LSTM-SVM"

Authors: Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li

Description: This article used 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, they extracted user's recent transaction type from browsing behavior as input to LSTM, to predict the probability of users' overdue behavior. Finally, they calculated 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.[2]

Results:LSTM-SVM model, after combining user dynamic data and static data, greatly improves the accuracy of our prediction and achieves 99% accuracy.

Model:Combine LSTM and SVM

3)**Paper Name:** Predicting Loan Approval of Bank Direct Marketing Data Using Ensemble Machine Learning Algorithms.

Authors: H. Meshref,

Description: In their research,they deployed various ensemble ML techniques such as AdaBoost, LogitBoost, Bagging, and Random Forest model to predict loan approval of bank direct marketing data. When compared to our study, the SMOTE technique we utilized to balance our dataset proved to be the key difference as our models achieved better performance.[3]

Results: Their research result showed that AdaBoost had the highest accuracy of 83.97%.

Model:Ensemble Methods

4)**Paper Name:** Loan prediction analysis using decision tree

Authors: Nikhil Madane and Siddharth Nanda

Description: Nikhil Madane and Siddharth Nanda in this paper reviewed credit scoring of mortgage loans and made the following conclusions: Credit applications that do not pass certain requirements are often not accepted because the probability of them not paying back is high. Low-income applicants are more likely to get approval, and they are more likely to pay back their loans in time.[4]

5) **Paper Name:** An Exploratory Data Analysis for Loan Prediction Based on Nature of the Clients

Authors: X.Francis Jency, V.P.Sumathi, Janani Shiva Sri

Description:The authors in this paper did an exploratory data analysis. The paper's main purpose was to classify and examine the nature of loan applicants. Seven different graphs were plotted and visualized and using these graphs the authors concluded that most loan applicants preferred short-term loans.[5]

6) In 2019, S. Z. H. Shoumo, M. I. M. Dhruva, S. Hossain, N. H. Ghani, H. Arif and S. Islam devised a machine learning model which helps in credit risk assessment system by showing great promise in identifying loan defaulters. The authors have also stated a comparative analysis of various machine learning algorithms and concluded that among all the models, the combination of tuned Support Vector Machine and Recursive Feature Elimination with Cross- Validation has shown more accurate results. The paper also introduces the five step proposed model by the authors. The support vector machines can outperform other tree-based models or regression models. Furthermore, in the debate over which dimensionality reduction technique to use, their model has shown us that recursive feature elimination with cross-validation can outperform other models[6]

7) In 2011, Lops, Pasquale & de Gemmis, Marco & Semeraro, Giovanni & Narducci, Fedelucio & Musto, Cataldo proposed how social networks like LinkedIn can be used to gather information about the users. The paper proposes an idea where user profiles can be made and based on their interests and specialties research papers can be recommended to the users. Various methods for recommendation of research papers are given and one of such methods is that researchers from the same area of field/interest tend to read the same research articles. The authors have used the LinkedIn extractor system of the users and their connections and found that this group of people tend to have similar interests. This same thing can be used in loan prediction where people in the connection can help to find whether the customer will default or not [7].

REFERENCES

- [1]Anchal Goyal , Ranpreet Kaur-“Loan Prediction Using Ensemble Technique”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016
- [2] Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li “Overdue Prediction of Bank Loans Based on LSTM-SVM”Jiangsu Key Lab of Big Data and Security and Intelligent Processing Nanjing University of Posts and Telecommunications, Nanjing, 210023, China
- [3] H. Meshref, "Predicting Loan Approval of Bank Direct Marketing Data Using Ensemble Machine Learning Algorithms."International Journal of circuits, systems, and signal processing. 2020, Vol. 14, pp. 914-922 DOI: 10.46300/9106.2020.14.117
- [4] Madane N and Nanda S 2019 Loan prediction analysis using decision tree Journal of The Gujarat Research Society 21 p p 214–21
- [5] Jency X F, Sumathi V P and Sri J S 2018 An exploratory data analysis for loan prediction based on nature of the clients Int. Journal of Recent Technol. and Engineering (IJRTE) 7 pp 176–9
- [6] S. Z. H. Shoumo, M. I. M. Dhruva, S. Hossain, N. H. Ghani, H. Arif and S. Islam, "Application of Machine Learning in Credit Risk Assessment: A Prelude to Smart Banking," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 2023-2028, doi: 10.1109/TENCON.2019.8929527.
- [7] Lops, Pasquale & de Gemmis, Marco & Semeraro, Giovanni & Narducci, Fedelucio & Musto, Cataldo. (2011). Leveraging the LinkedIn social network data for extracting content-based user profiles. RecSys'11 - Proceedings of the 5th ACM Conference on Recommender Systems. 293-296. 10.1145/2043932.2043986.