
IDENTIFYING CREDIT RISKS IN BANKING DOMAIN USING DECISION TREE ALGORITHM

Dr. B. Rangaswamy*¹, B. Sai Mahesh*², Md. Muskaan*³, P. Sagarika*⁴

*¹Professor, Department Of Computer Science And Engineering, ACE Engineering College,
Hyderabad, Telangana, India.

*^{2,3,4}Department Of Computer Science And Engineering, ACE Engineering College,
Hyderabad, Telangana, India.

DOI: <https://www.doi.org/10.56726/IRJMETS-NCASCTE202223>

ABSTRACT

Data Science can be used in various areas in financial sectors like customer segmentation, profitability, high risk loan applicants, credit analysis, fraudulent transactions, cash management and forecasting operations. There are numerous risks related to bank loans both for the banks and the borrowers getting the loans and the number of transactions banking sectors is rapidly growing and huge data volumes are available. Banks need to analyze their customers for loan eligibility so that they can be specifically target good customers. In this project, an automated the loan eligibility process model have been created, based on customer details such as Gender, Marital Status, Age, Occupation, income, Property etc. The Decision Tree algorithm have been used to create the loan approval classification model. As the number of transactions in banking sector is rapidly growing and huge data volumes are available. Instead of using standardized schemes, this classification model can be used by bank managers for analyzing the customer's behavior easily and the risks around loan can be reduced..

Keywords: Decision Tree, Loan Eligibility, Data Science, Machine Learning.

I. INTRODUCTION

Data Science is one of the most motivating and vital area of research with the aim of extracting information from tremendous amount of accumulated data sets. Going on is an era of simplifying almost all complicated works using computers. Manual processing makes the process slow and other problems such as inconsistency and ambiguity on operations. The analysis of risk in bank loans need understanding what is the meaning of risk. Risk in bank loans involve: credit risk, the risk that the loan won't be return back on time or at all; liquidity risk, the risk that too many deposits will be withdrawn too quickly, leaving the bank short on immediate cash; and interest rate risk, the risk that the interest rates priced on bank loans will be too low to earn the bank adequate money. Data Science techniques aid to distinguish between borrowers who pay back loans at the appointed time from those who don't. It also helps to expect when the borrower is at default, whether providing loan to a particular customer will result in bad loans. All processes related to banking sector could be analyzed using data mining to detect the customer's behavior. It also helps to analyze whether the customer will make prompt or delay payment if the credit cards are sold to them. Many analysis have been conducted based on data science in the field of financial and banking sector. The goal of this project, is to focus on the loan evaluation process. The Decision Tree algorithm have been used to create the loan approval classification model using factors such as Loan_ID, Gender, Married, Education, Self-employed, Applicant Income, Coapplicant Income, Loan Amount, Property_Area etc.,. This model helps the bank manager to make the decision of whether the loan have to be approved or not to a particular customer. So that, Bank manager can classify the customers easily and accurately using this classification model.

II. METHODOLOGY

Of course, sample data is required to support the implementation of machine learning utilising the Decision Tree algorithm. The information about different loan applications based on their configuration is shown in the table below.

Sample information was collected from Kaggle.com.

	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
1	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	0
2	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	1
3	Male	Yes	0	Not Graduate	No	2583	2358	120	360	1	Urban	1
4	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	1
5	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	1
6	Male	Yes	0	Not Graduate	No	2333	1516	95	360	1	Urban	1
7	Male	Yes	5	Graduate	No	3036	2504	158	360	0	Semiurban	0
8	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	1
9	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurban	1
10	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	1
11	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	1
12	Male	No	0	Graduate	No	1853	2840	114	360	1	Rural	0
13	Male	Yes	2	Graduate	No	1299	1086	17	120	1	Urban	1
14	Male	No	0	Graduate	No	4950	0	125	360	1	Urban	1
15	Female	No	0	Graduate	No	3510	0	76	360	0	Urban	0
16	Male	Yes	0	Not Graduate	No	4887	0	133	360	1	Rural	0
17	Male	Yes	0	Not Graduate	No	7660	0	104	360	0	Urban	0
18	Male	Yes	1	Graduate	No	5955	5625	315	360	1	Urban	1
19	Male	Yes	0	Not Graduate	No	2600	1911	116	360	0	Semiurban	0
20	Male	Yes	0	Graduate	Yes	9560	0	191	360	1	Semiurban	1
21	Male	Yes	0	Graduate	No	2799	2253	122	360	1	Semiurban	1
22	Male	Yes	2	Not Graduate	No	4226	1040	110	360	1	Urban	1
23	Male	No	0	Not Graduate	No	1442	0	35	360	1	Urban	0
24	Male	No	0	Graduate	No	3167	0	74	360	1	Urban	0
25	Male	No	1	Graduate	Yes	4692	0	106	360	1	Rural	0
26	Male	Yes	0	Graduate	No	3500	1667	114	360	1	Semiurban	1
27	Male	No	6	Graduate	No	12500	3000	320	360	1	Rural	0
28	Female	Yes	0	Graduate	No	3667	1459	144	360	1	Semiurban	1
29	Male	No	0	Graduate	No	4166	7210	184	360	1	Urban	1
30	Male	No	0	Not Graduate	No	3748	1668	110	360	1	Semiurban	1
31	Male	No	0	Graduate	No	3600	0	80	360	1	Urban	0
32	Male	No	0	Graduate	No	1800	1213	47	360	1	Urban	1
33	Male	Yes	0	Graduate	No	3941	2336	134	360	1	Semiurban	1
34	Male	Yes	1	Graduate	No	5649	0	44	360	1	Urban	1

The algorithm used as a Decision Tree Classifier that can be used to generate a decision based on a sample dataset begins by choosing the highest gain attribute as the tree's root, then creating branches for each value, dividing the cases into branches, and repeating the process for each branch until all the cases in it belong to the same class.

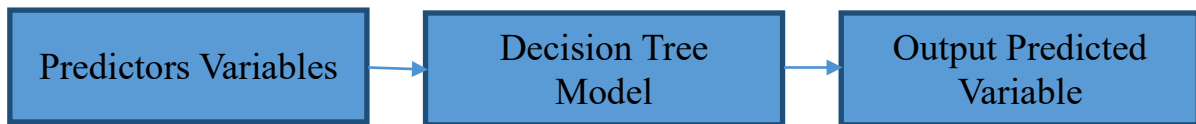


Figure: Flow of design and analysis

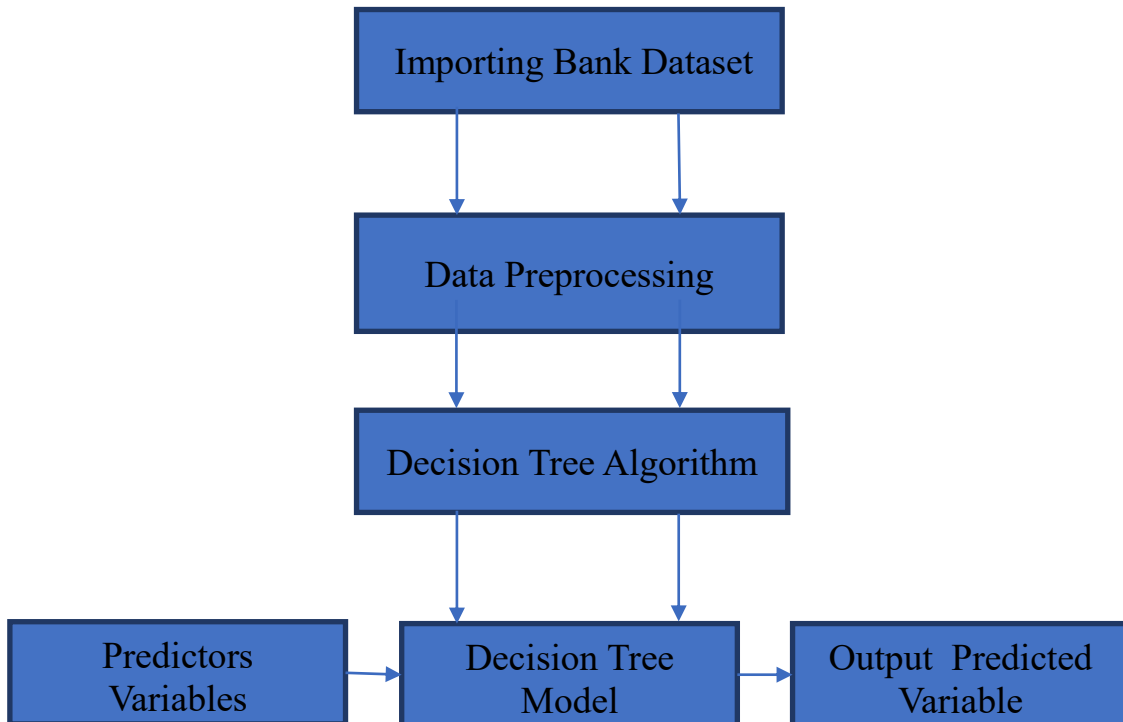


Figure: System Architecture

III. RESULTS AND DISCUSSION

By using decision tree algorithm and by importing bank dataset in r language the decision tree classifies the data and the output predicted and actual is given from 504 variables 296 has given loan approval and 208 has not given the loan

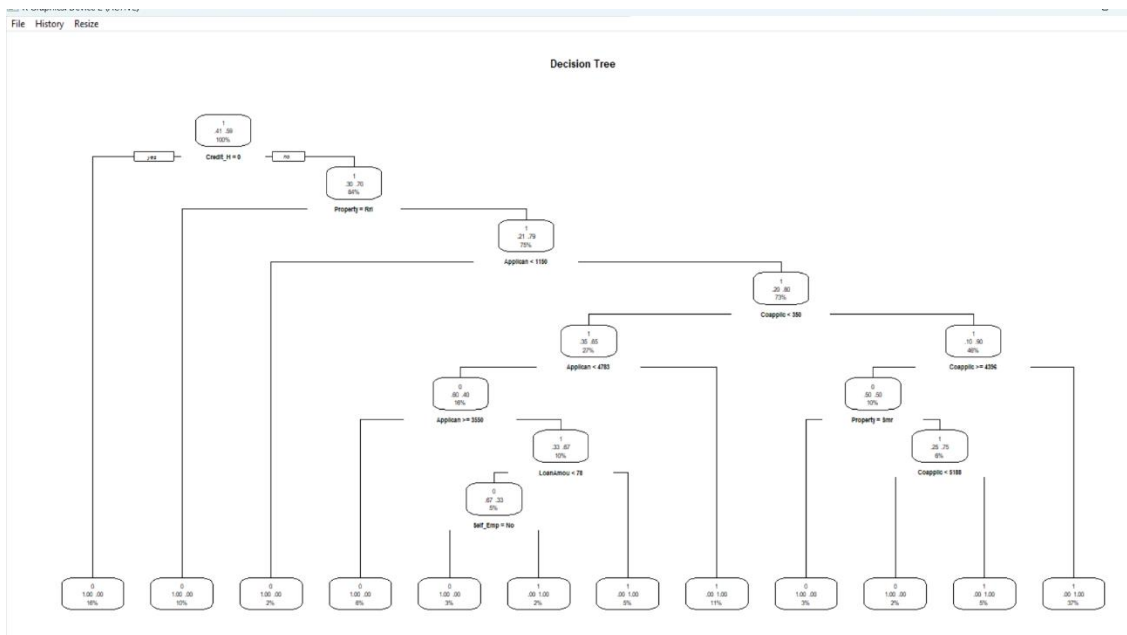


Figure: Decision Tree Classifying the loan dataset

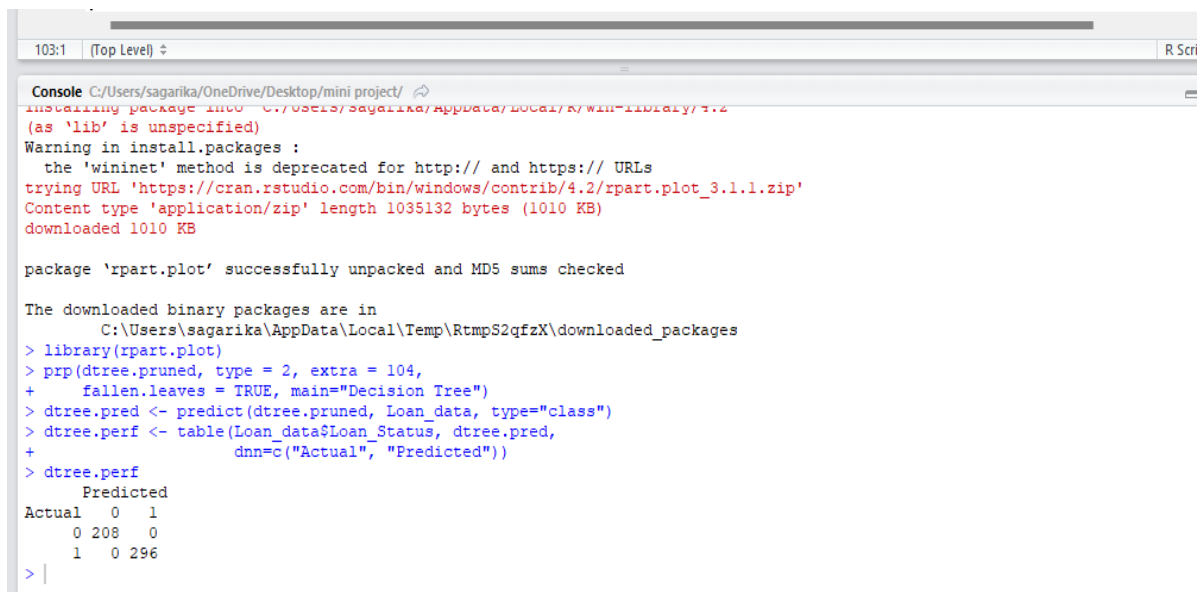


Figure: Predicted values for approving loans

IV. CONCLUSION

It is simple for banks to predict anything through the use of machine learning and the Decision Tree algorithm, the customers are going to pay back their loans or not. Decision Tree algorithm is used to build the classification and predictive model. Decision Tree Algorithm is very simple to understand and use. Accuracy is Increased by using this Decision Tree Algorithm. It is very handy and useful for the Bank Manager. We can Priorly Predict whether, the Customer can repay the Loan by using this Decision Tree Algorithm This model can be used to predict and classify the applications of loans that introduced by the customers to good or bad, by

investigate customer behaviours and previous pay back credit. Using this model bank manager can easily decide, whether to approve the loan request of the bank customers or not.

ACKNOWLEDGEMENTS

The Our sincere thanks to our guide Mrs. Talekar Rohini for his guidance. Also , we are grateful to our project coordinator Mr.PremKumar and Thank Ful to Dr. M.V. Vijay Saradhi, Head of the Department of Computer Science And Engineering, ACE Engineering College for his support and precious time.

V. REFERENCES

- [1] Steenackers A., and Goovaerts M.J. (1989): A credit scoring model for personal loans. Insurance Mathematics & Economics, 8, 31-34.
- [2] Henley W.E., Hand D.E.(1997). Construction of a k-nearest neighbor credit-scoring system. IMA Journal of Mana-gement Mathematics, 8, 305-321.
- [3] Davis, R. H., Edelman, D. B., and Gammerman, A. J.(1992):Machine Learning Algorithms for Credit-Card Applications,. IMA Journal of Mathematics Applied in Business and Industry (4), Pp: 43-51.
- [4] Desai V.S., Crook J.N., Overstreet G.A. Jr(1996): On comparison of neural networks and linear scoring models in the credit union environment. European Journal of Operational Research, 95(1), Pp: 24-37.
- [5] Baesens B., Setiono R., Mues C, Vanthienen J.(2003): Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation, Management Science, Volume 49 , Issue 3, March 2003, Pp:312 – 329.
- [6] West, D.(2000): Neural Network Credit Scoring Models, Computers & Operations Research, vol. 27, no. 11-12, pp. 1131-1152
- [7] Chi G., Hao J., Xiu Ch., Zhu Z. (2001): Cluster Analysis for Weight of Credit Risk Evaluation Index. Systems Engineering-Theory Methodology, Applications, 10(1), Pp. 64-67.
- [8] Lundy M.(1993): Cluster Analysis in Credit Scoring. Credit Scoring and Credit Control. New York: Oxford Uni-versity Press.
- [9] Luo Y.Z., Pang S.L., S.(2003): Fuzzy Cluster in Credit Scoring. Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, 2-5 November 2003, 2731-2736.
- [10] Staefanowski J., and Wilk S(2001): Evaluating Business Credit Risk by Means of Approach – Integrating Decision Rules and Case-Based Learning. International Journal of Intelligent Systems in Accounting, Finance and Management, 10, 97-114.