International Journal for Modern Trends in Science and Technology, 9(SI01): 86-90, 2023 Copyright © 2023 International Journal for Modern Trends in Science and Technology ISSN: 2455-3778 online DOI: https://doi.org/10.46501/IJMTST09SI0116

Available online at: http://www.ijmtst.com/vol9si01.html



A Method for Estimating the Likelihood of Receiving a Loan Approval that is primarily based on the Machine Learning

Y. Lakshmi Durga, Dr.V.Suma Avani, M. Vijay Kumar, M. Lakshmi Prasanna

Department of Computer Science and Engineering, Vijaya Institute of Technology for Women, Enikepadu, Vijayawada.

To Cite this Article

Y. Lakshmi Durga, Dr.V.Suma Avani, M. Vijay Kumar, M. Lakshmi Prasanna. A Method for Estimating the Likelihood of Receiving a Loan Approval that is primarily based on the Machine Learning. International Journal for Modern Trends in Science and Technology 2023, 9(SI01), pp. 86-90. <u>https://doi.org/10.46501/IIMTST09SI0116</u>

Article Info

Received: 26 January 2023; Accepted: 22 February 2023; Published: 26 February 2023

ABSTRACT

Getting a loan from a bank has become a relatively normal occurrence these days. In the form of interest, banks benefit from the loans they make to their customers. Many variables should be considered by banks when accepting a loan, including credit history and score, the person's reputation, the location of the property, and the relationship with the bank. Many people seek for loans such as home loans, automobile loans, and other types of loans. On the basis of the aforementioned criteria, no one can beaccepted. There are numerous instances where loan applications are denied by various financial institutions. For banks to maximise revenues, accurate forecasts about whether or not to issue a loan to aconsumer are critical. The goal of this research is to apply machine learning techniques to forecast whether or not acustomer will be able to obtain a loan from a bank.

 ${\it KeyWORDS:} Outlier, Prediction, Component, Training data, and Transformareal lterms that can be used to describe a loan of the second sec$

1.INTRODUCTION

This study used data from prior clients of multiple banks who had loans approved based on a set of criteria. To generate reliable results, the machine learning model is trained on that record. The primary goal of this study is to anticipate the loan's safety [1][3]. Loan safety is estimated using the logistic regression method. The data is cleansed first to avoid missing values in the data collection. Our model was trained using 1500 examples with ten numerical and eight categorical parameters.

Finance companies deals with all kinds of loans such as house loans, vehicle loans, educational loans, personal loans etc... And has a presence across areas such as cities, towns and village areas. A customer applies for a loan first, and then the Finance Company verifies the customer's eligibility for the loan. The applicants must fill out a form that includes information such as their marital status, gender, education, and number of dependents, as well as their income, loan amount, credit history, and other information.Therefore, a robust model is built taking those details as input to verify whether an applicant is eligible to apply for loan or not. The target variable here is Applicants "Loan Status" and the other variables are predictors. After building the Machine Learning model a Web Application is to be developed for a user interface that allows the user to see instantly if he/she is eligible to get a loan by entering the given details.

2. EXISTING SYSTEM

Machine learning methods such as decision tree and random forest are used in the existing system. Although the Random Forest and Decision Tree classifiers provide good efficiency, the Nave Bayes classifier provides excellent results. The lender must manually analyze each application, based on the applicant's primary points such as gender, legal status, education, number of dependents, income, loan amount, credit history, and so on, to determine if they are creditworthy or not, however this method is less accurate.

3. PROPOSED SYSTEM

The proposed model predicts whether a bank will give a loan to a customer. Because the goal of the model is classification, it is built using Logistic Regression using a sigmoid function. Preprocessing is the most time-consuming part of the model, followed by Exploratory Data Analysis, Feature Engineering, and finally Model Selection. Feeding the model with two independent datasets and then preceding the model.

Logistic regression

Logistic regression is a statistical machine learning technique/algorithm for classifying data by attempting to generate a logarithmic line that distinguishes between extreme outcome variables. It is feasible to create predictions using Logistic Regression in this way.

4. RESULTS AND DISCUSSIONS

	Bank	
	Loan Application Form	
Gender Female ~		
Married NO V		
Dependents 0 v		
Education Not Graduate 🗸		
Self_Employed NO V		
ApplicantIncome 0 to 5000		
Any Coapplicantincome NO V		
Expecting Loan Amount Btw 0 to 500		
Loan_Amount_Term In Days		
Credit_History 0 v		
Property_Area Urban		

After entering the values we will get the loan Status based theModel LR.

In this project we are using machine learning algorithm called Random Forest to predict loan eligibility and to train this random forest we are using below dataset

The Lots Mary Co	and Personal Period Tests Persons Mindee Miles	
Tall a la a	27 III BRXIOOSEB HIGH HIGH HIGH	- 5
Oractory Cleanet		-+89+0
vi ciyoor	Loan ID, Gender, Married, Dependents, Education, Self Employed, ApplicantIncome	.CoapplicantIncome.LoanAmount.L
AU *	2 LP001002, Male, No, 0, Graduate, No, 5849, 0, , 360, 1, Urban, Y	11
1301	3 LP001003, Male, Yes, 1, Graduate, No. 4583, 1508, 128, 360, 1, Rural, N	
April21	4 LP001005, Male, Yes, 0, Graduate, Yes, 3000, 0, 66, 360, 1, Urban, Y	
CardiacDisease	5 LP001006.Male.Yes.0.Not Graduate.No.2583.2358.120.360.1.Urban.Y	
Uataset	6 LP001008, Male, No. 0, Graduate, No. 6000, 0, 141, 360, 1, Urban, Y	
	7 LP001011.Male, Yes, 2. Graduate, Yes, 5417, 4196, 267, 360, 1. Urban, Y	
LardiacUseaserred ction.	EP001013.Male.Yes.0.Not Graduate.No.2333.1516.95.360.1.Urban.Y	
SCREENS.docx	9 LP001014, Male, Yes, 3, Graduate, No. 3036, 2504, 158, 360.0, Semiurban, N	
	10 LP001018.Male, Yes, 2, Graduate, No. 4006, 1526, 168, 360, 1, Urban, Y	
	11 LP001020, Male, Yes, 1, Graduate, No. 12841, 10968, 349, 360, 1, Semiurban, N	
	12 LP001024.Male, Yes, 2, Graduate, No. 3200, 700, 70, 360, 1, Urban, Y	
	13 LP001027.Male,Yes,2.Graduate.,2500,1840,109,360,1.Urban,Y	
	14 LP001028, Male, Yes, 2, Graduate, No. 3073, 8106, 200, 360, 1, Urban, Y	
	15 1.0001029 Male No. 0. Graduate No. 1853, 2840, 114, 360, 1. Rural, N	
	16 JP001030, Male, Yes, 2, Graduate, No. 1299, 1086, 17, 120, 1, Urban, Y	
	17 LP001032 Male No. 0 Graduate No. 4950 0 125 360 1 Hirban V	
	18 LP001034 Male. No. 1. Not. Graduate. No. 3596.0.100.240 Urban. V	
	19 J.D001036 Female No. 0 Graduate No. 3510 0. 76 360 0 Urban N	
	20 LD001038 Male Yes 0 Not Graduate No 4887 0 133 360 1 Bural N	
	21 LP001041 Male Yes 0 Graduate, 2600,3500,115, 1 Urban V	
	22 LD001043 Male Yes 0 Not Graduate No 7660 0 104 360 0 Urban N	
	23 ID001046 Malo Yog 1 Craduate No 5955 5625 315 360 1 Urban Y	
	24 LD001047 Male Yes 0 Not Graduate No 2600 1911 116 360 0 Comiurban N	
	25 TD001050 Vos 2 Not Graduate No 3365 1917 112 360 0 Dural N	
	24 ID001052 Malo Yos 1 Craduato 2717 2025 151 260 Comjurban M	
	20 DEVOLUSZ,Male, Les, 1, Graduate, 1, 511, 222, 151, 500, 1 Semiurban, M	
	2 DE001000, Mate, TeS, 0, Gladuate, TeS, 5500, 0, 151, 500, 1, Seminibally 1	Activate Windows
4/I Files (*.*)	(Go to Serbings to activate Windows.
lican-train.cov	* CardiacDiseasePredi * GropYield.py * Main.py * textData.cov * saledoncast.py	
or Help, press F1	In 12	2 col 58 616 00 PC ANSI
O Type her	e to search 🛛 🖟 🖸 🚖 😫 🖉 🦉 🦉 🦉 🖉 🏹	j 🔢 d ² ∧ 💹 d t⊃ 40 1211 📮

In above dataset in first row we can see dataset column names and in other rows we have dataset values and in last column we have class label as Y or N where Y means eligibleand N means not eligible and now we used above dataset to train machine learning model and after training we will upload test dataset and then application will predict class label Y or Nand





In above test data we don't have any N or Y class label and by analysing above records machine learning will predict eligibility.

To run the project, double-click the 'run.bat' file

PREDICTION OF LOAN ELIGBILITY OF THE CUSTOMER				- 0 ×
	PREDICTION 0	F LOAN ELIGIBILITY OF THE	CUSTOMER	
Upload Loan Dataset Preprocess Dataset	Generate Train & Test Data	Run Random Forest ML Model	Predict Eligibility using RF Model	
Random Forest Performance Graph	Exit			
			Activate Window	
			Go to Settings to activ	ate Windows.
O Type here to search	0 🖻 🔒 🤤	4 4 9 8 2 2 2	👱 📶 🔤 👩 🖻 📲 🛄	d0 28-04-2021 🛡

click on 'Upload Loan Dataset' button to load dataset



Select and upload the 'loan-train.csv' file on the above screen, then click the 'Open' button to load the dataset and get the screen below

Uploa	id Loan	Datase	¢		E:/2021/A	arest/April2	1/LounEl	igibility/I	Dataset/loa	a-train.csv						
Prepr	ocess I	Dataset		G	enerate T	rain & Test	Data	Ru	n Random	Forest ML	Model	Predict	Eligibility	using RF Mode	el	
Rando	om For	est Peri	formance G	raph Ex	it											
2021/krest	t/April2	/LoanE	ligibility/Data	iset loan-tra	ain.csv loa	ded										
Loan_ID C	Gender)	farried	Dependents	_ Loan_A	mount_Te	rm Credit_H	istory Pre	perty_Area	Loan_Stat	WS						
LP001002	Male	No	0.0	360.0	1.0	Urban	Y									
LP001003	Male	Yes	1.0	360.0	1.0	Rural	N									
EP001005	Male	Ves	0.0	360.0	1.0	Uman	v									
LP001008	Male	No	0.0	369.0	1.0	Urhan	Ŷ									
rows x 13 o	columns															

In above screen dataset loaded and all columns contains non-numeric values and machine learning will not accept non-numeric values so we need to convert all those values to numeric by assigning ID's to them where MALE will replace with 0 and FEMALE will replace with 1 and below graph showing number of different values in dataset



In above graph different colour lines represents counts of that column and you can see column names with colour in graph top right side. Now click on

'PreprocessDataset'buttontocleandataset

ress Dataset	Generate Train & Test Data	Run Random Forest ML Model	Predict Eligibility using RF Model
a Forest Performance	e Graph Exit		
ried Dependents Educa	ation Loan_Amount_Term Credit_History Pro	operty_Area Loan_Status	
0.0 0	360.0 1.0 2 1		
1.0 0 _	360.0 1.0 0 0		
0.0 1	360.0 1.0 2 1		
0.0 0	360.0 1.0 2 1		
lumns]			
	Forest Performance ied Dependent Edua 0.0 0 1.0 0 0.0 0 0.0 1 0.0 0 warmas]	Image: Second	Extr Extr for operation Extr 00 0.0 0.0 0.0 <t< td=""></t<>

In above screen all non-numeric data is replace with numeric values and now click on 'Generate Train & Test Data' button to split dataset into train and test part

Upload Loan Dataset	E:/2021/krest/April21/LoanEli	igibility/Dataset/loam-train.csv		
Preprocess Dataset	Generate Train & Test Data	Run Random Forest ML Model	Predict Eligibility using RF Model	
Random Forest Performance G	raph Exit			
al records used to test machine learnin	g algorithms are : 123			

In above screen dataset contains 614 records and using 491 records to train ML and 123 records to test ML accuracy. In below graph we can see importance of each attribute with other attribute by using graph correlation metric 🛞 Figure 1

andents	1		COL:	0.10			- 10
Uncome Dep	0.12	1	-0.12		0.045	-0.015	- 08
tincome Applican		-0.12	1	0.19	-0.06	-0.0021	- 06
Amou0bapplicant		0.57	0.19	1	0.039	-0.0084	- 04
nt_Term Loan		0.045	-0.06	0.039	1	0.0015	- 02
IIt_Historyn_Amou	Dependents	ApplicantIncome	Coapplicantincome	LoanAmount	Loan_Amount_Term	1 Credt_History	Activate Windows

In above graph whatever column in x-axis and y-axis having value >0 will be consider as important features or column. Now click on 'Run Random Forest Ml Model' to build random forest model on above dataset

₽

	PREDICTION O	F LOAN ELIGIBILITY OF THE	CUSTOMER
Upload Loan Dataset	E:/2021/krest/April21/LoanElk	gibility/Dataset/Ioan-train.csv	
Preprocess Dataset	Generate Train & Test Data	Run Random Forest ML Model	Predict Eligibility using RF Model
Random Forest Performance Graph	Exit		
ulum Forest Accuracy : 77.2377234 ulum Forest Neisland : 723853065257 ulum Forest Recall : 75.1375806447030 ulum Forest FSCORE : 73.34045326072	S 95		
			Activate Windows Go to Settings to admate Window

In above screen random forest model generated with 77% accuracy and we can see its precision, recall and FSCORE valueand now click on 'Predict Eligibility using RF Model' button to upload test data and perform eligibility prediction



andthenclicking the 'Open' button to load test data will yieldthe following predicted result.

	E. E	ibility/Dataset/loan-train.csv		
Preprocess Dataset	Generate Train & Test Data	Run Random Forest ML Model	Predict Eligibility using RF Model	
Random Forest Performance G	raph Exit			
it Record : [0.00000000e+00 0.000000	00e+00 7.58536872e-04 0.00000000e+00			
10244247e-02 0.00000000e+00 0.00000	143e-01 3.99499000e-02 1000e+00] Sorry! Not Eligible for Loan			
st Record : [0.00000000e+00 0.000000	00e+00 4.64557049e-04 0.00000000e+00			
32278525e-04 9.30507770e-01 3.54457	029e-01 3.90227921e-02			
36202689e-02 2.32278525e-04 2.32278	325e-04] Congratulation! You are Eligible for Lo	228		
est Record : [0.00000000e+00 0.000000	00e+00 5.91892456e-05 0.00000000e+00			
91892456e-05 7.60049103e-01 6.491870	546e-01 2.06570467e-02			
13081284e-02 5.91892456e-05 0.000000	100e+00] Sorry! Not Eligible for Loan			
est Record : [0.00000000e+00 0.000000	00e+00 6.06771234e-04 0.00000000e+00			
02205617-040 20022075-01 212260	932e-01 2.12369932e-02			
#338301/e-04 7./08337/3e-01 2.12307.				
09218822e-01 3.03385617e-04 3.033856	617e-04] Congratulation! You are Eligible for Lo	038		
n9218822e-01 3.03385617e-04 3.033856 st Record : [0.00000008e+00 0.000000	617e-04] Congratulation! You are Eligible for Lo)0e+00 6.39624743e-04 0.00000000e+00	028		

In above screen in square bracket we can see normalized test values and after square bracket we can see the prediction result as eligible or not eligible. You can scroll down above text area to view all predicted records and now click on 'Random Forest Performance Graph' button to getbelow graph



We can notice accuracy in the graph above, precision, recall and FSCORE values of random forest and graph y-axis represents %value where accuracy got 80% and Precision got 65%. Each metric bar colour name you can see from top rightside.

5. CONCLUSION

The prediction process includes cleaning and processing data, imputation of missing values, experimental analysis of the data set, model creation, and testing on test data. On the original data set, the best case accuracy was 0.811 on Data set. The following are the findings after determining that applicants with the lowest credit ratings will be denied loan approval due to a higher risk of defaulting on the loan. Applicants with a high income

and smaller loan requests aremore likely to be approved, which makes sense because they are more likely to repay their loans. Gender and marital status, for example, tend to be overlooked.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

- Toby Segaran, "Programming Collective Intelligence: Building Smart Web 2.0 Applications. "O'Reilly Media.
- [2] Drew Conway and John Myles White," Machine Learning for Hackers: Case Studies and Algorithms to Get you Started," O'Reilly Media.
- [3] Trevor Hastie, Robert Tibshirani, and Jerome Friedman,"The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer, Kindle
- [4] PhilHyoJin Do ,Ho-Jin Choi, "Sentiment analysis of real-life situations using loca- tion, people and time as contextual features," International Conference on Big Data and Smart Computing (BIGCOMP), pp. 39–42. IEEE, 2015.
- [5] Bing Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, May 2012.
- [6] Bing Liu, "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions," Cambridge University Press, ISBN:978-1-107-01789-4.
- [7] Shiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng, "CNN for
- [8] situations understanding based on sentiment analysis of twitter data," Procedia computer science, 111:376–381, 2017.CrossRef.
- [9] K I Rahmani, M.A. Ansari, Amit Kumar Goel, "An Efficient Indexing Algorithm for CBIR,"IEEE- International Conference on Computational Intelligence & Communication Technology ,13-14 Feb 2015.
- [10] Gurlove Singh, Amit Kumar Goel ,"Face Detection and Recognition System using Digital Image Processing", 2nd International conference on Innovative Mechanism for Industry Application ICMIA 2020, 5-7 March 2020, IEEE Publisher.
- [11] Amit Kumar Goel, Kalpana Batra, Poonam Phogat," Manage big data using optical networks", Journal of Statistics and Management Systems "Volume 23, 2020, Issue 2, Taylors & Francis.
- [12] Parvathi, D. S. L., Leelavathi, N., Ravikumar, J. M. S. V., & Sujatha, B. (2020, July). Emotion Analysis Using Deep Learning. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 593-598). IEEE.
- [13] Kumar, J. R., Sujatha, B., &Leelavathi, N. (2021, February). Automatic Vehicle Number Plate Recognition System Using Machine Learning. In IOP Conference Series: Materials Science and Engineering (Vol. 1074, No. 1, p. 012012). IOP Publishing.