World Journal of Social Sciences
Vol. 1. No. 3. July 2011. Pp. 132-141

Modeling Consumer Credit Scoring Through Bayes Network

Abbas Heiat*

In this research TAN Bayes, Markov Bayes and Markov-FS data mining methods were used to predict andcompare the performance of these three Bayes networks and to identify the inputs or predictors that differentiate "good credit" from "bad credit". The results indicated the Markov model is slightly better at predicting correctly; however, the Markov-FS model is only a couple of percentage points behind the Markov model. This may mean it would be better to use the Markov-FS model since it uses fewer inputs to calculate its results, thereby saving on data collection and entry time and processing time.

Keywords: Management Information Systems, Banking, Marketing

1. Introduction

In the financial industry, consumers regularly request credit to make purchases. The risk for financial institutions to extend the requested credit depends on how well they distinguish the good credit applicants from the bad credit applicants. One widely adopted technique for solving this problem is Credit Scoring. Credit scoring is the set of decision models and their underlying techniques that aid lenders in the granting of consumer credit.

In credit business, banks are interested in learning whether a prospective consumer will pay back their credit. The goal of this study is to model or predict a credit applicant can be categorized as a good or bad customer. In this study I have used three different Bayes network to identify the inputs or predictors that differentiate risky customers from others on the training data, and later deploy those models to predict new risky customers. As described in the next section TAN Bayes, Markov Bayes and Markov-FS data mining methods has not been used in previous credit scoring studies. Section 3 explains the source of data for this study followed by methodology and findings. Section 6 provides the conclusion for the study.

2. Review of the Literature

Credit scoring has become a critical and challenging business analytics issue as the credit granting businesses have been facing stiffer competition in recent years. Many statistical and data mining methods have been suggested to tackle this problem in the literature.Historically, discriminant analysis and linear regression have been the most widelyused techniques for building score-cards. Both have the merits of being

*Abbas Heiat, College of Business, Montana State University, Billings, MT 59101, USAE-mail: aheiat@msubillings.edu

conceptually straight forward and widely available in statistical software packages. Other techniques which have been used in the credit scoring field include logistic regression, probit analysis,nonparametric smoothingmethods, mathematical programming, markov chain models, recursive partitioning, expert systems, and genetic algorithms, neural networks, and classification models(Hand & Henley 19997). Hand and Henley examined statistical and data mining methods that have been applied to credit scoring and discussed advantages and disadvantages of these methods. Some researchers believe the fact that significant portion of credit scoring information is not normally distributed is a critical limitation in applying discriminant and linear regression analysis to credit scoring cases. However, other researchers on the basis of empirical observation of credit scoring problems concluded that non-normal distribution of credit scoring information may not be a significant limitation (Hand, Oliver & Lunn 1998). Discriminant analysis also suffers from another weakness that it shares with logistic regression. They merely minimize the number of accepted bad loans given an acceptance rate without any rule for picking this rate optimally. On theoretical grounds one may argue that logistic regression is a more appropriate method than linear regression since the goal is to classify good and bad loans. In a comparative study, however, Henley found that logistic regression was no better than linear regression (Henley 1995). Wiginton compared logistic regression to discriminant analysis. He concluded that logistic regression gave superior classification results. But neither method was sufficiently good to be cost effective (Wiginton 1980).

Decision treeisthe most commonly used approach to discovering logical patterns within data sets. Decision trees may be viewed as a simplistic approach to rule discovery because of the process used to discover patterns within data sets. Decision tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions, and then splitting it up further on each of the branches. Initially, you start with a training set in which the classification label (say, "bad credit" or "good credit") is known (pre-classified) for each record. All of the records in the training set are together in one big box. The algorithm then systematically tries breaking up the records into two parts, examining one variable at a time and splitting the records on the basis of a dividing line in that variable (say, income>50,000 or income<=50,000). The object is to attain as homogeneous set of labels (say, "good credit" or "bad credit") as possible in each partition. This splitting or partitioning is then applied to each of the new partitions. The process continues until no more useful splits can be found. The heart of the algorithm is the rule that determines the initial split rule (Williamson1987).The process starts with a training set consisting of pre-classified records. Pre-classified means that the target field, or dependent variable, has a known class or label: "diabetic" or "non-diabetic”. The goal is to build a tree that distinguishes among the classes. For simplicity, assume that there are only two target classes and that each split is binary partitioning. The splitting criterion easily generalizes to multiple classes, and any multi-way partitioning can be achieved through repeated binary splits. To choose the best splitter at a node, the algorithm considers each input field in turn. In essence, each field is sorted. Then, every possible split is tried and considered, and the best split is the one which produces the largest decrease in diversity of the classification label within each partition. This is repeated for all fields, and the winner is chosen as the best splitter for that node. The process is continued at the next node and, in this manner, a full tree is generated.

Artificial neural networks are defined as information processing systems inspired by the structure or architecture of the brain (Caudill & Butler 1990). They are constructed from interconnecting processing elements, which are analogous to neurons. The two main techniques employed by neural networks are known as supervised learning and unsupervised learnin*g*. In unsupervised learning, the neural network requires no initial information regarding the correct classification of the data it is presented with. The neural network employing unsupervised learning is able to analyze a multi-dimensional data set in order to discover the natural clusters and sub-clusters that exist within that data. Neural networks using this technique are able to identify their own classification schemes based upon the structure of the data provided, thus reducing its dimensionality. Unsupervised pattern recognition is therefore sometimes called cluster analysis. Supervised learning is essentially a two stage process; firstly training the neural network to recognize different classes of data by exposing it to a series of examples, and secondly, testing how well it has learned from these examples by supplying it with a previously unseen set of data. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. It provides projections given new situations of interest and answers "what if" questions. There are disadvantages in using ANN. No explanation of the results is given i.e. difficult for the user to interpret the results. They are slow to train due to their iterative nature. Empirical studies have shown that if the data provided does not contain useful information within the context of the focus of the investigation, then the use of neural networks cannot generate such information any more than traditional analysis techniques can. However, it may well be the case that the use of neural networks for data mining allows this conclusion to be reached more quickly than might ordinarily be the case(Arminger 1997).

Nonparametric methods especially nearest neighbor have been used for credit scoring applications. While the nearest neighbor method has some attractive features, they have not been widely used in credit scoring applications. One reason being the demand on the computer resources.

In general there is no overall “best” method for classification applications. The choice of the method or methods will depend on the nature of the problem, on the data structure, the variables selected, and the goal of classifications.

3. Data

In this research I have used the data set with information pertaining to past and current customers who borrowed from a German bank for various reasons in this research. The data set contains information related to the customers’ financial standing, reason to loan, employment, demographic information, etc. The German Credit data set(available at ftp.ics.uci.edu/pub/machine-learning-databases/statlog/) contains observations on 30 variables for 1000 past applicants for credit. Each applicant was rated as “good credit” (700 cases) or “bad credit” (300 cases).

New applicants for credit can also be evaluated on these 31 "predictor" variables. We want to develop a credit scoring rule that can be used to determine if a new applicant is a good credit risk or a bad credit risk, based on values for one or more of the predictor variables. All the variables are explained in Table 1. The original data set had a number of categorical variables, some of which have been transformed into a series of binary variables so that they can be appropriately handled by the data mining software (Shmueli, Patel, &Bruce 2010).

Table 1

	Variable Name	Description	Variable Type	Code Description
1.	OBS#	Observation No.	Categorical	Sequence Number in data set
2.	CHK_ACCT	Checking account status	Categorical	0 : < 0 DM
				1: 0 <= ...< 200 DM
				2 : => 200 DM
				3: no checking account
3.	DURATION	Duration of credit in months	Numerical	
4.	HISTORY	Credit history	Categorical	0: no credits taken
				1: all credits at this bank paid back duly
				2: existing credits paid back duly till now
				3: delay in paying off in the past
				4: critical account
5.	NEW_CAR	Purpose of credit	Binary	car (new) 0: No, 1: Yes
6.	USED_CAR	Purpose of credit	Binary	car (used) 0: No, 1: Yes
7.	FURNITURE	Purpose of credit	Binary	furniture/equipment 0: No, 1: Yes
8.	RADIO/TV	Purpose of credit	Binary	radio/television 0: No, 1: Yes
9.	EDUCATION	Purpose of credit	Binary	education 0: No, 1: Yes
10.	RETRAINING	Purpose of credit	Binary	retraining 0: No, 1: Yes
11.	AMOUNT	Credit amount	Numerical	
12.	SAV_ACCT	Average balance in savings account	Categorical	0 : < 100 DM
				1 : 100<= ... < 500 DM
				2 : 500<= ... < 1000 DM
				3 : =>1000 DM
				4 : unknown/ no savings account
13.	EMPLOYMENT	Present employment since	Categorical	0 : unemployed
				1: < 1 year
				2 : 1 <= ... < 4 years
				3 : 4 <=... < 7 years
				4 : >= 7 years
14.	INSTALL_RATE	Installment rate as % of disposable income	Numerical	
15.	MALE_DIV	Applicant is male and divorced	Binary	0: No, 1:Yes
16.	MALE_SINGLE	Applicant is male and single	Binary	0: No, 1:Yes
17.	MALE_MAR_WID	Applicant is male and married or a widower	Binary	0: No, 1:Yes
18.	CO-APPLICANT	Application has a co-applicant	Binary	0: No, 1:Yes
19.	GUARANTOR	Applicant has a guarantor	Binary	0: No, 1:Yes
20.	PRESENT_RESIDENT	Present resident since - years	Categorical	0: <= 1 year
				1<…<=2 years
				2<…<=3 years
				3:>4years
21.	REAL_ESTATE	Applicant owns real estate	Binary	0: No, 1:Yes
22.	PROP_UNKN_NONE	Applicant owns no property (or unknown)	Binary	0: No, 1:Yes
23.	AGE	Age in years	Numerical	
24.	OTHER_INSTALL	Applicant has other installment plan credit	Binary	0: No, 1:Yes

25.	RENT	Applicant rents	Binary	0: No, 1:Yes
26.	OWN_RES	Applicant owns residence	Binary	0: No, 1:Yes
27.	NUM_CREDITS	Number of existing credits at this bank	Numerical	
28.	JOB	Nature of job	Categorical	0 : unemployed/ unskilled - non-resident
				1 : unskilled - resident
				2 : skilled employee / official
				3 : management/ self-employed/highly qualified employee/ officer
29.	NUM_DEPENDENTS	Number of people for whom liable to provide maintenance	Numerical	
30.	TELEPHONE	Applicant has phone in his or her name	Binary	0: No, 1:Yes
31.	FOREIGN	Foreign worker	Binary	0: No, 1:Yes
32	RESPONSE	Credit rating is good	Binary	0: No, 1:Yes

Source: Shmueli, Patel, and Bruce, 2010.

4. Methodology

I was interested to use a data mining method that has not beenused regularly in business applications. I used Bayes Net method in this research. A Bayesian network is a graphical model that displays variables (often referred to as nodes) in a dataset and the probabilistic, or conditional, independencies between them. Causal relationships between nodes may be represented by a Bayesian network; however, the links in the network (also known as arcs) do not necessarily represent direct cause and effect. For example, a Bayesian network can be used to calculate the probability of a patient having a specific disease, given the presence or absence of certain symptoms and other relevant data, if the probabilistic independencies between symptoms and disease as displayed on the graph hold true. Bayes Networks are very robust where information is missing and make the best possible prediction using whatever information is present (Darwiche 2003). Bayesian networks are used for making predictions in many varied situations; some examples are:

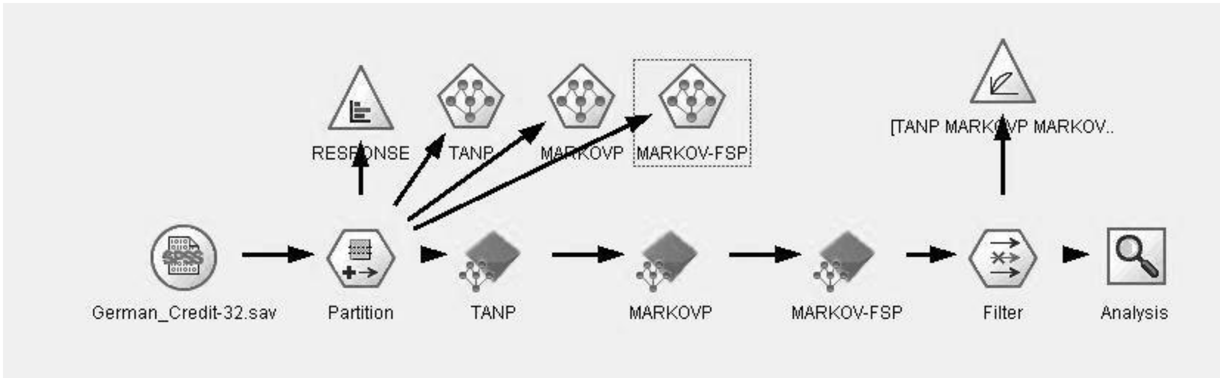
• Selecting loan opportunities with low default risk.
• Estimating when equipment will need service, parts, or replacement, based on sensor input and existing records.
• Resolving customer problems via online troubleshooting tools.
• Diagnosing and troubleshooting cellular telephone networks in real-time.
• Assessing the potential risks and rewards of research-and-development projects in order to focus resources on the best opportunities.

I used data mining software Clementine in this research. In the Clementine version I used (12.0 releases); the Bayes Net focuses on Tree Augmented Naïve Bayes (TAN) and Markov Blanket networks that are primarily used for classification.

5. Findings

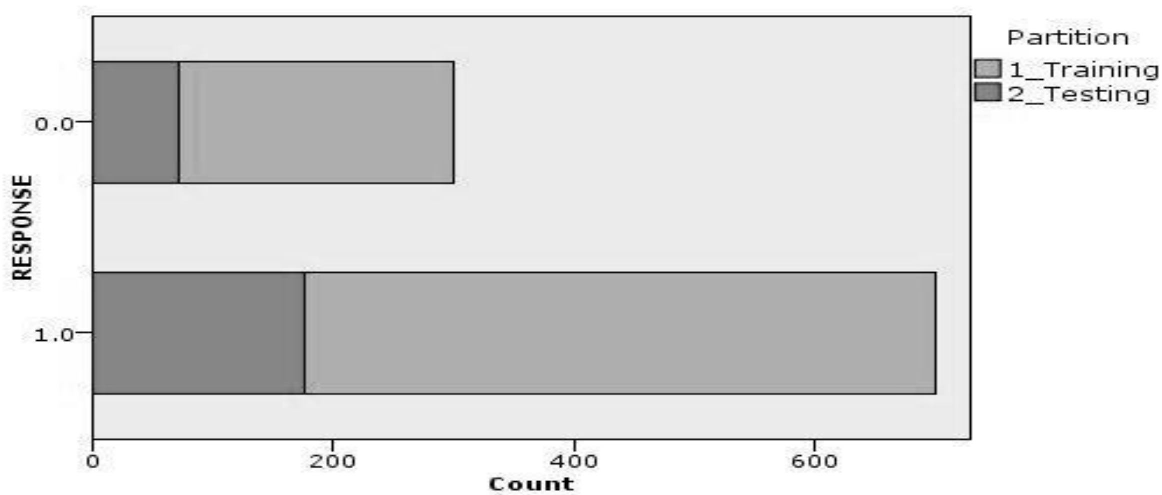
The Figure 1 shows the data mining model developed and used in this research.

Figure 1- The Data Mining Model for Credit Scoring



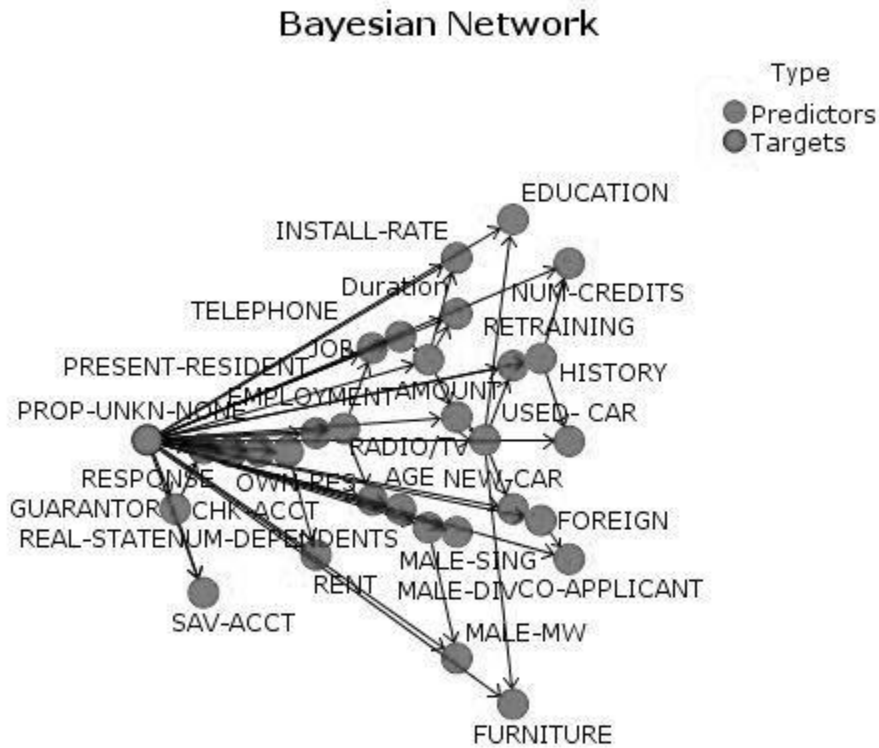
The diagram starts with selecting the data set for the analysis. Next, data was partitioned through partition node into training and testing (validation) sets. Figure 2 shows the configuration of partitioned data.

Figure 2- The Training and Testing Partition Proportions



The first model used is Augmented Naïve Bayes network (TANP). The lines in Figure 3 show the strength of effect of each variable on the response variable i.e. good or bad credit.

Figure 3- TANP Bayes Network



The second model used is Markov Bayes Network (MARKOVP). Figure 4 shows the Markov network and the variables determined to be important in differentiating good and bad credits. Duration, CHK-ACCT, History, and Amount variables are the most important variables in descending order.

The third model used is Markov Bayes Network with feature selection (MARKOV-FSP). This method tries to select variables before building the network. Figure 5 shows the Markov network with feature selection. CHK-ACCT, Duration, Amount, History, and Real State are determined by this model to be the most important variables in descending order.

The classification rate results for the three models indicate that the TAN Bayes network performance (78 % correct identification) is superior to Markov (66.8 %) and Markov-FS (73.2%) networks. These results are also confirmed by Evaluation Graph or Gain chart in Figure 6.

Figure 4- Markov Bayes Network and Important variables

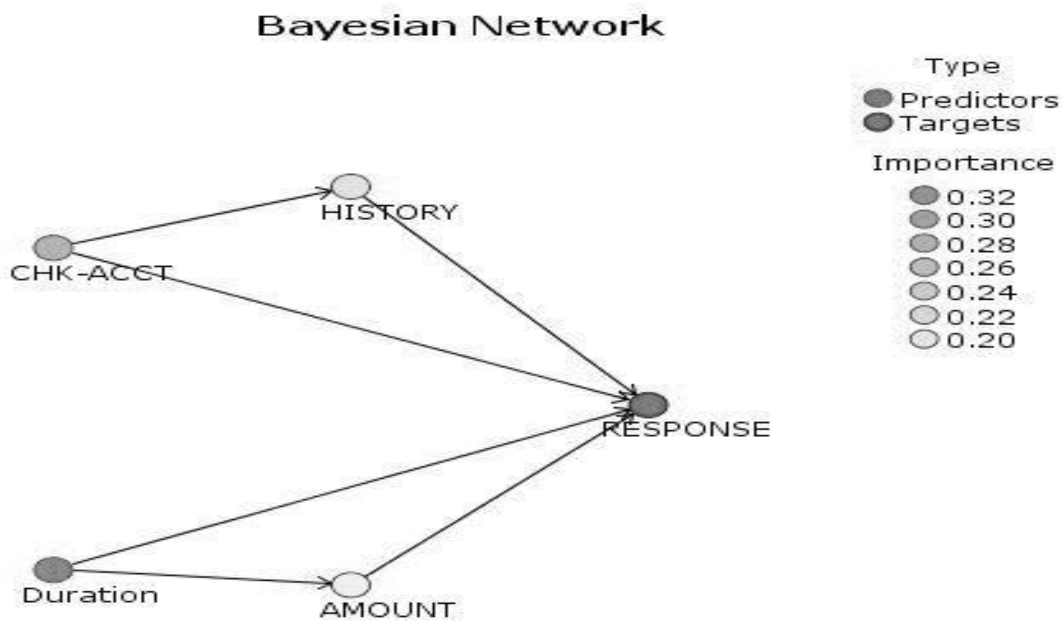


Figure 5- Markov Bayes Network with Feature Selection

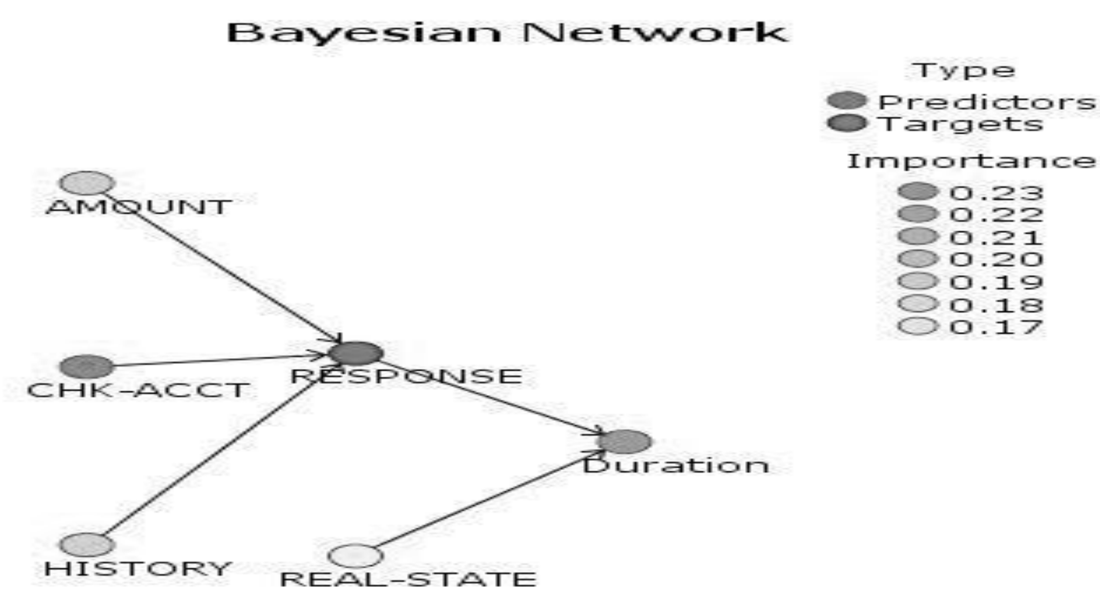
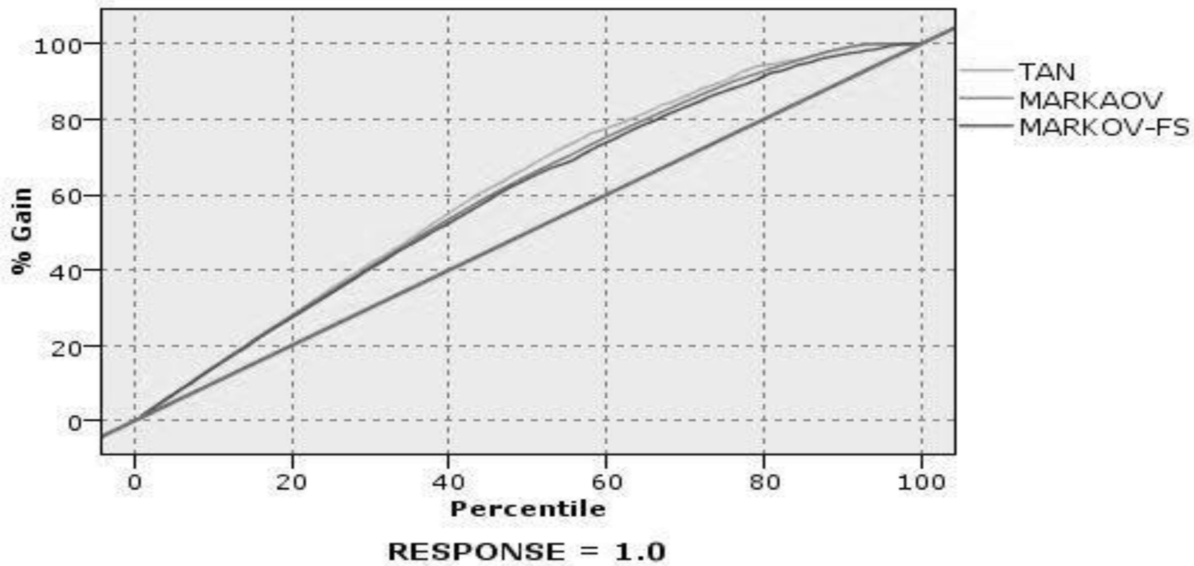


Figure 6- Evaluation Graph Comparing the three Bayes Models



6. Conclusion

The classification rate and the Evaluation graph shows that the Markov model is slightly better at predicting correctly good credits from bad credit. However, the Markov-FS model is only a couple of percentage points behind the Markov model. This may mean it would be better to use the Markov-FS model since it uses fewer inputs to calculate its results, thereby saving on data collection and entry time and processing time.

The number records with the good and bad credits are not equal in the German Credit dataset. To avoid bias, it is more appropriate to use equal number of each value of a binary target variable in classification applications. However, in our case the total number of records in our dataset was not large enough to sample equal numbers of good and bad credits. In future studies, using a larger dataset, equal number of good and bad credits should be used in the analysis. This would lead to much better results in accurately classifying the customers.

To evaluate the effectiveness of the Bayes Networks, in future researches it would be necessary to apply the common classification methods like Decision Trees or Neural Networks and compare the results with the performance of the Bayes network.

References

Arminger, GD & Bonne, T 1997, ‘Analyzing credit risk data: a comparison of logistic discrimination, classification tree analysis and feed- forward networks’, *Computational Statistics,* vol.12, pp. 293-310.

Caudill, M& Butler, A 1990, *Naturally Intelligent Systems,* The MIT Press, Cambridge.

Darwiche, A 2003, ‘A differential approach to inference in Bayesian networks’, *Journal of the ACM*, vol. 50, pp. 280-305.

Hand, DJ & Henley, WE 1997, ‘Statistical Classification Methods in Consumer Credit Scoring: a Review’, *Journal of Royal Statistics Society,* part 3, pp. 523-541.

Hand, DJ & Oliver, JJ &Lunn, AD 1998, “Discriminant analysis when the classes arise from a continuum”, *Pattern Recognition*, vol. 31, pp. 641-650.

Henley, WE 1995, *Statistical Aspects of Credit Scoring,* PhD theses, The Open University, Keynes, UK.

Shmueli, G, Patel, N, &Bruce, P 2010, *Data Mining for Business Intelligence*, Wiley, New Jersey.

Wiginton, JC 1980, ‘A note on the comparison of logit and discriminant models of consumer credit behavior’, *Journal of Financial Quantitative Analysis*, vol.15,pp. 757-770.

Williamson, S 1987, ‘Costly monitoring, loan contracts and equilibrium credit rationing’, *Quarterly Journal of Economics*, vol.102), pp. 135-145.