ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/05.2024-66751549/IJAEAST.2024.04.0005, April 2024, Volume-2, Issue-4, pp.139 - 143
International Journal of Advances in Engineering Architecture Science and Technology

Loan Default Risk Analysis

Malla Keerthi Vennela[1], S. Sai Druthi[2], CH. S.S.G. Deepak[3], K. Naresh[4]

[1,2,3,4]Department of CSDM, Dadi Institute of Engineering & Technology, Andhra Pradesh, India

vennelamalla2002@gmail.com[1]

Received: 21-02-2024 **Accepted:** 27-04-2024 **Published:** 29-04-2024

Abstract

Background: The loaning investigation has ended up a profoundly basic inquiry about range since it may help avoid credit defaults and give credits to those who would pay on time.
Objectives: Hence, it even though, that we formulated a method for machine learning known as the arbitrary woodland strategy, additionally the information was utilized in this.
Methods: Anything fundamental is accumulated from web destinations, and the information assembled is normalized sometime recently being utilized for inquiring about and anticipating yield, and it is at that point conveyed to the irregular timberland strategy, which is utilized in our inquiry.
Statistical Analysis: After that, we may utilize the program to decide on the off chance that an individual is qualified for an advance or not, and a bank might not only target the well-off.
Findings / Applications and Improvements: Clients are gotten to for advance purposes, but it too gets to other perspectives of a client, that play a critical part in credit-giving choices and loaning forecast assessment dodgers.

Keywords: Credit endorsement, Irregular Woodland calculation, Pandas, Matplotlib.

1. Introduction

Lending analysis has emerged as a crucial field of study, as it can aid in preventing loan defaults and granting loans to individuals who are likely to make timely payments. In this regard, we have developed a machine learning technique known as the random forest algorithm, utilizing data gathered from various sources. The collected data is pre-processed and normalized before being employed for analysis and output prediction through the random forest method.

2. Literature Survey

A benchmark figure is required in all commercial managing an account business to evaluate whether to allow a credit to a person candidate. The judgment call criteria don't get to be limited to a single property; they might include any number of qualities that must be taken under consideration. Cash loan specialists may supply datasets counting the related data for their shoppers. This dataset's properties will be utilized to build a calculation that will evaluate on the off chance that a credit ought to truly be endorsed for a certain client. There are two conceivable outcomes conceivable: appropriation or refusal. The built demonstrate must reach conclusions speedier than craved. Computer science may offer assistance with prediction, judgment, and learning with information. It has it possess enhance. Information is the foremost vital thing within the world, that have activated a renaissance within the reach of computer science. Machine learning procedures have created a wide extend of information item based. To procure information for this demonstrate, I examined a few articles. The journalists of the article pointed

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/05.2024-66751549/IJAEAST.2024.04.0005, April 2024, Volume-2, Issue-4, pp.139 - 143
International Journal of Advances in Engineering Architecture Science and Technology

to diminish the endeavours put forward by banks by developing a demonstrate utilizing a extend of calculations to memorize and laying out which of the methods can be right. The four factors of the paper were information collection, appraisal of different machine learning strategies on the information, giving total and testing. They utilized a mapper to figure the passages. Scholars were trying to find surveys within the investigate. Credit score of modern contracts and application criteria are made utilizing the inductive choice tree procedure. The credit score has an effect on credit endorsement. Analysts created a show to check in the event that credit authorizing is secure and it was found that restricted clients appear to be more likely to be affirmed for advances since they're more likely to reimburse them. This test was accumulated utilizing Kaggle. They utilized the randomized timberland strategy within the paper. They utilized the decision-tree approach inside the paper. A test set is utilized to certify the frame. The researchers utilized information collection to create a demonstrate within the ponder and the device is made up of three components.

3. Overview of the System

Existing System

The existing system relied on the decision tree method, which had limitations in terms of handling large volumes of data and providing accurate results. To address these shortcomings, we propose the implementation of the random forest algorithm, which is an ensemble of decision trees and offers improved performance and accuracy.

Proposed System

The proposed system follows a modular design, consisting of the following components:

- Dataset: Large datasets are collected, and the data is divided into training and testing sets. The training data is used to train the machine learning algorithm, while the testing data is employed to evaluate the model's accuracy.
- Preprocessing: Data preprocessing is performed to handle missing values, outliers, and any other anomalies present in the dataset. This step ensures consistent application of the random forest technique.
- Data Collection: A user-friendly web interface, developed using front-end technologies like HTML and PHP, facilitates the collection of user information required for loan eligibility assessment.
- Data Analysis: The trained model is utilized to determine whether an individual is eligible for a loan, providing a streamlined process for both bank employees and loan applicants with remarkable accuracy.

4. Architecture

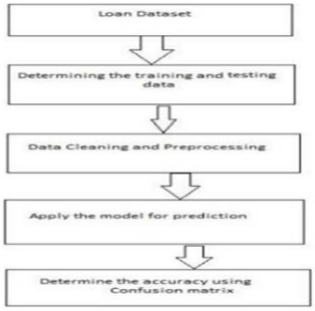
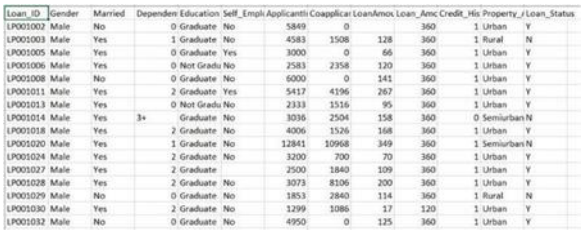


Figure 1. Frame Work of Loan Prediction

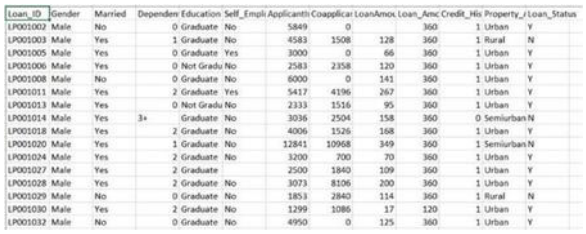
ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/05.2024-66751549/IJAEAST.2024.04.0005, April 2024, Volume-2, Issue-4, pp.139 - 143
International Journal of Advances in Engineering Architecture Science and Technology

5. Results Screen Shots



Loan_ID	Gender	Married	Dependen	Education	Self_Empl	Applicanti	Coapplicar	LoanAmou	Loan_Amo	Credit_His	Property_/	Loan_Status
LP001002	Male	No	0	Graduate	No	5849	0		360	1	Urban	Y
LP001003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
LP001006	Male	Yes	0	Not Gradu	No	2583	2358	120	360	1	Urban	Y
LP001008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
LP001013	Male	Yes	0	Not Gradu	No	2333	1516	95	360	1	Urban	Y
LP001014	Male	Yes	3+	Graduate	No	3036	2504	158	360	0	Semiurban	N
LP001018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
LP001020	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurban	N
LP001024	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
LP001027	Male	Yes	2	Graduate		2500	1840	109	360	1	Urban	Y
LP001028	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	Y
LP001029	Male	No	0	Graduate	No	1853	2840	114	360	1	Rural	N
LP001030	Male	Yes	2	Graduate	No	1299	1086	17	120	1	Urban	Y
LP001032	Male	No	0	Graduate	No	4950	0	125	360	1	Urban	Y

a) Training Dataset



Loan_ID	Gender	Married	Dependen	Education	Self_Empl	Applicanti	Coapplicar	LoanAmou	Loan_Amo	Credit_His	Property_/	Loan_Status
LP001002	Male	No	0	Graduate	No	5849	0		360	1	Urban	Y
LP001003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
LP001006	Male	Yes	0	Not Gradu	No	2583	2358	120	360	1	Urban	Y
LP001008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
LP001013	Male	Yes	0	Not Gradu	No	2333	1516	95	360	1	Urban	Y
LP001014	Male	Yes	3+	Graduate	No	3036	2504	158	360	0	Semiurban	N
LP001018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
LP001020	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurban	N
LP001024	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
LP001027	Male	Yes	2	Graduate		2500	1840	109	360	1	Urban	Y
LP001028	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	Y
LP001029	Male	No	0	Graduate	No	1853	2840	114	360	1	Rural	N
LP001030	Male	Yes	2	Graduate	No	1299	1086	17	120	1	Urban	Y
LP001032	Male	No	0	Graduate	No	4950	0	125	360	1	Urban	Y

b) Testing Dataset

c) Registration Page

d) Login Page

e) Upload Details

f) Predicted Result

Figure 2. Implementation Results

6. Model Comparison

As a part of the project, the models under our consideration are:

- Support Vector Machine
- Logistic Regression
- Random Forest Classifier

Table 1. Comparison

Model	Accuracy
Logistic Regression	classification report of training data precision recall f1-score support 0 0.95 0.45 0.61 154 1 0.80 0.99 0.88 337 accuracy 0.82 491 macro avg 0.87 0.72 0.75 491 weighted avg 0.84 0.82 0.80 491 classification report of testing data precision recall f1-score support 0 0.83 0.39 0.54 38 1 0.78 0.96 0.86 85 accuracy 0.79 123 macro avg 0.81 0.68 0.70 123 weighted avg 0.80 0.79 0.76 123

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/05.2024-66751549/IJAEAST.2024.04.0005, April 2024, Volume-2, Issue-4, pp.139 - 143
International Journal of Advances in Engineering Architecture Science and Technology

Support Vector Machine	cost of training model precision recall f1-score support 0 0.96 0.47 0.63 154 1 0.80 0.99 0.89 337 accuracy 0.83 491 macro avg 0.88 0.73 0.76 491 weighted avg 0.85 0.83 0.81 491 cost of testing model precision recall f1-score support 0 0.79 0.39 0.53 38 1 0.78 0.95 0.86 85 accuracy 0.78 123 macro avg 0.78 0.67 0.69 123 weighted avg 0.78 0.78 0.75 123
Random Forest Classifier	cost of training model precision recall f1-score support 0 0.99 0.90 0.94 154 1 0.95 1.00 0.98 337 accuracy 0.97 491 macro avg 0.97 0.95 0.96 491 weighted avg 0.97 0.97 0.96 491 cost of testing model precision recall f1-score support 0 0.68 0.39 0.50 38 1 0.77 0.92 0.84 85 accuracy 0.76 123 macro avg 0.73 0.66 0.67 123 weighted avg 0.74 0.76 0.73 123

From the above designed models, we have considered Random Forest Classifier as a model with 97% train accuracy to predict the eligibility criteria of loan applicants.

7. Conclusion

So, we created a system where we could send data directly to the home page and then transform the input into the data layer where we use random forests to analyse the information. Using random forest to obtain the customer's approval, this algorithm is a reliable and effective way to determine whether the customer will accept the loan. It has a high accuracy rate in predicting credit and provides an easy way to decide whether to be approved or not. In addition, since the system takes into account many negative factors when calculating mortgage loan costs, there is less chance of making mistakes. Additionally, the random forest method is flexible and can be applied to large data sets.

8. Future Enhancement

We may try to develop and improve the current techniques so that the correctness of the result is enhanced and the time required is decreased so that we can receive an outcome in a brief time, and we can attempt to integrate them for any active learning environment in order to ensure the banker's hard workload is lowered.

References

1. Kumar Arun, Garg Ishan, Kaur Sanmeet, May-Jun. 2016. Loan Approval Prediction based on Machine Learning Approach, IOSR Journal of Computer Engineering (IOSR-JCE).
2. Wei Li, Shuai Ding, Yi Chen, and Shalin Yang, Heterogeneous Ensemble for Default Prediction of Peer-to-Peer Lending in China, Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei University of Technology, Hefei 2009, China.
3. Short-term prediction of Mortgage default using ensembled machine learning models, Jesse C. Sealand on July 20, 2018.

4. Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Techniques, Dr. K. Kavitha, International Journal of Advanced Research in Computer Science and Software Engineering.
5. K. Hanumanth Rao, G. Srinivas, A. Damodar, M. Vikas Krishna: Implementation of Anomaly Detection Technique Using Machine Learning Algorithms: International Journal of Computer Science and Telecommunications (Volume2, Issue3, June 2011).
6. S.S. Keerthi and E.G. Gilbert. Convergence of a generalizes MO algorithm for SVM classifier design. Machine Learning, Springer, 46(1):351–360, 2002.
7. Shiva Agarwal, “Describe the concepts of data mining”, Data Mining: Data Mining Concepts and Techniques, INSPEC Accession Number: 14651878, Electronic ISBN:978-07695-5013-8, 2013.
8. 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.
9. A kindaini, Bolarinwa. “Machine learning applications in mortgage default prediction.” University of Tampere, 2017.
10. Amir E. Khand ani, Adlar J. Kim and Andrew Lo, “Consumer credit-risk models via machine learning algorithms and risk management in banking system. Bank Finance., vol. 34, no. 11, pp. 27672787, Nov. 2010.