ISSN 2278-3091
Volume 8, No.3, May - June 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse74832019.pdf
https://doi.org/10.30534/ijatcse/2019/74832019

A Fuzzy based Data mining Approach for the Loan Credibility Prediction System in Co-operative Banking Sector

Soni P M[1], Varghese Paul[2]
[1] Research Scholar, Bharathiyar University, Coimbatore, India, sonipm@sngist.org
[2] Professor, Dept.of IT, CUSAT, Kerala, India, varghesep@rajagiritech.ac.in

ABSTRACT

Loans especially mortgage loans play a vital role in the banking industry of many countries. Loans to individuals are considered as more risky than business loans in many situations. Due to these two reasons, the efficiency of prediction of loan repayment credibility is important for the welfare of both households and banking system and hence it is also important for the entire society. Data mining can solve the problem by analysing historical data and predict the behaviour of a customer. For prediction purpose various tools are used. Classification is the popular modeling technique to predict the loan repayment capability of a customer. Classification can be performed using different algorithms and accuracy also varies with the algorithms. Fuzzy logic system maps the nonlinear an input data set to a scalar output data. Fuzzy logic system consists of four main parts such as Fuzzification, Rules, Inference engine and Defuzzification. In this research, a fuzzy based loan repayment capability prediction model is introduced. The main objective of this paper is to implement fuzzy model that incorporates artificial intelligence approach for the prediction. The experiments were being done using python programming and the data was selected from a premier banking institution.

Key words*:* feature selection, classification, accuracy, performance

1. INTRODUCTION

Data mining technique involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data set. [1]. Today, Customer Relationship Management in banking industry is purely based on Data Mining techniques. The different areas in which Data mining Tools can be used in the banking industry are customer segmentation, Banking profitability, credit scoring and approval, Predicting payment from Customers, Marketing, Detecting fraud transactions and Cash management and Forecasting operations [2]. The bankers should vigilant to fraudsters because they will create more problems to the banking organization especially in the finance.. Banks hold huge volumes of customer behavior related data from which they are unable to arrive at a judgment if an applicant can be defaulter or not [2]. The approaches of machine learning to general AI are distinguished by data utilization and data patterns' discovery, and the application of machine learning can be seen in many areas, including weather forecast, fraud detection and medical diagnosis[8]. Classification is one of the data analysis method that predict the class labels [3]. There are more classification methods such as statistical based, distance based, decision tree based, neural network based, rule based [4]. Fuzzy logic was introduced by Lot_ Zadeh in 1965, to deal with uncertainty.

Fuzzy logic has successful application in many fields since Zadeh presented the theory in 1965[5]. There are many fuzzy based applications such as in medicine, artificial intelligence, decision theory and operations research. Fuzzy systems modeling incorporate a certain amount of human knowledge for creating fuzzy rules. Fuzzy theory deals with imprecision and vagueness and represent information from Zero (0) to One (1). It can effectively describe imprecise knowledge of human subjective judgment by linguistic term. Fuzzy logic has also been applied in predicting the loan repayment capability of a customer. It was assumed that, given the uncertainty of human behavior, it is very important to study the behavioral pattern of the clients with respect to their ability to repay the loan, if granted [6]

This paper is organized as follows. The next section explains about the dataset and data mining tools used for conducting the experiment. Section 3 discusses about the proposed experiment. Section 4 demonstrates the modeling using fuzzy logic. Section 5 explains about the results and discussion. Conclusion is given in section 6 followed by acknowledgement and references.

2. DATA SET AND TOOLS

Data was collected from a premier cooperative bank that provides loans to individuals, business firms, etc so as to meet the requirements of all type of customers[10]. On site observation and interview with the banking officer helped me to collect the relevant data]. A detailed study about the loan processing and banking transactions are also made for the same[10]. The data set consists of 15000 customer details of mortgage loan . In order to conduct the experiment Python programming was used.

Table 1: Attribute list of original 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

3. PROPOSED EXPERIMENT

The three main process involved in the fuzzy logic computation are fuzzification of inputs, fuzzy inference associated with the rule base and Defuzzification. The user interface is the first stage where the user interacts with the system. The graphical user interface helps the user to communicate with system effectively. The second stage is database that contains information about the customers of the bank. The database is normalized and efficiently supports queries and adhoc queries posed to the system. Third stage of the model is the Fuzzification process. The fuzzifier picks the data from the bank, converts it into fuzzy values and keeps it in the system for further processing. The rule base contains the various rules constructed in consultation with a knowledge expert. These rules form the actual working rules for the fuzzy controller. Based on the input and output membership functions, numbers of rules were generated. The fourth stage named as Inference system is the actual process of mapping with a given set of input variables and output through a set of fuzzy rules. Fuzzy rule includes statements of "IF.THEN." with two parts. The first part that starts with IF and ends before the THEN is referred to as the predicate (premise, antecedent) which combines in a harmonious manner the subsets of input variables. Consequent part comes after "THEN" which includes the convenient fuzzy subset of the output based on the premise part. This implies that there is a set of rules which is valid for a specific portion of the inputs variation domain. A fuzzy classification rule is a fuzzy if–then rule whose consequent part is a class label [7]

Defuzzification is the final stage of the model. The result obtained from the loan credibility prediction inference system is purely fuzzified output and therefore it is necessary to defuzzify the output for obtaining a value that can be understood by the user.

The architecture is also termed as FLRCPM (Fuzzy based Loan Repayment Capability Prediction Model) accurately using new proposed feature selection algorithm using wrapper model and fisher score concept.

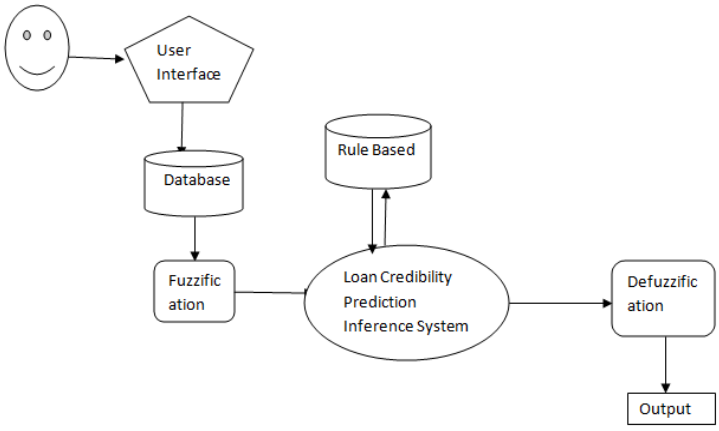

Figure 1: Block diagram of FLRCPM

Table 2 explains about the list of operations that is to be conducted in the experiment.

Table 2: Steps for the proposed model

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

Fuzzification is the process of conversion of the input numerical values into membership functions. The membership functions can be represented by three linguistic terms such as low/medium/high or bad/average/good or short/medium/long. The number of linguistic terms can be increased subjectively say 5 or 7 that will lead to high accuracy. Membership function can be defined as a function that specifies the degree to which a given input belongs to a set. The output value of a membership function is always in between 0 and 1. The output value is also known as a membership value or membership grade.

Membership functions are used in the Fuzzification and defuzzification steps of a FLRCPM to map the non-fuzzy input values to fuzzy linguistic terms and vice versa. The different types of membership functions are Triangular, Trapezoidal, Piecewise linear, Gaussian and Singleton .The grade of membership of x in class A is represented by the value $f_A(x)$, 1 is for complete membership and 0 is for non-membership[9].

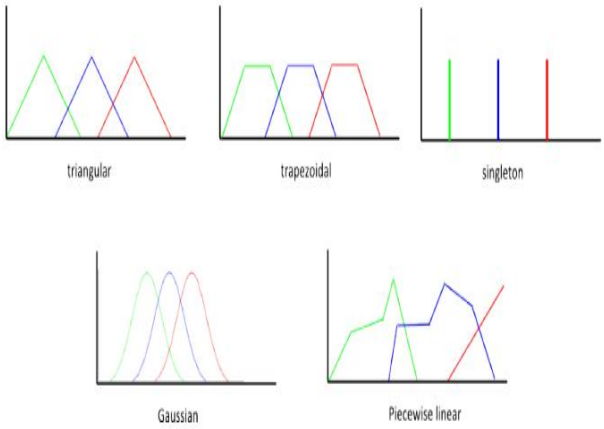

Figure 2: Types of membership functions

4. FUZZY MODELLING

For the implementation of FLRCPM, four features were selected based on their feature importance. Table 3 displays all features with their feature importance and Table 4 displays the four important features with their feature importance value. The feature importance is obtained by applying random forest feature selection method. To create fuzzy model, membership function for the important features are generated

Table 3: Feature importance of the original dataset

int_rcvd	0.141848
Opening	0.136317
fine_rcvd	0.124183
interest Rate	0.112922
Receipt	0.087235
Purpose	0.083576
Occupation	0.083099
Loan amount	0.082047
action	0.042612
days	0.036825
Payment	0.025731
gender	0.024916
Loan Balance	0.018689

Table 4: Important features selected for fuzzy model

# int_rcvd	0.141848
# Opening	0.136317
# fine_rcvd	0.124183
# interest Rate	0.112922

loan_df ['interest Rate'].describe ( ) command produced the statistical analysis of the feature interest Rate and it is displayed in table 5. The main components are count, mean, standard deviation minimum and maximum. This helps to know more about each feature statistically.

Table 5: Statistical analysis of the feature interest rate

count	3406.000000
mean	14.741045
std	2.128273
min	11.000000
25%	13.000000
50%	14.000000
75%	17.000000
max	18.500000
Name: interest Rate	

The most convenient way to take a quick look at a univariate distribution in seaborn is the `distplot()` function. By default, the function will draw a histogram and fit a kernel density estimate (KDE). The distribution of predictors of the feature interest rate is depicted in figure 3. The python code for the same is written in table 6.

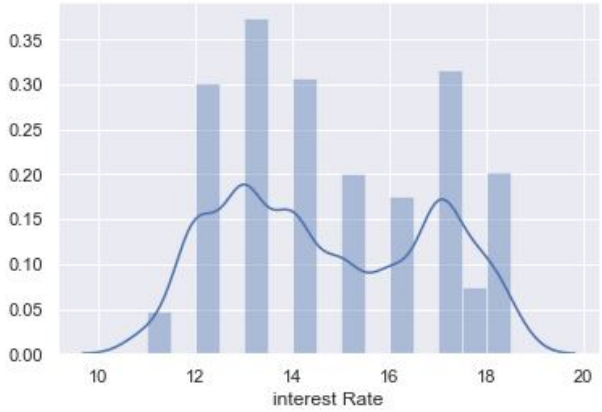

Figure 3: Distribution of predictors of feature interest rate

Table 6: Code segment for interest rate distribution graph

```
import seaborn as sns, numpy as np
sns.set(); np.random.seed(0)
sns.distplot(loan_df['interest Rate'])
```

The classification of customers into secured or unsecured using the feature interest rate can be done using the code segment of the table 7. The corresponding diagram is displayed in figure 4

Table 7: Code segment secured or unsecured

```
import matplotlib.pyplot as plt
%matplotlib inline
g = sns.FacetGrid(loan_df, col='secured')
g.map(plt.hist, 'interest Rate', bins=20)
```

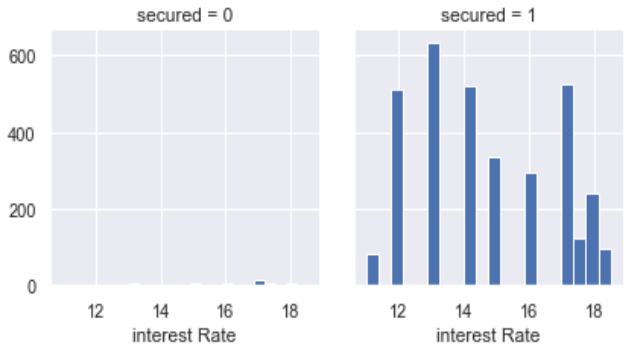

Figure 4: Classification based on interest rate

.

In fuzzy based approach, Membership functions are used to solve practical problems by experience rather than k knowledge. It represents the degree of truth. Here the membership function leads to develop a model that helps the banking officials to predict the loan repayment capability of a customer. Fuzzy based Membership functions are generally represented by graphical forms. The three different values taken by the most important features are poor, average and good. The class label secured produces two values say yes or no. Figure 5 is the graphical representation of membership function for "secured". Figure 6 is the graphical representation of membership function for "fine_rcvd". Figure 7 is the graphical representation of membership function for "interest_rate". Figure 8 is the graphical representation of membership function for "days". Figure 9 is the graphical representation of membership function for "int_rcvd".

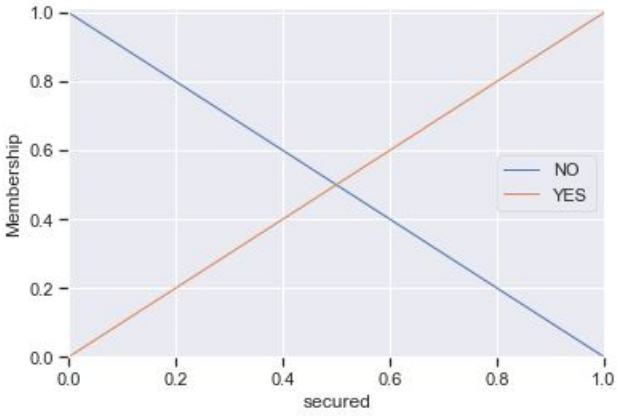

Figure 5: Membership function for "secured"

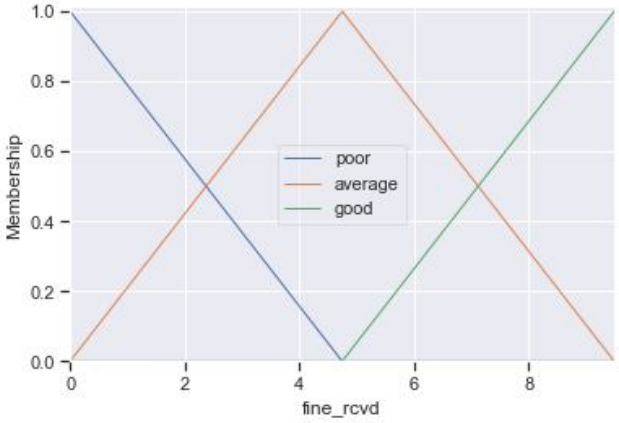

Figure 6: Membership function for "fine_rcvd"

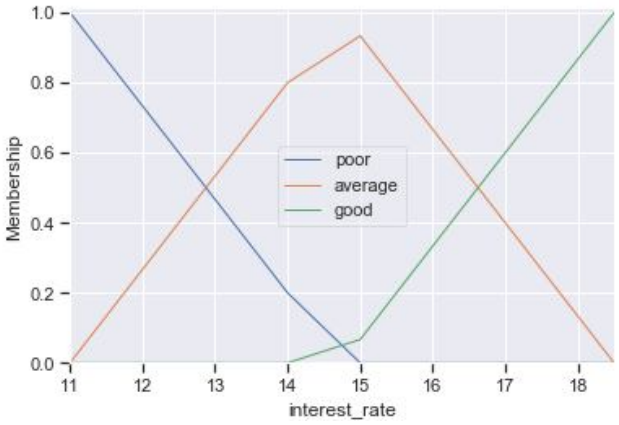

Figure 7: Membership function for "interest_rate"

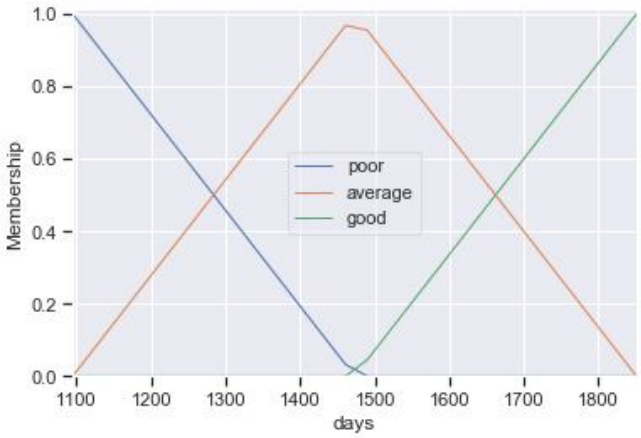

Figure 8: Membership function for "days"

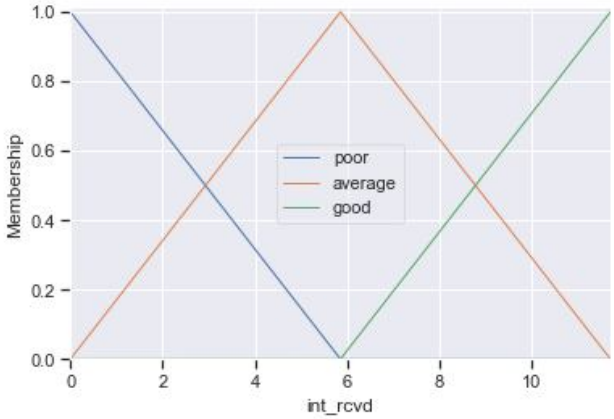

Figure 9: Membership function for "int_rcvd"

The generation of Fuzzy Rules begins with an empty Rule and construct interactively by setting *.antecedent*, *.consequent*, and *.label* variables. The code segment for defining the inference rules using fuzzy logic for predicting the loan behavior of a customer depicted in figure 10.

```
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 10: Code segment for defining rule in Python.

5. RESULTS AND DISCUSSION

Fuzzy based Loan Repayment Capability Prediction Model can help the banking officials to predict the loan repayment capability of a customer on mortgage loan data of one of the co-operative banks in India.

To pass inputs to the Control System using Antecedent labels with Pythonic API the following code segment in figure 11 can be applied

```
loan.input['interest_rate'] = 18
loan.input['int_rcvd'] = 18
loan.input['fine_rcvd'] = 11
loan.input['days'] = 1200
loan.compute()
```

Figure 11: Input to the fuzzy control system

The probability of approving or rejecting the loan based on the above input values are obtained as 0.46 and is depicted in figure 12.

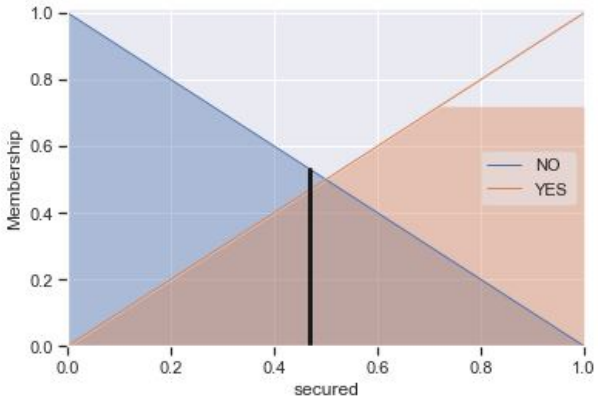

Figure 12: Probability of approving the loan

6. CONCLUSION

In this paper, fuzzy based approach is proposed to predict the loan repayment capability of a customer in a co-operative bank especially in the case of mortgage loans. It is very difficult for the bank officers to determine whether to approve loan applicants or not. Now the experiment has proved that a model that uses fuzzy based classification can help the bank officers to take decisions more accurately. This proposed methodology uses fuzzy based inference rules for identifying the customers whose repayment capability status is safe or risky. FLRCPM makes use of fuzzy logic in order to assign a partial membership degree to each attribute in the mortgage loan dataset that leads to more accurated result.

ACKNOWLEDGMENT

I would like to thank my research guide Dr. Varghese Paul, Professor in IT, Cochin University of Science and Technology, who assisted me to do the research work in a successful manner.

REFERENCES

1. Vivek Bhambri "**Application of Data Mining in Banking Sector**", *IJCSt* Vol. 2, ISSue 2, June 2011, ISSN : 2229-4333(Print) | ISSN : 0976-8491(Online)
2. L. Torgo, "**Functions and data for data mining with r"** *R package version* 0.2.3, 2012
3. Sudhamathy G. "**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
4. K. Chitra, B.Subashini, "**An Efficient Algorithm for Detecting Credit Card Frauds**", *Proceedings of State Level Seminar on Emerging Trends in Banking Industry*, March 2013
5. Chen, L-H., Chiou, T-W, (1999), "**A fuzzy credit rating approach for commercial loans A Taiwan cas**e". *Omega International Journal of Management Science* 27(1999), 407-419, 1999., https://doi.org/10.1016/S0305-0483(98)00051-6
6. Joseph Kobina Panford , James Ben Hayfron-Acquah " **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
7. Yi-Chung Hu a, Ruey-Shun Chen a, Gwo-Hshiung Tzeng, "**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
8. Omaia Al-Omari1, Nazlia Omari "**Enhanced Document Classification Using Noun Verb(NV)Terms Extraction Approach**" *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE),* Volume 8, No.1, January –February 2019 https://doi.org/10.30534/ijatcse/2019/26822019
9. Sri Hari Nallamala, Dr. Pragnyaban Mishra, Dr. Suvarna VaniKoneru, "**Qualitative Metrics on Breast Cancer DiagnosiswithNeuro Fuzzy Inference Systems" ,** *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE),* Volume 8, No.2, March-April 2019.