***Naivigyan Journal of Multidisciplinary Research* (NJMR)**
Volume 1, Issue 1, April 2025, Pages: 26-32
Original Article | ISSN (O): 3108-1487

Implementation of Random Forest and Fuzzy-Based Data Mining Classification Model for the Banking Domain

Soni P M[1]* and Jayakrishnan S[2]

[1]Assistant Professor, Naipunnya Institute of Management and Information Technology, Pongam, Thrissur- 680308, Kerala, India

[2]Associate Professor, Naipunnya Institute of Management and Information Technology, Pongam, Thrissur- 680308, Kerala, India

*Corresponding Author's Email: sonipm@naipunnya.ac.in

Abstract

Most countries rely on the banking industry that very much relies on loans particularly mortgage loans. Personal loans are generally regarded to be risky compared to business loans. To the households and the financial system and consequently to the society as a whole, the credibility forecasts of loan repayment is crucial since the two factors are crucial. Data mining can correct the situation predicting customer behaviour and analysis of historical data. Various tools are used to generate forecasts. The most commonly used modelling method in predicting the ability of a borrower to repay a loan is classification. To classify, one can use various different techniques, and the precision of such algorithms is also different. The primary aim of the paper is to compare the outcomes of the classification performed with the help of random forest algorithm and fuzzy-based modelling.

Keywords: Feature selection, Classification, Accuracy, Performance

Introduction

Data mining methods apply the most sophisticated data analysis tools to detect hitherto unknown, valid patterns and relationships within a large data set [Vivek Bhambri, 2011]. The banking industry can manage customer relationships through various data mining techniques. The different areas in which the banking industry can use data mining tools are customer segmentation, banking profitability, credit scoring and approval, customer payment prediction, marketing, fraud transaction detection, cash management, and forecasting operations [Sudhamathy G, 2016]. Bankers are supposed to be aware of fraudsters because they will create further problems to the banking institution, especially in the financial area. Although banks have so much data relating to consumer behaviour, they cannot identify whether an application is unable to make the payment [Sudhamathy G, 2016]. To address these problems, data mining techniques such as fuzzy-based modelling, classification modelling can be used. One of the data mining techniques is classification, which is used to predict the target class labels [Berson et al., 1999]. Different types of classification algorithms are rule-based, neural network-based, decision tree-based, statistical-based and distance-

based methods. [Chitra and B. Subashini, 2013] Lot_Zadeh proposed the concept of fuzzy logic as the solution to uncertainty in 1965.

Since Zadeh introduced fuzzy logic in 1965, it has been successfully used in numerous disciplines [Chen and Chiou, 1999]. Numerous fields, including operations research, artificial intelligence, medicine, and decision theory, have applications based on fuzzy logic. Fuzzy systems modelling involves a lot of human expertise in the creation of fuzzy rules. The problem of inaccuracy and vagueness of information in fuzzy theory can be represented by two quantities, zero (0) to one (1). It can adequately articulate clouded knowledge of human subjective judgment using a language expression. Fuzzy logic in data mining has been applied to forecast the ability of a customer to repay a loan. Examining the clients' behaviour patterns in connection to their ability to repay the loan, provided it was granted, was deemed essential due to the unpredictable nature of human behaviour. [Joseph Kobina and James Ben,2014].

The structure of this document is as follows. The dataset utilised for the experiments and the methods are described in the following section. The experiment's technique is covered in the section are Fuzzy logic and random forest classification. The findings and discussion are shown in the next session, followed by the conclusion and references.

Methods and Materials

Information was gathered from a leading cooperative bank that offers different types of loan amounts to individuals, businesses, and other entities to satisfy the needs of all kinds of clients. The necessary information was obtained by interviewing the banking officer and observing the site. A thorough analysis of financial transactions and loan processing was also conducted for the same. The data set includes 15,000 mortgage loan customer details. The data collected towards the mining process may contain noise, discrepancies or have missing values. Consequently, the mining process creates conflicting information. The process involving high-quality data will deliver good results from data mining. The data acquired will require pre-processing to improve the efficiency of the data mining process, data quality and finally the outcome of the mining. Data preprocessing is one of the most significant tasks in the data mining process, as it is concerned with the preparation of the original data set and its conversion into the final one. Some of the techniques used in changing the original data set to the final data set are data cleansing, data reduction, data transformation, and data integration. These are also called preprocessing techniques of data. After pre-processing data, it is possible to apply data mining methods such as classification or fuzzy-based modelling to the data to determine patterns and obtain insights. This information can be used by bank executives to assist them in making the correct decisions.

Classification is a data mining technique that assists in assigning a new data record to one of the target classes within a given data set [Berson et.al, 19993] stated that Classification aims to map a data item into one of several predefined categorical classes. For conducting this experiment, a random forest algorithm and a fuzzy-based classification algorithm were considered. One effective tree-learning method in machine learning is the Random Forest algorithm. During the training stage, it generates a number of Decision Trees.

Table 1- Credit Dataset

Sl	Attribute	Datatype
1	Loan No.	object
2	Loan Date	datetime64[ns]
3	Due date	datetime64[ns]
4	Loan amount	int64
5	Opening	int65
6	Payment	int66
7	Receipt	int67
8	int_rcvd	float64
9	fine_rcvd	float65
10	Mem No	object
11	action	object
12	secured	object
13	Loan Balance	int64
14	interest Rate	float64
15	Category	object
16	Purpose	object
17	gender	object
18	Occupation	object

To measure a random subset of characteristics in each partition, a random subset of the data set is used to build each tree. By introducing variety across individual trees, this randomness lowers the possibility of overfitting and enhances prediction performance overall. The algorithm combines the output of every tree in prediction, either by average (for regression tasks) or voting (for classification tasks). An example of reliable and accurate findings is provided by this cooperative decision-making process, which is aided by the insights of several trees.

Table 2 - Random Forest Algorithm

Step 1:	Choose randomly a combination of “k” features out of the entire “m” features.
Step 2:	In which **k << m**
Step 3:	Choose an ideal split point to determine the node **"d"** among the **"k"** features.
Step 4:	Split the node into **daughter nodes** using the **best split**
Step 5:	Repeat steps **1- 3** until the “l” number of nodes is obtained.
Step 6:	Build a forest by repeating steps **1-4** to create “n” trees.
Step 7:	End

For classification and regression tasks, random forests are frequently utilised because of their reputation for handling complex data, lowering. Table 2 provides a short explanation of the way the random forest classification algorithm is performed. Fuzzification is the process of transforming of the numeric values into membership functions. The membership functions can be described by using the three terms of language, that is, short/ medium/ long, bad/average/ good and low/ medium/ high. Subjectively, it can be accomplished through the inclusion of extra words of the language by five or seven to achieve excellent precision. A membership function is a function that indicates the degree to which an input belongs to a set. Table 3 briefly explains the steps for the fuzzification process.

Table 3 - Steps for fuzzy-based classification

Step 1 :	Read input values from the user and put into the database.
Step 2 :	Apply Fuzzification on the database to get the fuzzified input
Step 3 :	Using fuzzified input and inference rules, process the inference system for loan credibility prediction
Step 4 :	Apply Defuzzification to convert fuzzified output to user understandable output

The most important features and their corresponding feature importance that have been selected for the fuzzy classification algorithm are represented in Table 4.

Table 4 - Most important Features

Feature	Feature importance
int_rcvd	0.1418
opening	0.1363
fine_rcvd	0.1241
interest rate	0.1129

The output value of an information of a membership function is always between 0 and 1. In fuzzy logic, the output value of a membership function is also termed as membership grade. Figure 2 portraits the values of membership function for the feature “fine_rcvd”. The three grades are poor, average and good. Figure 3 represents the code in python to generate fuzzy rules.

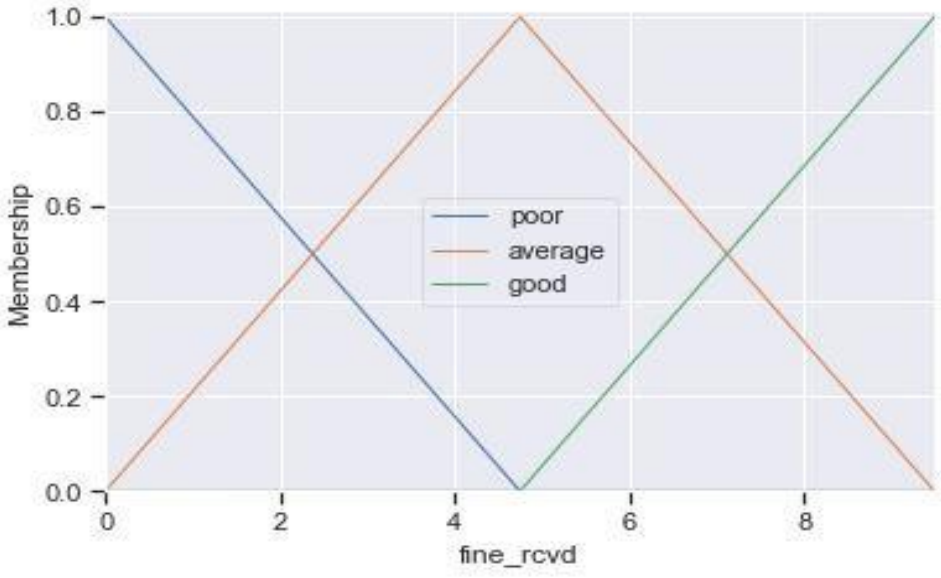

Figure 2 - Membership function for "fine _rcvd"

```
rule1 = ctrl.Rule (interest_rate ['good'] | fine_
rcvd ['good']| int_rcvd ['good']| days ['good'],
secured ['NO'])
rule2 = ctrl.Rule (interest_rate ['poor'] | fine_
rcvd ['poor']| int_rcvd ['poor']| days ['poor'],
secured ['YES'])
rule3 = ctrl.Rule (interest_rate ['average'] | fin
e_rcvd ['average']| int_rcvd ['average']| days ['a
verage'], secured ['YES'])
loan_ctrl = ctrl.ControlSystem ([rule1, rule2])
loan = ctrl.ControlSystemSimulation (loan_ctrl)
```

Figure 3 - Fuzzy Rules

Results and Discussions

The purpose of the experiment is to make a comparative analysis of the two points of the categorisation models, i.e., fuzzy-based modelling and random forest algorithm. Table 5 shows the accuracy which is attained by the two classifiers, and Figure 4 shows a graphical representation of the outcome.

Classifiers	**Accuracy (%)**
Random Forest	99.1
Fuzzy	84.6

Table 5 - Accuracy

Figure 4 depicts the input values given to the fuzzy classifier for evaluating the loan credibility behaviour of a customer.

interest_rate =16

interest_rcvd = 16

fine_rcvd = 12

days =1210

Figure 4 - Fuzzy Input

After giving the input values to the fuzzy classifier, it displayed the output as in Figure 5, which represents the probability of approving or rejecting a loan.

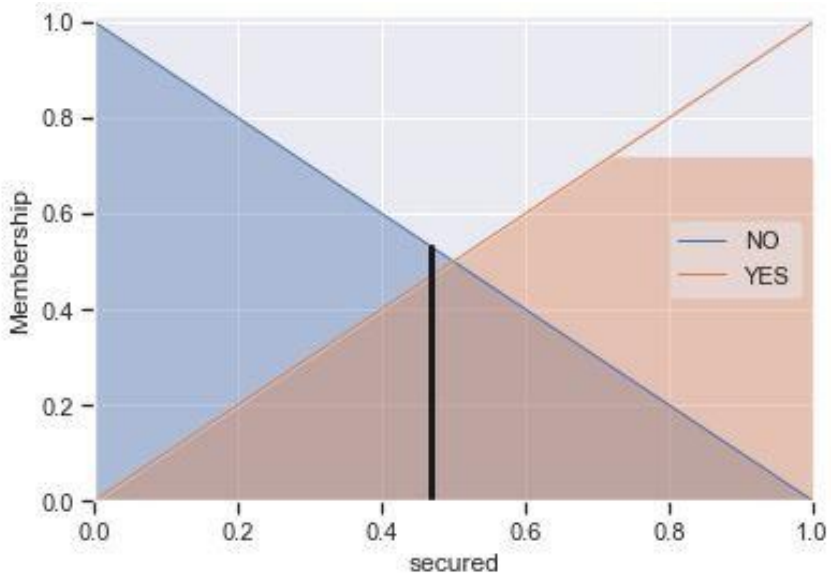

Figure 5- Probability of approving or rejecting the loan

Conclusion

The study compared the fuzzy-based and random forest methods of classification in data mining to predict the likelihood of a banking customer repaying a loan, especially in the case of a mortgage loan. Bank officers are confronted with a big problem as to whether to give loan approval to loan applicants or not. The research has now proven that the two strategies can assist the bank staff to make more accurate choices.

References

Berson, A., Smith, S., and Thearling, K. (1999). Building Data Mining Applications for CRM. McGraw-Hill, New York

Bhambri, Vivek (2011). "Application of Data Mining in Banking Sector", *IJCSt* Vol. 2, Issue 2, June 2011, ISSN : 2229-4333(Print) | ISSN : 0976-8491(Online)

Chen, L-H., Chiou, T-W, (1999), "A fuzzy credit rating approach for commercial loans A Taiwan case". *Omega International Journal of Management Science* 27(1999), 407-419, 1999, https://doi.org/10.1016/S0305-0483(98)00051-6

Chitra, K. and B.Subashini (2013). "An Efficient Algorithm for Detecting Credit Card Frauds", *Proceedings of State Level Seminar on Emerging Trends in Banking Industry*, March 2013.

Joseph Kobina Panford , James Ben Hayfron-Acquah (2014). " Fuzzy Logic Approach to Credit Scoring for Micro Finance in Ghana: A Case Study of KWIQPLUS Money Lending" *International Journal of Computer Applications*, May 2014

Omaia Al-Omari1, Nazlia Omari (2019). "Enhanced Document Classification Using Noun Verb(NV) Terms Extraction Approach", *International Journal of Advanced Trends in Computer Science and Engineering* (IJATCSE), Vol.8, No.1, January –February 2019 https://doi.org/10.30534/ijatcse/2019/26822019

Sri Hari Nallamala, Dr. Pragnyaban Mishra, Dr. Suvarna VaniKoneru (2019). "Qualitative Metrics on Breast Cancer DiagnosiswithNeuro Fuzzy Inference Systems", *International Journal of Advanced Trends in Computer Science and Engineering* (IJATCSE), Vol. 8, No.2, March-April 2019.

Sudhamathy G. (2026). "Credit Risk Analysis and Prediction Modelling of Bank Loans Using R" Vol 8 No 5 Oct- Nov 2016 https://doi.org/10.21817/ijet/2016/v8i5/160805414

Yi-Chung Hu a, Ruey-Shun Chen a, Gwo-Hshiung Tzeng (2003), "Finding fuzzy classification rules using data mining techniques", *Pattern Recognition Letters* 24 (2003), 509 –519 https://doi.org/10.1016/S0167-8655(02)00273-8