# Credit Assessment of Bank Customers by a Fuzzy Expert System Based on Rules Extracted from Association Rules

Hamid Eslami Nosratabadi, Ahmad Nadali, and Sanaz Pourdarab

*Abstract*—**Credit assessment is a very typical classification problem in Data Mining. A type of classification technique that has attracted an increasing number of attempts in recent years is finding classification rules based on association rule mining techniques. This paper aims to contribute to this kind of research by classifying the bank's customers via association rules with the use of the APRIORI algorithm and CRISP-DM methodology and considering the Experts' opinions to filter the obtained rules and define the Membership functions for the considered criteria, finally a Fuzzy Expert System is designed based on the selected rules from association rules to specify the Credit Degree of banks' customers. The presented steps have been studied in an Iranian Bank as empirical study.**

*Index Terms*—**Credit assessment, fuzzy expert system, classification, association rules.**

## I. INTRODUCTION

Credit scoring is a method of predicting potential risk corresponding to a credit portfolio. These models can be used by financial institutions to evaluate portfolios in terms of risk. Credit scoring tasks can be divided into two distinct types. The first type is application scoring, where the task is to classify credit applicants into ''good'' and ''bad'' risk groups [1]. The second type of tasks deal with existing customers and along with other information, payment history information is also used here. This is distinguished from the first type because this takes into account the customer's payment pattern on the loan and the task is called behavioral scoring[1].Classification and association-rule discovery are two of the most important tasks addressed in the data mining literature. In recent years, extensive research has been carried out to integrate both approaches. By focusing on a limited subset of association rules, i.e. those rules where the consequent of the rule is restricted to the class variables, it is possible to build more accurate classifiers [2].

In the literature, several studies used financial information in development of credit models. For example in a paper, The Ranked Multilabel Rule (RMR) algorithm is introduced as a new associative classification technique, which is generating the rules with multiple labels [3]. There is another study which presents a new fuzzy data-mining algorithm to extract both fuzzy association rules and membership functions by means of a genetic learning of the membership functions and a basic method for mining fuzzy association rules[4].A

method of building credit-scoring models using fuzzy rule based classifiers has been suggested in another research which the rule base is learned from the training data using a SOM based method and Then design a contextual classifier by incorporating the fuzzy k-nn rule is with it for better classification [1]. In a recent application in the field of credit scoring, a decision support system is developed for the banks to determine the credibility of manufacturing firms in Turkey which is based on the financial ratios and fuzzy TOPSIS approach [5].Another study applied neuro-fuzzy models to analyze consumer loan applications and compared the advantages of neuro-fuzzy systems over traditional statistical techniques in credit-risk evaluation [6]. A genetic fuzzy and a neuro-fuzzy classifier for credit scoring has been applied in another research [7]. There is a benchmark state-of-the-art classification algorithms for credit scoring also [8]. Another paper proposed a hybrid system based on clustering and neural network techniques [9]. A two-stage hybrid modeling procedure with artificial neural networks and multivariate adaptive regression splines has been studied in a research also [10]. There is a study which integrated the back propagation neural networks with traditional discriminant analysis approach [11]. A work involving two interesting credit analysis problems and resolves them by applying neural networks and genetic algorithms techniques has been presented in [12]. In another research it estimates the Label of Credit customers with the use of Fuzzy Expert system to classify them and then a Data Mining algorithms is run on the final data [13].In the field of Fuzzy Expert System, a paper presents a Fuzzy Expert System to evaluate Science and Technology Parks [14]. Recent studies have revealed that emerging artificial intelligent (AI) techniques, such as artificial neural networks (ANNs) [15-16-17], evolutionary computation (EC) and genetic algorithm (GA) [18] and support vector machine (SVM) are advantageous to statistical analysis and optimization models for credit risk evaluation in terms of their empirical results[19-20-21].

Some Banks make use of the Data-oriented methods like Data Mining and some of the Knowledge-oriented methods like Expert System to score the credits .There is a challenge to specify which method will give better results .Since the use of the extracted knowledge from Data Mining process to score the credit customers is not consistent with experts opinions sometimes and has some problems, It is not safe to ignore the experts' opinions. In the other side, ignoring the hidden extracted patterns in data (which is in the form of obtained rules from the classification algorithms) will separate the credit scoring process from the available facts in historical data and it is possible not to understand the value of these rules. In the suggested method the credit scoring of

bank customers is not only based on the extracted rules from Association rules as the Data Mining method and opinions of Bank's experts is considers here which is the advantage of this model. In addition, with considering of their experiences and expertise, a Fuzzy Expert System has been designed. As the result, the model has higher level of integrity (according to consider both of the Data-oriented and knowledge-oriented methods) and higher level of confidence, because of controlling the rules by the experts.

From the another point of view, with the existence of this composed model the evaluation of customers' credit will not be limited to the obtained rules from Data Mining and It is capable to assess the customers' credit in various situations of their attributes (as the Input variables) in the form of deductable rules from the FIS model. So we will have more extended and universal model. The scope of the current paper however is limited to the evaluation of Saman Bank. Section 2& 3 of this paper present a survey of the literature related to the Data Mining & Credit Scoring & Fuzzy Expert system. Section 4 provides an empirical study in Saman Bank, while Section 5 discusses the relative benefits and the managerial implications of the proposed methodology as the conclusion.

## II. CREDIT ASSESSMENT AND ASSOCIATION RULES

Data Mining (DM) is an iterative process within which progress is defined by discovery, either through automatic or manual methods. DM is most useful in an exploratory analysis scenario in which there are no predetermined notions about what will constitute an ''interesting'' outcome. The application of Data Mining techniques for financial classification is a fertile research area [22]. Data mining (DM) is the process for automatic discovery of high level knowledge by obtaining information from real data [4]. Data mining techniques, therefore, can help to accomplish such a goal by extracting or detecting hidden customer characteristics and behaviors from large databases. The generative aspect of data mining consists of the building of a model from data. Each data mining technique can perform one or more of the following types of data modeling: (1) Association; (2) Classification; (3) Clustering; (4) Forecasting; (5) Regression; (6) Sequence discovery; (7) Visualization. Here are some examples of some widely used data mining algorithms:(1) Association rule;(2) Decision tree;(3) Genetic algorithm;(4) Neural networks;(5) K-Nearest neighbor; (6) Linear/logistic regression[23].

One of the applications of Data Mining is Credit scoring. Credit scoring is a method of predicting potential risk corresponding to a credit portfolio. These models can be used by financial institutions to evaluate portfolios in terms of risk. Credit scoring tasks can be divided into two distinct types. The first type is application scoring, where the task is to classify credit applicants into ''good'' and ''bad'' risk groups. The data used for modeling is generally consisted of financial information and demographic information about the loan applicant. In contrast, the second type of tasks deal with existing customers and along with other information, payment history information is also used here. This is distinguished from the first type because this takes into account the customer's payment pattern on the loan and the task is called behavioral scoring [1].

Classification is concerned with the categorization of records in a data set, typically achieved by applying a set of Classification Rules (CRs), also sometimes referred to as prediction rules. A classification rule has the general form A--> C, where A, the antecedent, is the union of some set of attribute values of the records involved: for example age < 40 and status = married. The consequent, C, is the label of a class to which records can be assigned. Classification rules are typically derived from examination of a training set of records that have been previously annotated with appropriate class labels. Many techniques exist to generate CRs from a given training set including techniques based on decision trees, Bayesian networks , and Support Vector Machines .A method for generating classification rules that has attracted recent attention is to make use of Association Rule Mining (ARM) techniques. ARM is concerned with the discovery of probabilistic relationships, known as Association Rules (ARs), between binary valued attributes of database records. ARM algorithms, such as Apriori, Apriori-TFP and FPgrowth have been developed specifically to address the problem of finding ARs within very large and sometimes noisy datasets, possibly including very large numbers of attributes. Frequently, however, and especially when working with imprecise and noisy training data, it will not be possible to obtain a complete classification using only a small number of high-confidence rules. Most methods, therefore, generate a relatively large number of rules which are then pruned. Algorithms for generating Classification Rules can be broadly categorized according to when the required pruning is performed as follows: 1.Two-stage algorithms, which first produce a set of candidate CRs, by a Classification Association Rule Mining (CARM) process or otherwise (stage1).In a separate second stage these are then pruned and ordered for use in the classifier. Examples of this approach include CMAR, CBA 2. Integrated algorithms where the classifier is produced in a single processing step, i.e. generation and pruning is ''closely coupled''.The CBA algorithm exemplifies the ''two-stage'' approach, and was one of the first to make use of a general ARM algorithm for the first stage. CBA uses a version of the well-known Apriori algorithm, using user supplied support and confidence thresholds, to generate CARS which are then prioritized as follows (given two rules r1 and r2): (1)r1 has priority over r2 if r1.confidence > r2.confidence. (2) r1 has priority over r2 if r1.confidence == r2.confidence and r1.support > r2.support.(3) r1 has priority over r2 if r1.confidence == r2.confidence and r1.support == r2.support and |r1.antecedent| < |r2.antecedent|.Once the rules are generated the ordered set of rules is pruned (stage 2) as follows:1) For each record in the training set find the first rule (the one with the highest precedence) that correctly classifies the record (the cRule) and the first rule that wrongly classifies the record (the wRule). 2) For each record where the cRule has higher precedence than the wRule, the rule is included in the classifier. 3) For all records where the cRule does not have higher precedence than the wRule alternative rules with lower precedence must be considered and added to the classifier. CRs are added to the classifier according to their precedence. On completion the lower precedence rules are

examined and a default rule selected to replace these low precedence CRs. CBA illustrates a general performance drawback of two-stage algorithms; the cost of the pruning stage is a product of the size of the data set and the number of candidate rules, both of which may in some cases be large. It is clear, also, that the choice of support and confidence thresholds will strongly influence the operation of CBA. The ordering strategy seems to work well on some data sets but less well on others [2]. In classification, one builds a model (set of rules) from a group of classified training data objects in order to forecast the classes of previously unseen data objects. The data set used to learn the model is known as the training data set and the data set used to measure the quality of the model is known as the test data set. Classification techniques such as rule induction and covering derive the most obvious class correlated to a rule and simply ignore the other classes. They consider building the classifier in a heuristic way. Once a rule is created, all training data objects associated with it are removed, and thus a training object is covered by only one rule. Hence, these approaches usually produce small sized classifiers. In the last few years a new hybrid approach that uses association rule discovery methods to build classifiers called associative classification (AC), has been proposed. AC is a promising classification approach, which has been shown to build more accurate set of rules than traditional classification approaches. Most AC algorithms work in two phases, the first phase involves the discovery of frequent rule items (attribute values that occur with a class label above the user specified support threshold). In the second phase, a set of rules (classifier) is built from the frequent rule items found in the first one. In phase 2 and before constructing the classifier, most AC algorithms sort the rules discovered in phase 1 according to their confidence and support values and then apply pruning heuristics to discard redundant and useless rules [3]. A popular pruning method in AC mining, for building the classifier, is the database coverage. This method tests the set of ranked rules against the training data. For each rule starting with the top ranked rule, the database coverage heuristic tests if the selected rule covers correctly any training data object. In other words, it examines if the selected rule antecedent matches any training data objects. If the test turns to be true and both the selected rule and the training data object have a common class, then such a rule is considered a candidate rule in the classifier. If no training data objects match the selected rule or there was a match but no common class was found, then the selected rule will be discarded [3].

## III. FUZZY EXPERT SYSTEM

Fuzzy expert systems use fuzzy data, fuzzy rules and fuzzy inference, in addition to the standard ones implemented in the ordinary expert systems [14]. The fuzzy Inference Systems (FIS) are very good tools as they hold the nonlinear universal approximation. They are suitable to handle experimental data as well as a priori knowledge on the unknown solution, which is expressed by inferential linguistic rules in the form IF THEN whose antecedents and consequents utilize fuzzy sets instead of crisp numbers. Fuzzy inference systems can express human expert knowledge and experience by using

fuzzy inference rules represented in "if-then" statements. The fuzzy inference process has five steps: Fuzzify inputs, apply fuzzy operator, apply implication method, aggregate all outputs and Defuzzify. In order to obtain a good FIS it is necessary that the researchers possess domain knowledge; the knowledge has to be represented in a symbolic form, be complete, correct and consistent. Numeric analysis approach of fuzzy system was first presented by Takagi and Sugeno and then a lot of studies have been made [13]. A fuzzy expert system is simply an expert system that uses a collection of fuzzy membership functions and rules, instead of Boolean logic, to reason about data (Schneider et al., 1996). The rules in a fuzzy expert system are usually of a form similar to the following:

If A is low and B is high then X = medium, where A and B are input variables, X is an output variable.

Here low, high, and medium are fuzzy sets defined on A, B, and X respectively. The antecedent (the rule's premise) describes to what degree the rule applies, while the rule's consequent assigns a membership function to each of one or more output variables. Generally the two most popular fuzzy inference systems are introduced that have been widely deployed in various applications: Mamdani Fuzzy inference system and Takagi-sugeno fuzzy inference system. The differences between these two fuzzy inference systems lie in the consequents of their fuzzy rules, and thus their aggregation and defuzzification procedures differ accordingly [14]. Fuzzy expert system modeling can be pursued using the following steps: 1) Select relevant input and output variables. 2) Choose a specific type of fuzzy inference system 3) Design a collection of fuzzy if-then rules (knowledge base) [14].

## IV. EMPIRICAL STUDY

In this study, we have used from CRISP-DM (CRoss-Industry Standard Process for Data Mining) method which Its life cycle consists of six phases as Fig 1.
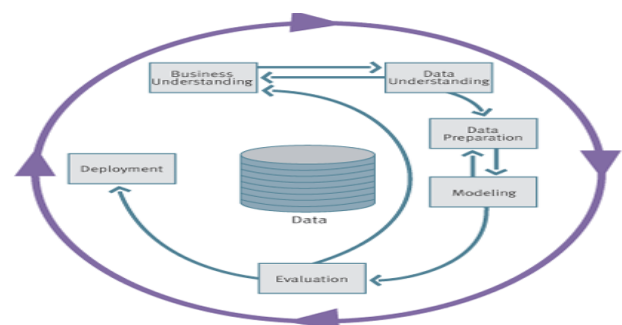


Fig 1. Phases of the CRISP-DM process model

To follow this standard, we have collected 436 records of the factual customers for The Saman Iranian Bank and have considered four financial criteria related to customers includes Current ratio, Debit ratio, Net Benefit Ratio, Claims collection period as the input fields of database and the credit degree of customers as the output field. Since the research variables are numerical and we need to define them as an Item Set for the Apriori algorithm, we should break the fields

into the form of spectrums. At First the spectrums for each field has been asked from the bank's experts (Table1). As we are going to use of the association rules to classify the data, we have defined the input data as Consequent and the class as Antecedent before running the model on data. At last, we have used the association rules and the Apriori algorithm with minimum Support equal to 0.05 and the minimum Confidence equal to 0.08 and classified the customers and have obtained the final rules. In this research, the Clementine software is used to run the Apriori algorithm.

TABLE I: DISCRETE RANGES OF FIELDS FOR NUMERIC VALUES

| Type of field | Name of field | Low | Medium | High |
|---|---|---|---|---|
| Input1 | Current Ratio | 0-1.2 | 1.2-1.4 | > 1.4 |
| Input2 | Debit Ratio | 0-0.65 | 0.65-0.8 | > 0.8 |
| Input3 | Net Benefit ratio | 0- 0.1 | 0.1-0.2 | > 0.2 |
| Input4 | Claims Collection period | 0-90 | 90-150 | > 150 |
| Output | Credit Degree of Customer | 0-0.4 | 0.4-0.7 | 0.7-1 |

Then 12 Rules from the obtained rules, which had higher Support and confidence criteria, have been selected according to the views of experts as the useful and proper rules. Since these selected rules, were similar to the views of banks' experts (Table II).

TABLE II: THE SELECTED RULES FOR FUZZY EXPERT SYSTEM

| | Current Ratio | DEBIT RATIO | Net Benefit Ratio | Claims Collection period | Credit Degree of Customer |
|---|---|---|---|---|---|
| 1 | Medium | HIGH | Low | Medium | Low |
| 2 | High | LOW | Medium | Medium | Medium |
| 3 | Medium | LOW | High | - | High |
| 4 | Medium | HIGH | - | Low | Medium |
| 5 | High | LOW | Low | High | Medium |
| 6 | Low | MEDIUM | Medium | High | Low |
| 7 | - | - | High | Low | High |
| 8 | High | LOW | Medium | Medium | Medium |
| 9 | High | MEDIUM | High | Low | High |
| 10 | Medium | HIGH | Medium | Low | Medium |
| 11 | - | HIGH | Low | - | Low |
| 12 | Low | MEDIUM | Medium | Medium | Medium |

Finally financial criteria including four variables as the input and the credit degree of customers as the output of Mamdani expert system in the MATLAB software have been defined (fig 2-6). The Experts defined the membership functions for these variables based on their knowledge and their opinions in the resulted rules from Data Mining process is run .Then the concluded system has been created based on the filtered rules   from Data mining process.
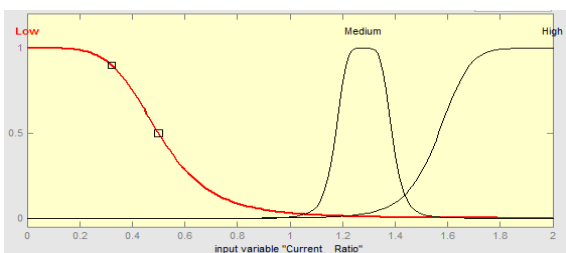

Fig. 2. Three Gbell membership function for current ratio
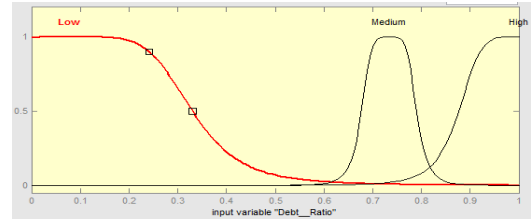

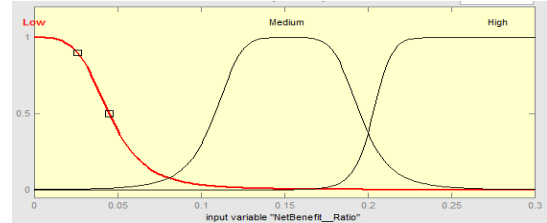Fig. 3. Three Gbell Membership function for Debit Ratio


Fig. 4. Three Gbell Membership function for Net Benefit ratio
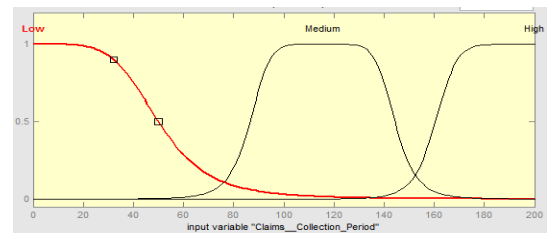

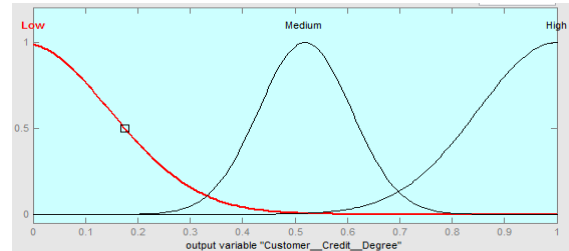Fig. 5. Three Gbell membership function for claims collection period


Fig. 6. Three Gaussian membership functions for credit degree of customer

Regarding to the proposed fuzzy expert system, we have evaluated a credit customer of bank as follows:

Current Ratio:1.34          Debt ratio:0.359
NetBenefit_ratio:0.226      Collection period:66.2

As a result Credit Degree will be 0.81.

## V. CONCLUSION

To build the credit scoring models from a credit database can be taken as a task of data mining. In this paper for evaluating the credit degree of Bank's customers, we have used of the combination of extracted rules of association rules by Data Mining Process and the knowledge of experts. We have classified the customers by the Association Rules after preparing and formatting the data. The obtained rules by the Bank's experts have been selected according to the importance. The designed Fuzzy Expert system is based on the rules of association rules and the defined Membership functions by the experts. This Fuzzy Expert system is used as the final predicting model for credit assessment of the bank's customers. One of the advantages of this model is the combination of the knowledge of the Bank's experts and the rules extracted from association rules to the fuzzy expert system.

REFERENCES

[1] A. Laha, "Building contextual classifiers by integrating fuzzy rule based classification technique and k-nn method for credit scoring", *Advanced Engineering Informatics* ,2007, 21 (3), pp. 281-291.

[2] F. Coenen and P. Leng," The effect of threshold values on association rule based classification accuracy", *Data & Knowledge Engineering,* 2007, 60 (2), pp. 345-360.

[3] F. A. Thabtah andP. I. Cowling, "A greedy classification algorithm based on association rule", *Applied Soft Computing*, 2007, 7 (3), pp. 1102-1111.

[4] J. A. Fdez, R. Alcalá, M. J. Gacto, and F. Herrera,"Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms", *Fuzzy Sets and Systems*, 2009, 160 (7), pp. 905-921.

[5] Y. T. İÇ and M. Yurdakul, "Development of a quick credibility scoring decision support system using fuzzy TOPSIS", *Expert Systems with Applications*, 2010, 37 (1), pp. 567-574.

[6] R. Malhotra and D. K. Malhotra "Differentiating between good credits and bad credits using neuro-fuzzy systems", *European Journal of Operational Research*, 2002, 136(1), pp. 190–211.

[7] F. Hoffmann, B. Baesens, J. Martens, F. Put, and J. Vanthienen, "Comparing a genetic fuzzy and a neurofuzzy classifier for credit scoring". *International Journal of Intelligent Systems*, 2002, 17(11), 1067–1083.

[8] B. Baesens, T. V. Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen,"Benchmarking state-of-the-art classification algorithms for credit scoring" *Journal of the Operational Research Society*, 2003,54(6), pp. 627–635.

[9] N. C. Hsieh," Hybrid mining approach in the design of credit scoring models," *Expert Systems with Applications*, 2005, 28(4), pp. 655–665.

[10] T. S. Lee and I. F. Chen, "A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines," *Expert Systems with Applications*, 2005, 28 (4), pp. 743-752.

[11] T. S. Lee, C-C. Chiu, C-J Lu, and I-F. Chen, "Credit scoring using the hybrid neural discriminant technique", *Expert Systems with Applications*, 2002, 24 (4), pp. 433-441.

[12] M. C. Chen and S. H. Huang, "Credit scoring and rejected instances reassigning through evolutionary computation techniques". *Expert Systems with Applications*, 2003, 24 (4), pp. 433-441.

[13] A. Nadali , S. Pourdarab, and H. Eslami Nosratabadi,"Labeling the class of Bank Credit's customers by a fuzzy Expert System For Credit Scoring with Data Mining Approach", ICKD 2011, IEEE.

[14] S.Pourdarab, H. Eslami Nosratabadi, M. Abbasian,"Design a Fuzzy Expert System to Evaluate Science and Technology parks", *4th international conference of Fuzzy Information and Technology,*2010.

[15] K. K. Lai, L. Yu, S. Wang, and L. Zhou "Credit risk analysis using a reliability-based neural network ensemble model", *Lecture Notes in Computer Science*, 2006, pp.682-690, DOI: 10.1007/11840930_71.

[16] K. K. Lai, L. Yu, L.G. Zhou, and S.Y. Wang, "Neural network meta learning for credit scores", *Lecture Notes in Computer Science*, 2006, 4113, pp. 403–408.

[17] R. Malhotra and D.K. Malhotra, "Evaluating consumer loans using neural networks". *Omega*, 2003, 31 (2), pp. 83-96.

[18] M. C. Chen and S. H. Huang, "Credit scoring and rejected instances reassigning through evolutionary computation techniques". *Expert Systems with Applications*, 2003, 24 (4), pp. 433-441.

[19] T. V.Gestel, B. Baesens, J. Garcia, and P. V. Dijcke,"A support vector machine approach to credit scoring". *Bank en Financiewezen*, 2003, (2), pp.73–82.

[20] Z. Huang, H. Chen, C-J. Hsu, W-H. Chen, and S. Wu "Credit rating analysis with support vector machines and neural networks: a market comparative study". *Decision Support Systems*, 2004, 37(4), pp. 543-558.

[21] K. K. Lai, L.Yu, W.Huang, and S. Wang,"A novel support vector machine metamodel for business risk identification". *Lecture Notes in Artificial Intelligence*, 2006, 980-984, DOI: 10.1007/978-3-540-36668-3_118

[22] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data Mining techniques for the detection of fraudulent,financial statements" , *Expert systems with applications*, 2007, 32 (4), pp. 995-1003.

[23] E. W. T. Ngai, L. Xiu, and D. C. K. Chau,"Application of data mining techniques in customer relationship management:A literature review and classification", *Expert Systems with Applications*, 36 (2 PART 2), pp. 2592-2602.