ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR292

Loan Approval Prediction System Using Machina Learning

P V V S V Prasad [1] ; P.V Nageswara Rao[2]
[1, 2]Departement of Computer Science and Engineering, GITAM School of Technology,
GITAM (Deemed to be University)Visakhapatnam, Andhra Pradesh, India

Abstract:- With the growth of the banking sector, the identification of reliable borrowers that must maintain the natural core income and asset accumulation becomes a key issue. Despite all security measures, reliability of customers remains an unclear question. To tackle this barrier, banking management which is directed towards customer loan repayment consistency is required. Credit approval is a significant aspect of economy since it determines the allocation of credit-linked funds. Today, machine learning is known for its power to automate and scale up the processing of application for loans.

This project will begin with data collection which will consist of data on historical (regarding the past) loan applications and the borrower profiles. The dataset has features of the credit score, income, previous work experience, debt-to-income ratio, and loan repayment record. This way, the models learn through the strengths to find good features and the reasons for accepting a loan. They are the experts in these areas and can forecast the potential patterns and connections of the data. Within the scope of this work, the supervised algorithms used are logistic regression, decision trees, random forests, and support vector machines. These algorithms are applied to the dataset available to often produce results like binary classification and regression. The adoption of machine learning among financial institutions is intended for a faster processing of loans which is their benefit. What is credit scoring, it is a tool which automate manual loan application review thereby increases efficiency. The machine-learning algorithms that analyse applications for loans could cut down on the possibilities of human biases and mistakes which are an inherent part of the process. Also, ML uses the model to recognize borrowers who may default and subsequently lower the likelihood of default. Part of the task involve utilizing historical credit market data and implementing ML algorithms to develop a highly accurate and reliable loan approving system based on trained-data, random forests, the stream of loans and reliable clients.

Keywords:- *Construction of Data, Creation of Correlation, Banker Loans, Customer Safety.*

I. INTRODUCTION

For banks, lending is the main area of work lending, money changing, taking deposits from customers, and earning profits by making interest from customers' loans. The goal here is the funding of creditworthy borrowers, but standard procedures through lending banks fail to provide protection for selected borrowers. To solve this problem, we designed a credit finding algorithm with the help of Python and began to get credit approvals from banks. This project does an eligibility analysis of a client by a process of evaluating different parameters including marital status, income, and expenditure to determine whether they meet the capacity to repay the loan or not. We use the creates trained data sets as a model that we directly then test the data to output a result of "yes" or "no" only to determine if a customer or not. can repay the loan. It goes without saying that This loan approval process is good business sense.

Automating approvement reduces the use of electricity, upgrades the quality, and makes the customer satisfaction better. It can do so through machine learning and hence can do more precise and consistent loan decisions. This approach facilitates just and accurate loan approvals. ML models are a solution to fair analysis and cut off human biases and errors in manual data analysis.

II. METHODOLOGIES AND TECHNIQUES

The supervised learning methods are employed to train the prediction model in this loan approval ML task. We research and study different machine learning algorithms such as logistic regression, decision trees, random forests, and support vector machines. Feature selection methods and data preparation techniques are used to improve model performance and learning. Undefined.

The execution of the mentioned program may be beneficial for banks and borrowers at the same time. The automated loan approval process simplifies decision-making, saves manual work and improves efficiency in the financial institutions' operations. ML model facilitates risk management through precise identification of high-risk customers, leading to fewer defaults and losses. Borrowers appreciate a fair and clear loan approval system, in which many factors are taken into consideration and credit scores are not the key consideration in a holistic credit evaluation. This paves the way for the ones with little or no credit history or non-financial characters. ML techniques will be

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR292

used as a basis for developing a loan approval mechanism that is more precise and operationally efficient for both lenders and borrowers.

In this sphere we face data biases from historical cases showing the gap in practices while the machine learning models are accurate in the quantitative domain but may be less context sensitive and very complex. On top of being difficult to comply with, there exists the risk of over–reliance on models.

To cope with these issues a fairness must be ensured, and a coordinated and comprehensive strategy should be used that includes both human qualified experts. The loan approval process involves a detailed consideration. It considers contextual factors and complies with regulatory guidelines. The objective in mind is making choices between risks and opportunities wisely.

III. LITERATURE SURVEY

A. An Approach Using Machine Learning Algorithms for Loan Approval Prediction

The study by Mohammad Ahmed Sheikh, Amit Kumar Goel and Tapas Kumar highlights the importance of predicting defaults of banking systems to reduce non-performing assets (NPAs) and increase profitability. The paper compares different forecasting methods, with a particular focus on Logistic Regression models.

Through the study it is concluded that the addition of individual characteristics to traditional checking account information improves forecasting accuracy. This highlights the importance of considering a range of consumer factors for loan approval decisions.

B. Default Forecasting of Loans Through Data Mining

Bhumi Patel, Harshal Patil, Jovita Hembram, and Sri Jaswal propose the use of data mining algorithms to predict loan defaults. By using datasets containing home loan applications, the paper aims to help banks make more rational decisions.

The study highlights the need for accurate accounting to minimize losses and increase loan approvals.

C. Forecasting Loan Position in Commercial Banks by Machine Learning Classifier

G. Arutjothi, Dr. C.S. Senthamarai will focus on building credit scoring models for commercial banks using machine learning techniques. The paper recommends a machine learning classification-based approach, where Min-Max normalization and K-Nearest Neighbour (K-NN) algorithm are combined the goal is to predict the loan condition more accurately, thus helping in making credit decisions.

D. Bank Loan Overpayment Prediction Based on LSTM-SVM

Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li propose a new method for bank loan default prediction using long-term short-term memory (LSTM) and support vector machine (SVM) algorithms combining so Provides predictive accuracy.

E. Forecast Default for Networked- Guaranteed Loans

Dawei Cheng, Zibin Niu, Yi Tu, and Liqing Zhang address the challenge of predicting the risk of guaranteed loans in a network by applying an imbalanced approach to the spread of a network with its neighbour a well-weighted (pwkNN) to be used to forecast cooperative risk in their proposed model) and algorithms are included.

F. Personalized Credit Rating Using Artificial Intelligence Technology for Federal Student Loans

Jian Hu describes the development of a loan management system designed for federal student loans using Artificial Neural Network (ANN) technology. Leveraging the adaptive capabilities of ANN, proposed back propagation neural network reveals effective credit rating evaluation for college students despite the presence of credit history.

IV. DATA ANALYSIS

The course starts with a thorough examination of a historical credit data that contains attributes of applicants as well as their credit outcomes. The data set has been carefully crafted and cleaned for missing values, redundancies, and class imbalance issues.

Exploratory data analysis methods are used to study the distribution of loan permissions among different demographic groups, types of loans, and risk types, and feature engineering techniques to get additional knowledge and the accuracy of ML models rises.

The study starts with a comprehensive study of historical credit data, which includes various components on applicants' financial background and credit outcomes The data set is carefully crafted and pre-processed to improve its standards of relevance, duplicates, and class imbalance.

With exploratory data analysis we determine how the loan approvals are distributed through different demographic groups, loan types and risk types and by applying feature engineering techniques we gain useful information and improve our ML models performance.

V. MODEL DEVELOPMENT

The fundamental part of the research is the development of ML algorithms and the estimation of their accuracy in classification of loan applicants. The algorithms used, such as logistic regression, decision trees, random forest, and gradient boosting, are many.

With the use of boosting the machines and neural networks are getting trained and improved with cross-validation method. Hyperparameter tuning and ensemble learning approaches are included to increase the model accuracy, stability, and generalization capability.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR292

Similarly, model interpretability is critical as a variable in this regard as it allows the processes of a loan to be understood and the procedure to remain transparent.

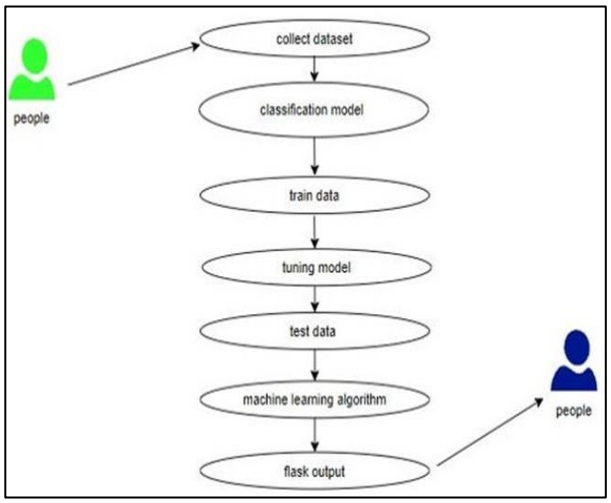

Fig 1: Workflow of Proposed System

VI. IMPLEMENTATION AND RESULT

- *PyTorch:*
  A deep learning framework that consists of developing and training neural network models.

- *NumPy:*
  Essential for doing numerical computations and handling arrays.

- *Pandas:*
  Enables data management and analysis, including the handling of large data sets.

- *Matplotlib:*
  Introduces data visualization and performance metrics of the model.

- *Scikit-Learn:*
  Include evaluation metrics and data pre-processing while undertaking machine learning operations.

Table 1 Implementation

NumPy	Numerical computations andarray manipulation
Pandas	Data manipulation andanalysis
Matplotlib	Data visualization
Scikit-Learn	Machine learning tools andevaluation metrics

❖ ***RESULTS***

The distribution mentioned for each model serves as a reminder of its ability to accurately classify loan approvals. The logistic model demonstrated the accuracy of 72%. Its accuracy, recall, and F1-score were 73%, 72%, and 72% respectively. The decision tree classification model achieved an accuracy of above 90%, with precision, recall and F1 scores of 0.91, 0.90 and 0.90 respectively.

F1 scores of 91%, 90%, and 90% for each dataset. Random forest and the k-nearest neighbour classification algorithms had precision, recall, and F1 scores of 92% and 82% respectively which did not significantly vary from those of performance levels of other classifiers. These results provide financial institutions which are seeking to implement loan approval prediction algorithms an effective framework for decision making.

Table 2 Results

	Metric	**Value**
Logistic Regression	Accuracy	72%
	Recall	86%
	Precision	70%
	F1 Score	65%
Decision Tree	Accuracy	90%
	Recall	95%
	Precision	95%
	F1 Score	91%
Random Forest	Accuracy	92%
	Recall	95%
	Precision	95%
K-NN	Accuracy	92%
	Recall	95%
	Precision	95%

VII. CONCLUSION

Finally, this study compares classification models applicable for loan approval prediction and develops implications that relate to the models' performance and suitability. The research results show that each model possesses unique strengths and weaknesses, but it is revealed by the data that both decision trees classification and forest random forests classification are predictors of loan outcomes that could be most reliable and accurate. Employing these models, financial institutions can streamline the decision-making processes, automate credit scoring, and conduct risk assessments, enhancing their customer satisfaction and operational profits.

FUTURE RESEARCH

The future studies should concentrate on sophisticated machine learning techniques like neural networks for loan approval predictions. On the other hand, the introduction of new data sources and advanced feature engineering techniques can lead to a greater precision of models. In

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR292

addition to that, research efforts can better be aimed at creating machine learning models that are interpretable to increase the transparency and confidence in the loan approval decision-making processes.

REFERENCES

[1]. K. Hanumantha Rao, G. Srinivas, A. Damodhar, M. Vikas Krishna. "Implementation of Anomaly Detection Technique Using Machine Learning Algorithms." International Journal of Computer Science and Telecommunications, Vol. 2, Issue 3, 2011.
[2]. S.S. Keerthi, E.G. Gilbert. "Convergence of a Generalized SMO Algorithm for SVM Classifier Design." Machine Learning, Springer, Vol. 4, Issue 1, pp. 351-360, 2002.
[3]. Andy Liaw, Matthew Wiener. "Classification and Regression by Random Forest." Machine Learning, Vol. 2, Issue 3, pp. 9-22, 2002.
[4]. Ekta Gandotra, Divya Bansal, Sanjeev Sofat. "Malware Analysis and Classification: A Survey." Journal of Information Security, Vol. 05, Issue 02, pp. 56-64, 2014.
[5]. Rattle Data Mining Tool. [Online] Available: http://rattle.togaware.com/rattle-download.html.
[6]. Aafer Y., Du W., Yin H. "Droid APIMiner: Mining API-Level Features for Robust Malware Detection in Android." Security and Privacy in Communication Networks, Springer, pp. 86-103, 2013.
[7]. J. R. Quinlan. "Induction of Decision Tree." Machine Learning, Vol. 1, No. 1, pp. 81-106.