Effective Prediction of Bankruptcy based on the Qualitative factors using FID3 Algorithm

A. Martin
Research Scholar
Department of Banking Technology
Pondicherry University,
Puducherry, India

S.Balaji
PG Student
Department of CSE
Sri Manakula Vinayagar Engineering College,
Puducherry.

V. Prasanna Venkatesan
Associate Professor
Department of Banking Technology
Pondicherry University,
Puducherry.

ABSTRACT
Bankruptcy is one of the most important issues in Financial Management and investment. Numerous studies on Bankruptcy Prediction have been carried out considering Quantitative factors and they applied different techniques on it to predict Bankruptcy, while only fewer studies have proposed and considered Qualitative factors for prediction of Bankruptcy and even then failure of bankruptcy persists. This paper proposes a model involving Experts decision and Fuzzy ID based algorithm to predict Bankruptcy in an effective manner. In Fuzzy ID3 the evaluation of Entropy and Information Gain helps to rank the qualitative parameters and the membership function evaluation is used to generate prediction rules in qualitative Bankruptcy prediction. The result of the prediction provides the most important factors that have more impact on the Bankruptcy. Since, the prediction is carried out with the experts listed factors the prediction accuracy is raised along with better performance.

General Terms
Data Mining, Bankruptcy, Prediction, Fuzzy Algorithm.

Keywords
Bankruptcy Prediction, Qualitative factors, Fuzzy ID, Information Gain.

1. INTRODUCTION
Bankruptcy prediction is one of the most important research issues in finance, giving a great potential to reduce the risk involved in the financial process. The challenging task is developing an effective prediction model to estimate the potential distress of companies.

Diverse studies on bankruptcy prediction only use quantitative approach along with the financial variables for predictions. But, only fewer studies on bankruptcy prediction have reported about the qualitative approaches. Qualitative data are subjective and difficult to measure. They associate to things such as management expertise, business location, product innovation, product development, etc. Because qualitative data are not easy to assess, several models have been developed to assist business leaders in knowing what kind of information should be pertained and how it should be evaluated. And also due to its subjective nature, it is difficult to understand how or what type of qualitative information is related to success. Moreover, there is only limited research evaluating whether qualitative measures are related with key success factors or not. A foremost and yet unanswered question is whether qualitative information, where properly measured and analyzed and whether they can be used to measure a firm's success or not. The next but related question is, which qualitative information should be measured for prediction. Here, for bankruptcy prediction we use FID3 algorithm.The use of fuzzy sets to partition universes can have significant advantages over more traditional approaches and when combined with classical decision tree induction methods (ID3) can help to address many of the difficulties involved in financial distress problems. The ID3 algorithm introduced by Quinlan [13] has proved to be an effective and popular method for finding decision tree rules to express information contained implicitly in discrete valued data sets.

In this paper the first part deals with the factors involved in bankruptcy and their importance in prediction and the later part of the paper deals with the technique for prediction of the bankruptcy.

2. QUALITATIVE IN CORPORATE BANKRUPTCY PREDICTION
Qualitative information which is used for the default risk estimation process consists of various risk components. Qualitative information is subjective, so based on the previous experiences; the consulted person can have keen knowledge about it. As the type and characteristics of the firm gets changed, the qualitative factors also changes and so the framed rules also gets changed. So, for the qualitative problem-solving, we need knowledge from expert's decisions. They are based on the qualitative factors or parameters we are using in the prediction. So the success depends on both the framing of the appropriate rules with the appropriate parameters.

Qualitative factors that we analyzed from the previous researches are involved for prediction. The prediction performance varies depending upon the selected qualitative parametric values. So, the selection of parameters has an importance in qualitative bankruptcy prediction.There are only fewer studies on qualitative bankruptcy using various methods such as Electri Tri[5] method and New Venture Template Model (NVM) which uses the parameters such as Operational Risk, Competitiveness, Credibility Analysis, Training of employees, Quality, etc.By using the above mentioned models other qualitative parameters such as Innovation, value, persistence, scarce, prevent, flex for Qualitative bankruptcy prediction.

The following components can be categorized as follows: Industry risk, Management risk, Financial flexibility, Credibility, Competitiveness, and Operating risk which are in short represented as IR, MR, FF, CR, CO AND OP. IR is evaluated by the stability, growth of the industry, the degree of competition, and the overall conditions of the industry. MR is involved with the efficiency, stability of management and organization structure. It is evaluated by ability of the management, stability of top management, stability of organization structure, management performance, and

feasibilities of business plans. FF denotes the firm financing ability either from direct or indirect financial market and other sources. CR is involved with the reputation of a company associated with credit history, reliability of information provided by the company, and the relationship with the financial institutions. CO means the degree of competitive advantage determined by market position and the capacity of core technology. OP is the volatility and stability of procurement, the efficiency of production, the stability of sales, and the efficiency of collection policy of accounts receivable.

2.1 Experts Decision in Bankruptcy Prediction

Experts use their subjective knowledge to handle qualitative information in the following analysis steps: preliminary analysis, financial analysis, industry level analysis, financing strategy evaluation, and default risk estimation [2][12]. Qualitative information that is used for the prediction process consists of numerous risk components. Experts evaluate the qualitative risk factors through the risk estimation process using their subjective knowledge and they assign appropriate levels such as positive level, average and negative levels to these factors. Finally, they classify various cases into categories such as high, low and reject. Here the experts' decision is carried out by providing a set of questionnaire and the questionnaire is given to the experts for decision making and they judge the factors, the part of questionnaire is shown in table 1, where the factors are rated by value 1. And the resulting factors and values are used for further evaluation processes.

Table 1. Questions provided to and rated by Experts

FACTORS	QUESTION	SBI< 50% YES	SBI> 50% YES	NO
Financial Flexibility	Caused change/ affected		1	
Direct financing	caused change/ affected		1	
Indirect financing	Caused change/ affected		1	
Other financing (Affiliates, Owner, Third parties)	Caused change/ affected			1
Credibility	Caused change/ affected	1		
Credit history	Caused change/ affected	1		
The reliability of information	Caused change/ affected	1		
The relationship with financial institutes	Caused change/ affected			1

3. FUZZY IN BANKRUPTCY

Fuzzy logic mathematically has truth values ranging between 0 and 1. Fuzzy logic corresponds to the degrees of truth.There exist many methods to do decision analysis. Each method has its own advantages and also its disadvantages. In machine learning, decision tree learning is one of the most popular techniques for making classifications decisions. The fuzzy Interactive Dichotomizer3 is based on a fuzzy implementation of the ID3 algorithm. One of the classification algorithm used in case of evaluating uncertainty data is Fuzzy Interactive Dichotomizer 3 Algorithm.

Interactive Dichotomizer 3 (ID3 for short) algorithm is one of the most used algorithms in machine learning and data mining due to its simplicity to use and effectiveness. It builds a decision tree from some fixed or historic symbolic data in order to learn, classify and to predict the classification of new data. The data must have several attributes with different values. In the meantime, the data must belong to several predefined, discrete classes (i.e. Yes/No). Decision tree chooses the attributes for decision making by using information gain (IG).

3.1 Steps in FID3 Algorithm

ID3 (Examples, Target-attribute, and Attributes) were Examples denotes the training examples. Target- attribute is the attribute whose value is to be predicted from the tree. Attributes are a list of other attributes that may be tested by learned decision tree.

1. Generate a root node for the tree with the membership functions with value 1.
2. In case, if a node t with a fuzzy set of data satisfies the following conditions, then it is a leaf node and it is assigned by the class name.
3. If all the Examples are positive, return the single-node tree Root, with label = +.
4. If all the Examples are negative, return the single-node tree Root, with label = -.
5. If the Attributes are empty, return the single-node tree Root, with label = most common.
6. Or else Begin
7. A is the attribute from Attributes that best* classifiers Examples
8. The decision attribute for Root is A.
9. For each of the possible value v_i of A. Add a new tree branch below Root, corresponding to the test A = v_i.
10. Let the Examples for v_i be the subset of examples that have value v_i for A.
11. If Examples v_i is empty, then below this new branch add a leaf node with label = most common.
12. Else add the sub-tree to the new branch.
13. ID3 (Examples, Target-attribute, Attributes-{A}).
14. End

It is formed of one root node, which is the tree top, or starting point, followed by series of other nodes. Terminal nodes are leaves (Sub factors). Each node corresponds to a split on the values of one input variable. This variable is chosen in order to reach a maximum of homogeneity amongst the examples that belong to the node, relatively to the output variable. The steps involved in experimental result can be used to understand the algorithm in a better way. The best attribute is the one with highest information gain which is used to predict Bankruptcy.

4. PROPOSED SYSTEM WITH FUZZY ID3

In this system for all the available qualitative factors we are going to apply the FID3 algorithm. Here we consider negative values as the least values and positive values as the highest values in order to specify the interval for each factor (i.e.) attribute. This interval is used to find the membership value for the factors. Then we are going to apply the Fuzzy ID3 based algorithm.

Fuzzy decision tree is an effective method to obtain knowledge in uncertain classification related problems. It uses the fuzzy set theory to represent the data set and also it combines the tree growing and pruning in order to determine the structure of the tree. Usually, there exist two different kinds of attributes one is discrete and the other is continuous. Frequently many algorithms require data with discrete value. It is not easy to substitute a continuous domain attribute with a discrete one. This also involves some partitioning and clustering. It is also problematic to define the boundaries of the continuous attributes.

Thereby ,the process of finding the values for each attribute and the Information gain, we can identify the best attribute among the available attributes and depending upon the information gain value we can list the attributes that has more impact on Bankruptcy. The following is an example to find the value for one attribute. Let us consider Stability of transaction as our attribute.

Let us consider the space of the Stability of transaction attribute x is approximately between (-100,100).Then the membership functions for the attribute x are worst, better, good μ_x.

$\mu_w(x) = 1$ x<0

$\mu_b(x) = 1\text{-}x/100$ 0<x<100

$\mu_g(x) = 0$ x>100

Now, we have partitioned the sample set into different intervals.

The partition is complete when (each domain value is belong to at least one subset) and inconsistent (a domain value can be found in more than one subset) [10][11].Then we calculate the Fuzzy entropy and Information Gain using the below formula.

(1) $\boldsymbol{H(S) = \sum_i^N -P_i * \log_2 P_i}$

(2) $\boldsymbol{G(S,A) = H(S) - \sum \frac{|S_v|}{|S|} H(S_v)}$

In equation (1), H(S) represents the entropy of the set *S* of training examples in the node. Similarly, equation (2) G(S, A) represent the Information gain of attribute A. Modulo S_v is the size of the subset S_v of training examples x_j with v attribute. Modulo S presents the size of set S.

This is used to find the Fuzzy entropy value and Information Gain [2]. Then, by using this Information Gain (IG) value we rank the factors in increasing order such that the best attribute will be the attribute with highest information gain which have more impact on bankruptcy.

4.1 EXPERIMENTAL RESULT

The following are the steps for finding the Information Gain value.

- Initially we have 37 factors to be considered.
- As a result of Experts Judgment we have got 17 positive values and 20 negative values.
- Now we apply the values for finding the Entropy value in order to find Information Gain.

Where, S – Sample set with 37 members' altogether

$$H(20+,17-) = -\left(\frac{20}{37}\right) * \log_2\left(\frac{20}{37}\right) - \left(\frac{17}{37}\right) * \log_2\left(\frac{17}{37}\right)$$

H(S) =H (17+, 20-) =0.2995

- The value of Entropy will always be lesser than 1.Now we calculate the Information Gain (G) value,
- Now here, let us find the value of G for Financial Flexibility.
- We consider the 8 examples, in which 6+ and 2- values. Now we look out for two functions where one is bad and other is good.

S (BAD) = [6+, 2-]

S (GOOD) = [11+, 18-] i.e. (the remaining after removing the 8 examples from (17+, 20-))

- Now to find value of IG for FF, we have to calculate the entropy value for bad and good function,
- The following is the step to calculate the entropy value for Bad and good functions.

$$H(6+,2-) = -\left(\frac{6}{8}\right) * \log_2\left(\frac{6}{8}\right) - \left(\frac{2}{8}\right) * \log_2\left(\frac{2}{8}\right)$$

H (B) =0.2442.

$$H(11+,18-) = -\left(\frac{11}{29}\right) * \log_2\left(\frac{11}{29}\right) - \left(\frac{18}{29}\right) * \log_2\left(\frac{18}{29}\right)$$

H (G) =0.2881

- Now apply the entropy values to find value of G, $G(S,FF) = 0.2995 - \left(\frac{8}{37}\right) * H(Bad) - \left(\frac{29}{37}\right) * H(Good)$
- = 0.2995-(0.0528)-(0.2258), G(S, FF) =0.0209

Thus the Information Gain value of Financial Flexibility is 0.0209.Similarly we calculate the entropy and Information Gain values for all the considered factors.

5. Fuzzy Implementation using MATLAB

The following are the steps involved in the implementation of fuzzy using MATLAB.First launch the MATLAB tool then in the command window type the command as fuzzy; this incorporates the fuzzy toolbox of MATLAB. Then the Fuzzy Inference System window appears where we provide the number of input factors and the number of expected outputs. Under the name of the FIS, on the left side, are the pop-up menus that allow us to modify the various parts of the inference process. On the right side, below is the area that displays both the names of input and output variable, its associated membership function type, and also its range.

Then select any input factor and click on it, thereby the Membership Function Editor window appears as shown in Figure1, here we provide the membership values and the display range.

The Fuzzy Logic Toolbox does not limit the number of inputs. If the number of inputs that is the number of membership functions is too large, then it may also be hard to analyze the Fuzzy Inference System using the other graphical interface

tools. The Membership Function Editor in fuzzy logic toolbox defines the shapes of all the membership functions related to each variable. The Rule Editor performs the task of editing the list of rules that defines the behaviour of the system. The Rule Viewer is used for looking at, as opposed to editing, the Fuzzy Inference System. The Rule Viewer is a display of the fuzzy inference diagram. Using as a diagnostic, it will show which rules are active and also how individual membership function shapes are determining the results.

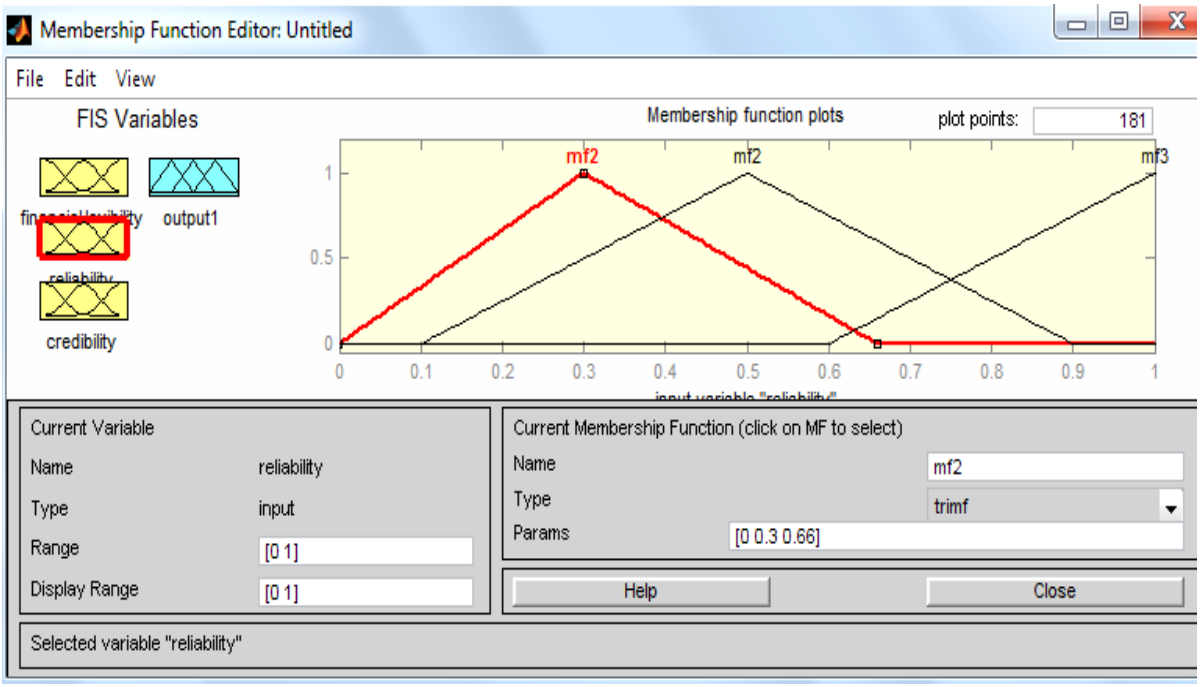

Fig 1: Membership Function Editor

Similarly we open the rule editor window and here we generate the rules based on the membership functions of the considered factors and after generating the rules in the rule editor window and we can also perform the actions of removing window and changing the previous generated rules in this rule editor. After generating the rules in the rule editor the rule viewer window shown in figure 2, is used to view the graphical view of the generated rules which displays the accuracy level and performance of the proposed system.

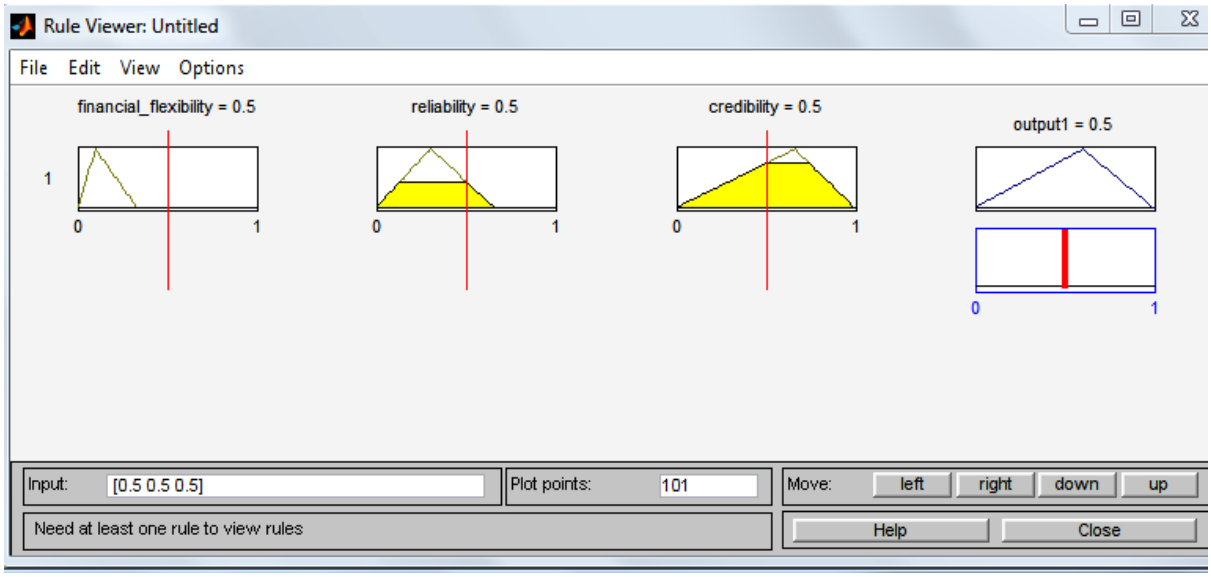

Fig 2: Rule Viewer Window

6. CONCLUSION

The Bankruptcy prediction can be done more accurately by considering the most important Qualitative factors. But only fewer works have been carried on predicting the Qualitative factors. The proposed system is applied to a successful bank and we clearly listed the most impacting factor to the least impacting factor and with the evaluated membership function we generated rules for the prediction and the prediction performance is viewed using the matlab tool. The predicted result shows the accuracy and performance level of prediction of Bankruptcy.

7. ACKNOWLEDGMENTS

Our Sincere thanks goes to the experts who have provided their suggestions and related data's towards the development of this proposed system.

8. REFERENCES

[1] The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms.Myoung-Jong Kim*, Ingoo HanB. 2003.

[2] "Applying Fuzzy ID3 Decision Tree for Software Effort Estimation ", Ali Idri and Sanaa Elyassami, International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011.

[3] W. H. Beaver, M. F. McNichols and J.W. Rhie,"Have financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy", Review of Accounting Studies, pp. 93 – 122, 2005.

[4] Nasir, M.L., John, RA., Bennett, S.C., Russell, D.M., Patel, A."Predicting Corporate Bankruptcy using Artificial Neural Networks'' Journal of Applied Accounting Research, Spring.2000

[5] "The prediction of bankruptcy using fuzzy classifiers" R. Nogueira, S.M.Vieira and J.M.C. Sousa, ieee 2005.

[6] Jan Vandeweyer, "Electre Tri as a method for bankruptcy prediction", Department Handels wetenschappen en Bedrijfskunde Strategisch KMO- en Retail management International Business, Academiejaar 2004-2005.

[7] LUAN Fugui, LI Jia," The Construction of Assessment System: Bankruptcy Reorganization Value of corporate", ORIENT ACADEMIC FORUM, 2009.

[8] "Using partial least squares and support vector machines for bankruptcy prediction", Zijiang Yang a, Wenjie You b, c, Guoli Ji Elsevier 2011.

[9] P.R. Kumar and V., Ravi, "Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques–A Review, "European Journal of perational Research, 180(1), 2007.

[10] "M. Umano, H. Okamoto, I. Hatono, H. Tamura, F. Kawachi, S. Umedzu, J. Kinoshita. "Fuzzy Decision Trees by Fuzzy ID3 algorithm and Its Application to Diagnosis Systems".IEEE Conference on Fuzzy Systems, vol. 3, June, 1994.

[11] Olaru C., Wehenkel L." A complete fuzzy decision tree technique. Fuzzy set and systems", pp. 221-254, 2003.

[12] Caouette, J. B., Altman, E. I., & Narayanan, P. "Managing credit risk: The next great financial challenge. New York: Wiley & Sons Inc", 1998.

[13] J.R.Quinlan, "Introduction of Decision Trees", Machine Learning 1 pp81-106, 1986.