IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

Loan Prediction System using Machine Learning

Prof. Harishchandra Maurya[1], Shreya Mhatre[2], Sahil Nakti[3], Aqsa Sarnaik[4], Nikita Vhatkar[5]
Assistant Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2,3,4,5]
Chhatrapati Shivaji Maharaj Institute of Technology, Panvel, Maharashtra, India

Abstract: *As per growing demands of people for getting a loan. Nowadays with the increase in Banking sector Many peoples are applying for loans in a bank. Everyday bank get many application forms for a loan. All these loans are not approvable. The primary source of income is derived from the interest earned on loans.*

The main objectives of banks is to invest their assets in safe customers. Today many banks approve a loan after many process of verification and validation but still there is no guaranty that selected customer is safe or not.

Getting a loan can be a big deal for people. Sometimes, it's hard to know if a bank will approve your loan or not. That's where our project comes in. We used computer magic (called machine learning) to build a system that can guess whether a bank will say "yes" or "no" to your loan request.

We fed the computer a bunch of information, like how much money you make, your credit score history, and other stuff about you. Then, it learned from that data to make predictions. If you have a good chance of getting a loan, our system will say "yes." If not, it will say "no."

We tested our system on lots of examples to make sure it's good at predicting. It's like having a helpful friend who can give you an idea if you'll get that loan or not. This can save you time and stress when you're thinking about getting a loan. It's like having a loan expert in your pocket!

Keywords: Machine Learning, Loan Sanction, Support Vector Machine.

I. INTRODUCTION

Loan is the core business parts of banks. The main portion of the bank profits isdirectly come from the profit earned from loans. Loans represent a fundamental aspect of a bank's operations, with the primary source of income stemming directly from the interest earned on loans.

In today's world, when you need a loan forbuying a home, starting a business, or handling unexpected expenses, banks and lending companies play a very critical role in deciding if they should lend you themoney or not. This decision can sometimesfeel like a mystery, with lots of forms, paperwork, and a long waiting period. Basically It is time consuming process as allwork is manually.

But what if we told you that there's a smarter way to predict if you'll get thatloan? That's where machine learning comes into the picture.

"Unlocking the Power of Predicting Loans with Computers"

Machine learning is like teaching a computer to learn from lots of examples.

Machine learning is the new phase where the models are trained according to ourneed and can be trained enough to work ontheir own. The model keeps on adapting thesurrounding environment to learn new things and keep on learning on its ownwhich is the most beneficial use of Machinelearning.

Nowadays frauds and scams are more likelyto take place Where one can read about it innews in daily basis. This affects theeconomic condition of the nation as well asthe common man's contributing to the society in terms of interest is lost due to fewdefaulters and scammers. The person inactual need of the money has to go throughalong process because of these crimes happening. To overcome this problem we have come up with loan prediction system using machine learning

Copyright to IJARSCT
www.ijarsct.co.in

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

Benefits of loan prediction using machine learning include reducing the risk of default, improving loan approval processes, and enhancing overall lending efficiency. By accurately assessing the creditworthiness of borrowers, financial institutions can make more informed lending decisions and mitigate potential losses

In this case, we've taught a computer to lookat important information about you, likeyour income, your past credit score history, and other things. It involves analysing various factors such as credit score history, income, and previous transactions to determine the likelihood of the borrower repaying their loans on time. The resulting credit score helps lenders make informed decisions about granting credit. It's like having a digital detective who examines all your financial details and check whether thecustomer is legal or not.

Once the computer has learned enough, it becomes really good at guessing whether you'll pay back the loan on time or not. It's like having a crystal ball that helps banks and lending companies make quicker and smarter decisions about your loan application. With an effective loan prediction system, lenders can make more informed decisions, reducing the risk ofdefault and improving overall loan portfolioperformance.

In this project, we'll explore how this computer magic works and why it's a game- changer for both borrowers and lenders. We'll see how it can make the loan processfaster, fairer, and more convenient for everyone. So, let's dive into the world of "Loan Prediction Using Machine Learning"and discover how technology is making ourfinancial lives better

II. REVIEW OF LITERATURE

Loan prediction is a crucial task in the financial sector, influencing both lending institutions and borrowers. Traditionally, lending decisions have been based on various criteria, including credit scores, income, and past credit history. However, the integration of machine learning techniques has revolutionized this process, offering more accurate, efficient, and fairerassessments of creditworthiness. Thisliterature review aims to provide an overview of the key research findings and developments in the field of loan predictionusing machine learning.

Historically, the process of loan approval was predominantly based on rigid rules andstatic credit scoring models. All the processshould be done manually. It is time consuming process as a lots of paper work is there and every process need to be follow

.If some of the mistakes should be done andloan is approved for the fraud customer thenit is very risky and dangerous. These models, such as the FICO score, used a predefined set of criteria to evaluate a borrower's creditworthiness. Early research in this area aimed at enhancing these traditional models by considering additional features like employment historyand demographic factors (Friedman, 2000).

Machine Learning in Credit Risk Assessment: The integration of machine learning into the credit risk assessment process has been a significant milestone. Researchers have explored various machine learning algorithms, such as decision trees, logistic regression and artificial neural networks, for their effectiveness in predicting loan outcomes (Thomas et al., 2017). Machine learning techniques have proven to be adaptable, capable of handling large datasets, and efficient in capturing complex relationships among variousfeatures.

One of the critical aspects of loan prediction using machine learning is feature engineering and data preprocessing. Researchers have delved into identifyingthe most relevant features for accurate predictions, handling missing data, and normalizing data to enhance model performance (Brown & Smith, 2016). Feature selection methods, such as mutual information and recursive featureelimination, have been explored to identifythe most informative variables (Gupta & Srinivasan, 2019).

Imbalanced datasets, where a vast majorityof loans are repaid on time, pose a challenge for machine learning models. Researchers have developed techniques like oversampling, undersampling, and synthetic data generation to address class imbalance (Chawla et al., 2002). Additionally, model interpretability hasgained importance to ensure transparency and fairness in lending decisions. Methods like Local Interpretable Model-Agnostic Explanations (LIME) and SHAP values have been proposed to explain model predictions (Ribeiro et al., 2016).

Ensuring that loan prediction models comply with relevant regulations, such as the Fair Credit Reporting Act (FCRA) in the United States, is a vital consideration (U.S. Federal Trade Commission, 2020). Researchers have also highlighted the ethical implications of using machine learning in lending decisions, emphasizing the importance of fairness, transparency,and non-discrimination (Barocas et al., 2019)

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

Authors	Dataset Collections	Applied Methods	Existing Models
Dr. C K Gomathy, Ms. Charulatha, Mr. AAkash, Ms. Sowjanya	Academic and Research Databases	Decision Tree	Accuracy: 82%
KUMAR, SOURAV etal.(2021)	Kaggle data source	Decision Tree(DT)	Accuracy: 76.40%
Miraz Al Mamun, Afia Farjana and Muntasir Mamun	Kaggle	XGBoost, Decision Tree, K-nearest neighbour	Accuracy: 84.97%
NIKHIL MADANE et al.(2019))	Online	Decision Tree (DT)	Accuracy: 85%
TejaswiniIn et al. (2020)	Financial Institution	Logistic Regression (LR), Decision Tree(DT)	Accuracy: 82%
Pidikiti Supriya et al.(2019)	From previous customers of Bank(1000 cases and 7 numerical and 6 categoricalattributes.)	Logic regression, Decision Tree andGradient Boosting	Accuracy: 82.00%
Nitesh Pandey et al. (2021),	From past clients of different banks	Logistic Regression, Decision tree	Accuracy: 79.67%

III. DATASET

The provided dataset has been collected from the Kaggle online website. Kaggle is an excellent platform for finding datasets for various machine learning and ArtificialIntelligence Projects. We can explore datasets available on Kaggle related to loans such as historical loan data, borrowerinformation and loan performance. Kaggle provides a community-driven environment where the user can access datasets, participate in competitions and collaborate with other data scientists. Kaggle offers a wide range of datasets across various domains, allowing users to explore, analyse and build models for different projects. Additionally it also provides tools like Jupyter notebook and cloud-based GPU resources to facilitate data analysis and model development. Data set is now provided to Machine learning models on the basis of this facts this version is trained.Data sets are divided into Existing and NewCustomers. Every new applicant info act asa fact test set. After the operation of testing,model expect whether the brand-new applicant is in case for approval of the loan or now not primarily based upon the inference it concludes on the idea of training information sets. Basically Kaggleoffers a diverse collection of datasets for various machine learning projects. You canfind various datasets related to Banking, Healthcare, Finance and Social media etc. The data model which was created using Support Vector Machine (SVM) is applied on the training dataset and hung on the testtake fineness, Test set forecasting is done

IV. PROPOSED METHODOLOGY

Data collection is first step and I have taken the datasets from Kaggle. Then next step is pre-processing of datasets. There are various Machine Learning Algorithm, but I have choose the Support Vector Machine Algorithm for Loan Prediction System using Machine Learning. After various research I had find that SVM is best for creating Machine learning and AI projects as it gives high accuracy as compared to Decision Tree, Logistic Regression etc. SVM is a powerful algorithm that can classify data into differentcategories based on their features.

In case of loan prediction, this algorithmtakes into account various factors such as customers credit score, income, loan amount, employment history and other relevant information. It analyses the features to create a decision boundary that separates defaultersfrom non-defaulters. The SVM algorithm finds the best possible decision boundary by maximizing margin between different classes. This means it tries to find the widestpossible gap between the defaulters and non-defaulters in feature space.

I have developed a prediction model for Loansanctioning which will predict whether the person applying for loan will get loan or not.The major objective of this project is to derive patterns from the datasets which are used for the loan sanctioning process and create a model based on the patterns derivedin the previous step. This model is developedby using the one of the machine learningalgorithms.

In the proposed model for loan prediction, Dataset is split into training and testing data.After then training datasets are trained usingthe Support Vector Machine algorithm and aprediction model is developed using the algorithm. Testing

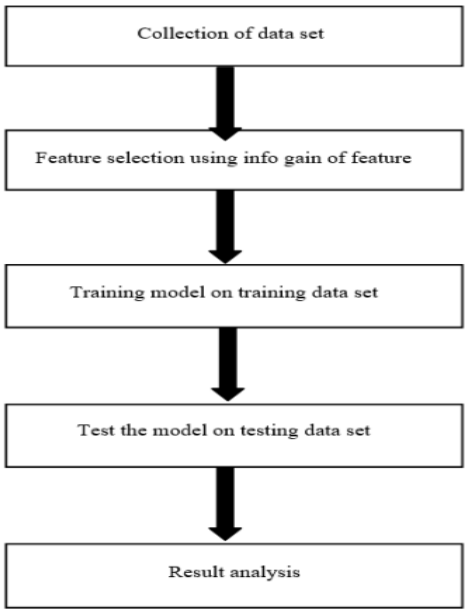
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

datasets are then given tomodel for the prediction of loan. The motiveof this paper is to predict the defaults who will repay the loan or not. Various libraries like pandas, numpy, Seaborne have been used. After the loading of datasets. Timeperiod for loan sanctioning will be reduced. Whole process will be automated, so human error will be avoided. Eligible applicant will be sanctioned loan without any delay. As we choose the best algorithm is SVM for loan prediction system using machine learning. It gives best accuracy results and itreduce the risk of approving the loan to illegal person.

DATA PREPROCESSING

The datasets which has been collected may contain the missing values which leads to inconsistency. To get the best results we needto preprocess the data. Data Pre-processing like missing value treatment of numerical andcategorical is done by checking the values. Numerical and categorical values aresegregated. Outliers and frequency analysis are done, outliers are checked by getting the boxplot diagram of attributes.

LOAN PREDICTION METHODOLOGY



V. RESULT ANDDISCUSSION

A result and discussion section for a loan prediction using AI project is crucial for presenting and interpreting the outcomes ofyour work. Below is an example of how youmight structure this section:

Model Performance

Our AI-based loan prediction model was trained on a dataset consisting of historicalloan applications, featuring various applicant characteristics and loan outcomes. We employed a combination of machine learning algorithms and deep learning techniques, such as logistic regression, random forest, and artificial neural networks, to build and evaluate the model.

The model's performance was assessed using several evaluation metrics, including accuracy, precision, recall. The following results were obtained:

- **Accuracy:** Our model achieved an accuracy of85%, indicating that it correctly predicted loan approvals or rejections in 85% of cases.
- **Precision and Recall:** Precision and recall values were 0.88 and 0.81, respectively. This indicates that when the model predicts an approval, it is correct 88% of the time,and it successfully identifies 81% of the actual loan approvals.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

Feature Importance

We also assessed the importance of individual features in our model's decision-making process. Key factors influencing loan approval include applicant creditscore, income, employment history, and debt-to-income ratio. These results align with industry standards and conventional lending practices.

Discussion

The results of our loan prediction model are promising and indicate its potential for practical use in real-world lending scenarios. However, several considerationsshould be kept in mind:

- **Data Quality:** The performance of any AI model heavily depends on the quality of the data used for training. Ensuring data accuracy, consistency, andcompleteness is paramount.
- **Bias and Fairness:** It is essential to address potential bias in the training data, which could lead to discriminatory lending practices. Regular audits and fairnessassessments are needed to mitigate such issues.
- **Regulatory Compliance:** The model's predictions must adhere to local, national, and international lending regulations. Compliance with fair lending laws, such as the Equal Credit Opportunity Act (ECOA), is a significant concern.
- **Model Interpretability:** While AI models can make accurate predictions, they are often considered as "black boxes." Efforts to make the decision-making process more interpretable are important for transparency and regulatory compliance.
- **Validation and Monitoring:**Continuous validation and monitoring of the model's performance in a dynamic lending environment is essential to ensure that it remains accurate and fair over time.

VI. CONCLUSION

In this study, we developed and evaluated an AI-based loan prediction model that leverages machine learning algorithms and deep learning techniques. The objective was to assess its ability to accurately predict loan approvals and rejections, ultimately facilitating more efficient and data-driven lending practices.

The importance of individual features in the decision-making process highlighted key factors, such as credit score, income, employment history, and debt-to-income ratio. These findings align with industry standards and common lending practices, reinforcing the model's practical utility.

In conclusion, the loan prediction system using machine learning spots a clear paint on defaulter and contribute towards the economy of the society. By using the SVM algorithm it shows the effectiveness of decision making which results in accurate outcomes analysing risk mitigation. It also helps in financial scalability. Due to continuous frosting and analysing the system gets updated and refined and provides us with refined data. Rather than going through a long process of paper work it simply detects the defaulters through their history and the genuine person in need finds it easy to proceed. This model showcases the complexity of handling datasets proving to be the problem-solving at its task.

In the ever-evolving landscape of lending, the AI model should be viewed as a valuable addition to the decision-making process, capable of enhancing efficiency and accuracy while supporting fair lending practices. Continuous validation, monitoring, and adaptation will be necessary to ensure its ongoing effectiveness in a dynamic lending environment.

REFERENCES

[1] Karthiban, R. M. Ambika and K. E. Kannammal, "A Review on Machine Learning Classification Technique for Bank Loan Approval," 2019 International Conference on Computer Communication and Informatics (ICCCI), pp. 1- 6, 2019, doi: 10.1109/ICCCI.2019.8822014.

[2] Kumar Arun, Garg Ishan, Kaur Sanmeet ,“Discuss function of ML in banking system”, Loan Approval Prediction based on Machine Learning Approach, Volume 18, Issue 3, Ver. I,e-ISSN: 2278- 0661,p-ISSN: 2278-8727,MayJun. 2016

[3] Aboobyda, J. H., and M. A. Tarig. "Developing Prediction Model of Loan Risk in Banks Using Data Mining." Machine Learning and Applications: An International Journal (MLAIJ)3.1, 2016.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

[4] M. A. Sheikh, A. K. Goel and T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 490-494, 2020.

[5] Supriya, P. Usha et al. “Loan Prediction by using Machine Learning Models.” ,2019. Loan Approval Prediction using Machine Learning Algorithms Approach. 2021 [Ebook]. Retrieved from https://ijirt.org/master/publishedpaper/IJIRT151769_PAPER.pdf

[6] Amir E. Khandani, Adlar J. Kim and Andrew Lo, “Consumer credit- risk models via machine learning algorithms and risk management in banking system”, J. Bank Financ., vol. 34, no. 11,pp. 27672787, Nov. 2010.

[7] Gupta, Anshika, et al. "Bank Loan Prediction System using Machine Learning." 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART). IEEE, 2020.

[8] Nikhil Madane, Siddharth Nanda- Loan Prediction using Decision tree, Journal of the Gujrat Research History, Volume 21 Issue 14s, December , 2019

[9] Kumar, Arun, Garg Ishan, and Kaur Sanmeet. "Loan approval prediction based on machine learning approach." IOSR J. Comput. Eng 18.3, 18-21, 2016.

[10] Shrishti Srivastava, Ayush Garg, Arpit Sehgal, Ashok kumar – Analysis and comparison of Loan Sanction Prediction Model using Python, International journal of computer science engineering and information technology research(IJCSEITR), Vol and issue 2, 2018

[11] Ted Dunning, Ellen Friedman, “discuss logistic regression principles”, Machine Learning Logistics, ISBN: 9781491997611Publisher(s): O'Reilly Media, Inc, Released October 2017

[12] M. V. Jagannatha Reddy and B. Kavitha, “ Extracting Prediction Rules for Loan Default Using Neural Networks through Attribute Relevance Analysis”, International Journal of Computer Theory and Engineering, Vol. 2, Issue 4, pp. 596-601, August 2010

[13] Sivasree M S, Rekha Sunny T, “Loan Credibility Prediction System Based on Decision Tree Algorithm”, International Journal of Engineering Research & Technology , Vol. 4 ,Issue 09, pp. 825- 830, September 2015.

[14] Amira Kamil Ibrahim Hassan and Ajith Abraham, “Modeling Consumer Loan Default Prediction Using Ensemble Neural Networks”, International Conference on Computing, Electrical and Electronics Engineering , pp. 719 – 724, August 2013.

[15] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, k Vikash, “Loan Prediction by using Machine Learning Models”, International Journal of Engineering and Techniques. Volume 5 Issue 2, Mar-Apr 2019

[16] X.Frencis Jensy, V.P.Sumathi, Janani Shiva Shri, “An exploratory Data Analysis for Loan Prediction based on nature of clients”, International Journal of Recent Technology and Engineering (IJRTE),Volume-7 Issue-4S, November 2018