

Smart Lender - Applicant Credibility Prediction for Loan Approval

Industry Mentor(s) Name: Sowjanya, Sandeep Doodigani

Faculty Mentor(s) Name: Dr. Ponsy R K Sathia Bhama

Team Members: Bhavani Venkata Karthik K (2019503511)

Amaresh Saddish (2019503004)

Tanooj Cheekati (2019503056)

Shreyas Karthik Ramesh (2019503050)

Literature Survey

Introduction

Big Data's main advantage is its ability to analyse a set of data in order to draw out valuable information. Volume, Velocity, and Variety are the attributes used to describe the data collected by various organisations and businesses. Sometimes, the precision and accuracy of the data depend not only on how accurate the data is but also on when it is delivered. Big data has the greatest potential in the financial industry. These data can be quickly transformed into decision-making tools with added value by using efficient analysis techniques.

The answer to the query of how to analyse this data is machine learning. Artificial intelligence offers features that cost less in terms of both time and resources when used. Additionally, it provides novel classification methods and new insights into data. Machine learning is currently being actively used in banking and financial institutions for a variety of purposes, including trading, portfolio management, fraud detection, and loan approval. Artificial intelligence is used in the financial industry to create "robo-advisors," intelligent portfolio systems that can adjust to the risks and objectives of the user. To make decisions very quickly, decision trading support systems, also known as algorithmic trading, are used. Machine learning has brought promising new approaches to analyse user behaviour and

spot fraudulent transactions for fraud detection. Loan providers, especially big businesses like banks and insurance companies, train machine learning algorithms using a tonne of consumer data and financial lending outcomes. which can be used to decide wisely when it comes to insurance and lending. In this study, we focus on the latter: the decision-making process for loan approval.

According to a data analysis perspective, determining whether to approve a loan is a "good" or "bad" risk depends on the analysis and classification of a set of loan applicant data. A set of data is divided into two classes using the binary classification method, which is a straightforward classification case. When we want to predict a specific outcome that can only take two distinct values, we typically use this type of classification. Typical examples include detecting spam, diagnosing illnesses, looking for credit card fraud, or in our case, approving loans. Binary classification is a very basic problem, despite being fairly straightforward. Different paradigms are applied when learning binary classifiers, such as: Support Vector Machines, Decision Trees, K-nearest neighbourhood, Bayesian Classification, Logistic regression, Neural Networks, and more recently, deep learning.

Literature Review

[1] Using various classification algorithms, a machine learning model was constructed using the historical data of the candidates. The main goal of this study is to use machine learning models trained on historical data to predict whether a new applicant will be approved for a loan or not.

Advantages: Able to take a wide range of inputs.

Disadvantages: Not ideal for special circumstances such as an economic depression or bankruptcy situation.

Algorithms: XGBoost, Random Forest

[2] With the help of well-known algorithms like Decision Tree, Logistic Regression, and Random Forest, the main goal of this project is to predict which of the customers will have their loan paid off or not. To ascertain and comprehend how loan systems function in order to predict future loans using demographic data from various factors that together make up the nature of approval using machine learning algorithms and concepts, and ultimately to deploy this model on cloud-based platforms.

Advantages: Using this method, the bank can easily separate the necessary information from huge amounts of informational collections, which helps in accurate advance forecasting and reduces the number of bad credit issues.

Disadvantages: Accuracy is mediocre

Algorithms: Decision Tree, Logistic Regression, Random Forest

[3] In this study, machine learning (ML) algorithms are used to retrieve patterns in predicting future loan defaulters from a common loan approved dataset. The analysis will be done using historical customer data, including their age, income, loan amount, and length of employment. Several ML algorithms, including Random Forest, Support Vector Machine, K-Nearest Neighbor, and Logistic Regression, were used to identify the most pertinent features, or those that had the greatest influence on the prediction outcome.

Algorithms: Random Forest, Support Vector Machine, K-Nearest Neighbor, and Logistic Regression

[4] Swindling tools by combining the CatBoost algorithm with a document verification module that uses Tesseract and Camelot as well as a personalised loan module, financial institutions can reduce the risk associated with providing loans to defaulters and unapproved customers.

Advantages: High accuracy

Algorithms: CatBoost

[5] In this study, logistic regression is used as a tool to forecast a borrower's eligibility for a loan. Kaggle data is gathered for analysis and forecasting.

Algorithms: Logistic Regression

[6] The goal of this work was to create a decision tree-based high performance predictive model for loan approval prediction. Experiments were conducted using a variety of tree methods, from the simplest and most understandable decision tree to the most intricate random forests. Due to the highly correlated and complex feature space, which resulted in an overly simplified tree that was impractical to use, the results showed insufficient performance

with respect to simplified decision trees. But boosting won out in terms of effectiveness, applicability, and interpretation. As a result, it was advised to use a boosting-based decision-tree predictive model to make decisions about the eligibility of loan applicants based on those applicants' characteristics.

Advantages: Decision making with respect to the most important factors.

Algorithms: Decision Tree, Bagging, Boosting, Random Forest

[7] In this project, the model is built to reject risk factors, computational time, disorganised registration documents, and cross-verification of all edge communications in order to make decisions to sanction loans from lawyers, bank managers, and legal perspectives, and to approve loans quickly within the allotted time using a decision tree approach designed with lawyers in mind.

Algorithms: Decision Tree

[8] The main objective of this research is to analyse a bank's lending approval system to improve the total efficiency of the bank's lending process by using the Fuzzy Miner algorithm technique. The data included complex processes that were related to loan requests of the clients of the bank. the system has received a loan request. The primary motivation of the study was based on several complaints regarding handling the request cases with delays. The use of the Fuzzy Miner technique helped recognize and identify the bottleneck areas and investigate the resources causing the delays. The study results can lead to the development of the bank's lending systems by reducing problems that cause delays in the Bank's operations.

Algorithms: Fuzzy Miner

References

[1] V. Singh, A. Yadav, R. Awasthi and G. N. Partheeban, "Prediction of Modernized Loan Approval System Based on Machine Learning Approach," *2021 International Conference on Intelligent Technologies (CONIT)*, 2021, pp. 1-4, doi: 10.1109/CONIT51480.2021.9498475.

[2] H. Ramachandra, G. Balaraju, R. Divyashree and H. Patil, "Design and Simulation of Loan Approval Prediction Model using AWS Platform," *2021*

International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 53-56, doi: 10.1109/ESCI50559.2021.9397049.

[3] P. Tumuluru, L. R. Burra, M. Loukya, S. Bhavana, H. M. H. CSaiBaba and N. Sunanda, "Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms," 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2022, pp. 349-353, doi: 10.1109/ICAIS53314.2022.9742800.

[4] S. Barua, D. Gavandi, P. Sangle, L. Shinde and J. Ramteke, "Swindle: Predicting the Probability of Loan Defaults using CatBoost Algorithm," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1710-1715, doi: 10.1109/ICCMC51019.2021.9418277.

[5] V. Singh, A. Yadav, R. Awasthi and G. N. Partheeban, "Prediction of Modernized Loan Approval System Based on Machine Learning Approach," 2021 *International Conference on Intelligent Technologies (CONIT)*, 2021, pp. 1-4, doi: 10.1109/CONIT51480.2021.9498475.

[6] M. Alaradi and S. Hilal, "Tree-Based Methods for Loan Approval," 2020 *International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, 2020, pp. 1-6, doi: 10.1109/ICDABI51230.2020.9325614.

[7] D. P. Rajesh, M. Alam, M. Tahernezehadi, C. Vikram and P. N. Phaneendra, "Real Time Data Science Decision Tree Approach to Approve Bank Loan from Lawyer's Perspective," 2020 *19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2020, pp. 921-929, doi: 10.1109/ICMLA51294.2020.00150.

[8] K. Sirisong, P. Palangsantikul, P. Arpasat, S. Intarasema and S. Tumswadi, "Analysis of a Bank's Lending Approval System using Process Mining," 2021 19th International Conference on ICT and Knowledge Engineering (ICT&KE), 2021, pp. 1-4, doi: 10.1109/ICTKE52386.2021.9665411.