Asian Economic and Financial Review
ISSN(e): 2222-6737
ISSN(p): 2305-2147
DOI: 10.18488/journal.aefr.2017.79.846.857
Vol. 7, No. 9, 846-857

URL: www.aessweb.com

BAYESIAN NETWORK MODELING: A CASE STUDY OF CREDIT SCORING ANALYSIS OF CONSUMER LOAN'S DEFAULT PAYMENT

Lobna ABID[1+]
Soukeina Zaghdene[2]
Afif Masmoudi[3]
Sonia Zouari Ghorbel[4]

[1,4]*University of Sfax, Department of Economic Development, Faculty of Economic and Management of Sfax, Tunisia*
[2]*University of Carthage, Department of Economic Development, Institute of Business Studies of Carthage, Tunisia*
[3]*University of Sfax, Laboratory of probability and Statistics. Faculty of Sciences of Sfax, Tunisia*

(+ Corresponding author)

Article History
Received: 3 April 2017
Revised: 2 June 2017
Accepted: 4 July 2017
Published: 24 July 2017

Keywords
Bayesian network
Credit scoring
Tunisian commercial bank
Consumer credit.
Structure learning
Parametric learning.

JEL Classification
C1, C11.

ABSTRACT

This paper deals with the issue of predicting customers' default payment. The Bayesian network credit model is applied for the prediction and classification of personal loans with regard to credit worthiness. Referring to credit experts and using K2 algorithm for learning structure, we set up the dependency conditional relations between variables that explain default payments. Then, the parametric learning is adopted to detect conditional probabilities of customers' default payment. The parameters are estimated on the basis of real personal loan data obtained from a Tunisian commercial bank. The Bayesian network analysis has revealed that customers' age, gender, type of credit, professional status, and monthly repayment burden and credit duration have an important predictive power for the detection of customers' default payment. Therefore, our findings allow providing an effective decision support system for banks in order to detect and reduce the rate of bad borrowers through the use of a Bayesian Network model.

1. INTRODUCTION

The encouraging of credit policy in Tunisia has contributed to the increase in the rate of allocated loans to individuals and, subsequently, to the increase in the rate of overdue loans. This increase in household non-performing loans (NPLs) is mainly due to the rise in unpaid loans which are induced by an enhancing loan policy. According to Mouley (2014) NPLs, notably the real estate loans, has reached a rate of 21% between 2005 and 2009 compared to 6% for productive sectors.

Therefore, banks have become more likely to accurately predict the future solvability of loans' applicants in many ways, such as credit limit and the annual rate of allocated loans to customers. In addition, banks are concerned with improving the efficiency of their services through automating their decisions with regard to granting credits.

In this respect, financial institutions have attempted to deal with these challenges by adopting predictive-scoring models. According to Brigham (1992) the credit scoring systems are part and parcel of company risk management since they serve the alleviation of bad debt loss through identifying, analyzing and monitoring consumer credit risk. So, they are concerned with assessing default risk payment. For this, they have classified customers with regard to individual propensities to default payment.

Credit scoring is based on either statistical research methods or on operational ones. The statistical tools consist of (1) the discriminant analysis which is essentially a linear regression (Lobna *et al.*, 2016) (2) the logistic regression (Maja *et al.*, 2004) and (3) the classification trees (Deschaine and Francone, 2008).

The operational research techniques consist of variants of linear programming. Many researchers (Oreski and Oreski, 2014; Fatemeh, 2015; Vaclav, 2015) adopt one of these techniques to develop credit scoring models. These techniques, which are known as artificial intelligence, involve expert systems, fuzzy systems, neural networks, genetic algorithms etc. (Davis, 1987; Fan *et al.*, 2015).

In this study, to adopt a credit-scoring model that serves to predict the probability of customers' default payment and the assessment of credit risks in the Tunisian banking sector, we opted for the application of a Bayesian Network scoring model.

This model is assumed to be a Directed Acyclic Graphs (DAG) of nodes and arcs. In our context, nodes represent variables such as monthly repayment burden (MRB), type of credit, credit amount, etc. Arcs refer to the conditional dependencies between variables.

For their ability to encode conditional probability distribution, BNs are considered as one of the most complete and consistent formalism for the acquisition, representation and modeling of complex systems. Thus, they are the result of a convergence between statistical methods that ensures the transition from observation to description and those of artificial intelligence. They are useful for classification purposes when interactions between variables can be modeled by conditional probability relationships.

In the current paper, conducting a study on a sample consisting of 5993 borrowers from a commercial bank in Tunisia, we have developed a decision support system detecting borrowers' default payments. So, we initially applied both structural learning by using the K2 algorithm (Monte Carlo simulations) and parametric learning in BNs serving the estimation of numerical parameters.

In fact, the K2 algorithm is implemented in Matlab with the Bayesian Dirichlet equivalence uniform (BDeu) score and the parametric learning is performed by using Matlab. However, we do not need to specify any information a priori that has been proven to be highly performing and robust as applied in standard methods (logistic regression, discriminant analysis, etc.). Hence, we have focused on simulation results.

This paper is organized as follows: section 2 introduces some basic concepts of BNs. Section 3 describes the data and outlines the methodology. Section 4 deals with results and discussions. Section 5 concludes the paper.

2. BAYESIAN NETWORKS APPROACH

It is confirmed that BNs could naturally account for a modeling of different phenomena.
According to Cowell (1999); Pearl (1988) and Castillo *et al.* (1997) Directed Acyclic Graphs (DAGs) perform BNs which are probabilistic graphical models. For Pearl (1988) BNs are set up in accordance with the concept of conditional independence amongst variables, resulting in the factorization of the probability distribution of the random variable of the n-dimensional $(X_1,......., X_n)$ which is presented as follows:

$$P(X_1,......., X_n) = \prod_{i=1}^{n} P(X_i | Pa(X_i)) \tag{1}$$

Where $Pa(X_i)$ indicates the parents of the random variables X_i . Consequently, the precise identification of the joint probability distribution of a BN takes place via providing former probabilities for all root nodes (nodes with no predecessors) and conditional probabilities for all other nodes, taking into account all potential combinations of their straight predecessors.

These variables, coupled with the DAG, state the BN thoroughly. Therefore, the established network represents an efficient tool so as to carry out a probabilistic inference.

Assigning values to certain variables (i.e. attributing an instantiation to a list of variables), probabilistic inference is made up of getting posterior probability of one or more of the unassigned variables (Henrion, 1988).

This probabilistic reasoning inside the net could be applied by both exact methods and approximated methods. Yet, there is evidence that this computation is assumed to be a NP-challenging problem (Cooper, 1990).

As a result, the behavior of a BN is set by two parameters: its structure (the nodes and the links between them) and the probability tables combined with the nodes. In fact, the structure and the conditional probabilities which are vital for characterizing the network may be supplied externally by experts or derived from an algorithm inducing them systematically (Cheng *et al.*, 2002). According to Heckerman and Geiger (1995) for an expert to construct a BN structure, he/she has to design the network with reliance on his /her own knowledge about the relationships between the variables.

On the basis of the fact that modeling the knowledge of an expert is not accurate and even a time-consuming task, research has assumed that structure learning algorithms is a more reliable practice accounting for the independence relationships among variables. As a major area of research, algorithms aims to bring about the structure of the BNs that better reflects the conditional independence relationships underling a set of variables (Friedman and Koller, 2003).

In general, two components are necessary for the structural learning methods. They are the learning algorithm and the assessment metric which measures the goodness of the net with respect to each learning phase (score+search). For Heckerman and Geiger (1995) most outstanding approaches of the structural learning are linked to several connected networks and are classified depending on the extent to which they are essential for imposing order on the variables.

Establishing an order amongst the variables means that a variable X_i could get the variable X_j as a parent provided that X_j proceeds X_i in the displayed order. With this restriction, the space cardinality holding all the structures is determined by $2^{\binom{n}{2}}$, where n is the number of nodes in the graph (Acid and Huete, 2001).

According to Robinson (1977) the research space cardinality becomes larger, and the number of networks increases rapidly when there is no established order between nodes.

3. DATA AND METHODOLOGICAL ISSUES

3.1. Data

Data taken from a Tunisian commercial bank contain information about households' credits over the period 2014 - 2015. The original data contain 5393 customers where 84.49 % of them are good customers while 15.51% are bad debtors. Data include demographic variables such as age, gender and profession, and credit variables like type of credit, allocated amount, credit duration, monthly repayment burden (MRB), and outstanding credit. These variables are identified in Table 1 (see appendix 1).

3.2. Learning Bayesian Networks

It is essential to specify a structure (defined by a DAG) and the conditional probabilities attributed to each node of the DAG in order to get a BN. Hence, two tasks are required for learning a BN: (i) Structural learning which implies the identification of a BN topology, and (ii) parametric learning which refers to the rough calculation of numerical parameters (i.e., conditional probabilities).

3.2.1. Structural Learning

The structure, which is the DAG in a BN, can be set either by expert knowledge or by learning algorithms that are elaborated in the domain of machine learning techniques.

Indeed, since there exist lots of structures that are compatible with the same set of independencies, finding out the structure from a data set is a hard and challenging task (Heckerman *et al.*, 1995; Margaritis, 2003; Koller and Friedman, 2010).

It is noteworthy that the more the number of variables increases, the more the identification of the structure becomes difficult (Sucar and Martinez-Arroyo, 1998; Cheng *et al.*, 2002).

Among the methods that bring about a BN topology, two major approaches are worth considering: (1) Search-and-score structure learning methods (Chickering, 2002; Korb and Nicholson, 2010) and (2) Constraint-based structure learning methods (Scheines *et al.*, 1998; Spirtes *et al.*, 2001).

In *score-based* structure learning, each BN structure is given a score on the basis of the extent to which the model fits well the data, which allows looking for the model structure with the most elevated score.

The application of these methods rests upon two main parameters: (1) a scoring metric which is needed to assess the value of every candidate BN in a data set, and (2) a search procedure which is required to delve into a very vast space of possible networks.

In constraint-based learning, the input is a set of relations of conditional independence between subsets of variables. The adoption of learning algorithms serves the establishment of a BN that shows a high percentage of these relations (Spirtes *et al.*, 2001).

Thus, the application of the constraint-based learning procedure yields an undirected graph.

Subsequently, an additional independence test is used to obtain a BN.

In fact, both Score-based and Constraint-based algorithms constitute hybrid algorithms. Thus, in order to lower the search space, a conditional independence test is applied; and to find the most favorable network in the lowered space, a network score is adopted.

According to Cooper and Herskovits (1992) K2 algorithm is employed in learning structure of MATLAB (2014 a), assuming that an ordering on variables is available and all structures are, a priori, likely to be equal.

Thus, variables are divided into four blocks: (1) *background variables* = (age, gender), (2) *conditional variables* = (profession, credit amount), (3) *intermediate variables* = (credit duration, type of credit, amount of credit), and (4) *diagnostic variables* = (MRB).

Drawing on Hojsgaard *et al.* (2012) work, the model selection process is restricted by blacklisting arrows that point from a later block to an earlier one.

So, obtaining the structure rests upon two options. We either select a single best model or obtain some average model which is known as model averaging (Claeskens and Hjort, 2008).

The search has been conducted by using available learning algorithms that are included in the BNT package. Furthermore, to select the learning algorithm, we have to look for the plausibility of the model and the scattered graphs. The final model has been learnt with hill climbing (hc), and score-based algorithm K2.

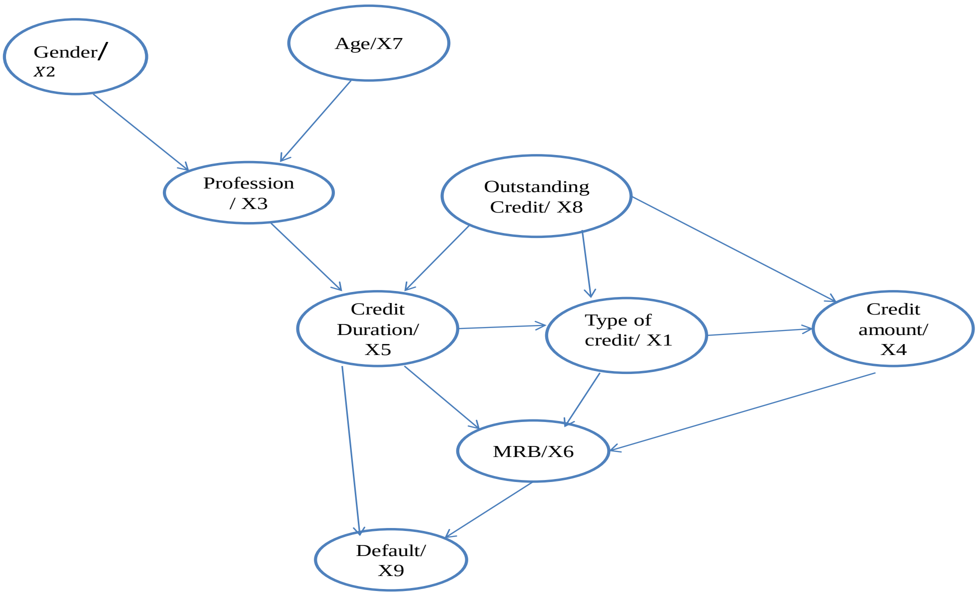

Figure-1. Structure obtained by using algorithm K2 from BNT package in MATLAB

Using the Kolmogorov-Smirnov statistical test, we assume that the obtained Bayesian network model is adequate for data with a p-value = 0.00021.

As Figure 1 indicates, some relationships between variables are easy to decode. For instance, both the credit duration and the MRB have a direct effect on default payment. The parents of the credit duration variable are the outstanding credit and profession. The credit amount has a direct effect on the MRB which, in turn, acts on default payments. The type of credit, the outstanding credit and the credit amount indirectly act on default payments. Besides, profession which is affected by gender and age has an indirect effect on default payment. Furthermore, the credit amount variable could indirectly discriminate between good and bad borrowers. Moreover, credit duration reflects the borrowers' intention, risk aversion, or self-assessment of repayment ability. This reinforces the fact that the longer a borrower is a client of a given bank, the more information the bank has about his/her banking behavior. Hence, allocating credits to ancient clients may reduce the probability of default payment. However, credit duration needs to be updated regularly due to adverse and unexpected changes in the borrowers' situation.

3.2.2. Parametric Learning

According to Neapolitan (2003) when a Bayesian parameter estimation is carried out through the adoption of the Dirichlet distribution, learning parameters have been obtained with the BN package in MATLAB (2014 a). Hence, we obtain a conditional probability distribution for each node.

Let D be a dataset and N_{ijk} the number of cases in *D* in which the node *i* is in state *k* and its parents are in state *j*;

that is $X_i = x_i^k$ and $Pa(X_i) = x_i^j$.

The distribution of $(N_{ij1},..., N_{ijri})$ is multinomial with parameters $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$ and $\theta_{ij} = (\theta_{ij1},..., \theta_{ijr_i})$, where

$\theta_{ijk} = |P(X_i = x|Pa(X_i) = x^j)$ and the Bayesian estimation of O_{ijk} is given by:

$$\hat{\theta}_{ijk} = \frac{N_{ijk} + \alpha_{ijk}}{N_{ij} + \alpha_{ij}}$$

In our study, we consider that $\alpha_{ijk} = 1$.

A conditional probability distribution is obtained for each node. Table 2 illustrates an example of conditional probability distribution. It also indicates that the joint probability distribution of the BN requires the specification of 9 conditional probabilities, one for each variable conditioned to its parents' set. Thus, the dependencies are easily translated into the probabilistic model.

With regard to the parents (X_5: credit duration and X_6: MRB) of the default payment variable (X_9) and with reference to the analysis of the relationships between the variables, it is clear that most people having default payment of 30.22% are those who have obtained consumer credit. Their repayment period ranges from 0 to 84 months and their MRB varies from 0 to 100 dinars. Besides, 25% of respondents who have default payment are those who were allocated consumer credits, the repayment period of which ranges from 0 to 84 months and their MRB is greater than 1000 dinars. Furthermore, 24.22% are those who have been granted housing credits with a duration exceeding 84 months and an MRB ranging from 100 to 200 dinars. Finally, 22.96% of applicants get housing credits whose repayment duration exceeds 84 months and whose MRB varies between 100 and 200 dinars.

Hence, based on parametric learning, consumer credit and MRB are the two most important indicators of default payments. This indicates that most Tunisian creditors belong to a middle class, proving the increase in consumer credit rates in late 2007, reaching a volume of 2,559 MD (total credits of 6.395 MD). Therefore, the Tunisian Central Bank has attempted to decrease the allocation of this type of credit (Circular N°. 2012-17).

4. RESULTS AND DISCUSSIONS

4.1. Inference

In a BN, to observe the relationships between variables and to detect their conditional independences, the conditional probability distribution of one or more variables is computed, taking into account the values of some other variables. The observed variables are subcategorized into three kinds of variables: a query variable X_i , the evidence variables **E** and the unobserved variables **U.** In a BN, making inferences implies computing new probabilities as soon as new data is displayed (Butz *et al.*, 2009). Thus, inference refers to finding the probability of any variable X_i conditioned on **e**, i.e., $p(x_i|e)$. However, when there is no evidence, we have recourse to prior probabilities $p(x_i)$. In this case, inference consists in combining evidence from all parts of the network and making any sort of query.

In BNs, *abductive inference* refers to detecting values of a group of variables explaining well the identified evidence. *In total abductive inference*, we solve $\arg\max_{\mathbf{u}} p(\mathbf{u} \mid \mathbf{e})$, to find the most probable explanation (MPE), whereas in *partial* abductive inference, we aim to find better explanation of the observed evidence for a subset of variables in **u** (i.e.,the explanation set), which is termed partial maximum a posteriori (MAP). These problems involve not only computing probabilities but also solving an optimization problem.

With regard to the type of credit allocated to borrowers, Table 3 indicates that the probability of default payments is slightly higher for those who have been granted housing loans (58.55%) than those who have obtained consumer credits (41.45%).

Our results indicate that younger households increase the credit worthiness since they have fewer commitments compared to elder applicants (76, 88%) (Table4).

Then, Table 5 shows that retired customers having been granted loans of less than 5000 dinars constitute 52.62% of the total default payment rate. Besides, 44, 88% of unpaid loans are owing to customers who have been granted credits ranging from 5000 to 30000 dinars (Table 6). Finally, 52.68% of bad debtors have MRB that does not exceed 100 Dinars (Table 7).

4.2. Classification and System Decision Support

The use of Bayesian network model has allowed us to identify the most probable states of borrowers having default payment (Table 8). Based on applicants' profiles, we can build a decision support system that aims to distinguish between good and bad borrowers.

As Table 8 indicates, there are only 10% of female default borrowers compared to 90 % of male bad debtors. So, females have less probability of default than males. Thus, our results indicate that gender is a significant predictor for the classification of bad borrowers.

Therefore, our results corroborate literature on indicators of default repayments. In fact, compared to male customers, it has been argued that female borrowers have a higher repayment rate (Viganò, 1993; Salazar, 2008; Roslan and MohdZaini, 2009). In fact, because females stick to the culture of financial discipline and are commonly averse to risk, they make less default payment (Pitt and Khandker, 1998; Bhatt and Tang, 2002; Croson and Gneezy, 2009). In addition, our findings also indicate that those who obtain housing credits have more probability to default payment (60%) compared to consumer credits (40%). Thus, our results are fine-tuned with other authors' findings on customers' credit worthiness in connection with the risk associated to housing credit repayment (Schreiner, 2004; Dinh and Kleimeier, 2007).

As Table 8 illustrates, we also conclude that elder debtors are more likely to default payment than less aged ones. This finding confirms the results of Boyle *et al.* (1992); Arminger *et al.* (1997) and Thomas (2000) that assume that applicants' age is one of the most used socio-demographical variables, enabling to detect default payment among customers. Conversely, this finding does not corroborate the results of Boyle *et al.* (1992) and Thomas (2000) indicating that elder borrowers are more averse to risk, and therefore are less likely to default. This explains the extent to which banks are more reluctant to grant loans to old-aged borrowers because they appear more averse to risk. It is also interesting to note that most bad borrowers are retired customers, which proves that they have the highest risk of falling into default payment. It is also interesting to note that most bad borrowers (55%) are retired customers. This predictive power of applicants' professional category describing customers' default payment was also underlined in a study based on logistic regression and discriminant analysis (Lobna *et al.*, 2016). The significance of this variable lies in the fact that it is highly correlated with the variable of income as long as an applicant's profession may indicate whether he/she has a high and stable income. Finally, credit duration is considered as a predictive variable of default payment. Thus, as Table 8 signals, 90% of debtors who have made default payments are those whose loan duration ranges from 0 to 84 months. This implies that the shorter the repayment duration is, the easier customers' credits are paid back.

5. CONCLUSION

Recent research on credit scoring emphasizes the importance of not only distinguishing 'good' customers from 'bad' ones, and anticipating when customers can make default payments. Such a prediction enables banks to take certain measures in order to prevent customers' undesirable behavior and therefore protect itself from potential borrowers with high defaults risks in a timely manner. In this study, Bayesian Networks (BNs) have been applied in order to produce an intuitive, transparent and graphical representation of the investigated interdependencies between variables indicating default payment. The proposed decision model can serve building bank decision support systems enabling to deal with the issue of credit risk in Tunisian commercial banks. With reliance on an available dataset, the suggested BN modeling is based on structural and parametric learning which shows the accuracy and robustness of the decision model. The BN structure was built by using K2 algorithm for learning structure which assumes apriori an ordering on the variables. The probabilities distribution was estimated from the dataset using the Dirichlet distribution. Parametric learning implying the estimation of numerical parameters yields a typology of bad debtors. Furthermore, the implemented BN was adopted to make inferences predicting scenarios of having default payment. We recall that our data consist of 5393 Tunisian commercial bank customers. Our results indicate that 15.51% of those customers are bad debtors. Applying a BN modeling on this dataset, the developed decision system shows that gender is one of the most used socio-demographic variables determining default payment since male customers are those who have the greatest likelihood of being bad debtors. Equally important, the majority of unpaid loans are due to those who belong to the older age group and mainly to those who are retired. This suggests that the age of the borrower is a vital predictor of default payment. Similarly, as for the type of credit, those who were granted housing credits are more prone to make default payments in the Tunisian banking sector. The main contribution of this paper was the development of a decision model for banks in the context of credit allocation systems using BN. Another contribution was a description of a BN modeling process which can be extended to other banking services. Finally, the current paper can be perceived as an attempt to detect bad borrowers and reduce the rate of credit risk. However, a future research in this field can include a larger database that incorporates additional variables such as the effect of interest rate, customers' perception of how to pay back loans in case of default payment and also varying the types of banks.

Appendix

List of Tables

Table-1. Variables lists

Variables list and summary statistics	State of variables
Type of credit (X_1)	1=Consumer credit 2=Housing credit
Gender (X_2)	1=Male 2=Female
Profession (X_3)	1= worker 2= middle excecutive 3=senior excecutive 4=retired 5=liberal fonction
Credit amount (X_4)	1- < 5000 2 - 5000 << 30000 3 - 30000 << 50000 4- > 50000
credit duration (X_5)	1 - 0 - 84mois 2- > 84mois
(MRB) (X_6)	1 - 0 - 100 2 - 100 - 200 3 - 200 - 500 4 - 500 - 1000 5 - > 1000
Age (X_7)	2 - 0 - 50years 1- > 50years
Outstanding Credit (X_8)	1- < 5000 2 - 5000 << 30000 3 - 30000 << 50000 4- > 50000
Default payment (X_9)	1 - default 2 - non-default

Source: Author's calculation by using Matlab

Table-2. Expected values of probability distribution of default variable conditional on combinations of its parent's values

$Pa(X_5;X_6)$ / X_9	$X_9 = 1$ Default Payment	$X_9 = 2$ Non payment Default
$P(X_5 = 1; X_6 = 1)$	30.22%	69,77%
$P(X_5 = 1; X_6 = 2)$	8.399%	91,60%
$P(X_5 = 1; X_6 = 3)$	9.43%	90,56%
$P(X_5 = 1; X_6 = 4)$	7.4%	92,59%
$P(X_5 = 1;X_6 = 5)$	25%	75%
$P(X_5 = 2;X_6 = 1)$	22.96%	77,04%
$P(X_5 = 2;X_6 = 2)$	24.22%	75,77%
$P(X_5 = 2;X_6 = 3)$	14.49%	85,50%
$P(X_5 = 2;X_6 = 4)$	2.127%	97,87%
$P(X_5 = 2;X_6 = 5)$	50%	50%

Source: Author's calculation by using Matlab

Table-3. Conditional probability distribution of X_1 given the payment default(X_9= 1)

$P(X_1/X_9)$	Consumer Credit	Housing Credit
	0.4144956	0.5855044

Source: Author's calculation by using Matlab

Table-4. Conditional probability distribution of X_7 given the payment default(X_9= 1)

$P(X_7/ X_9)$	> 50 years	0- 50 years
	0,76886955	0,23113045

Source: Author's calculation by using Matlab

Table-5. Conditional probability distribution of X_3 given the payment default(X_9= 1)

$P(X_3/X_9)$	worker	middle executive	Senior executive	retired	Liberal function	Senior executive	retired	Liberal function
	0.164770	0.181609	0.122337	0.526281	0.0050	0.12233	0.526281	0.0050

Source: Author's calculation by using Matlab

Table-6. Conditional probability distribution of X_4 given the payment default(X_9= 1)

$P(X_4/X_9)$	< 5000	5000 << 30000	30000 << 50000	> 50000
	0,5230303	0,44885482	0,01933639	0,0087785

Source: Author's calculation by using Matlab

Table-7. Conditional probability distribution of X_6 given the payment default(X_9 = 1)

$P(X_6/X_9)$	0-100 Dinars	100-200 Dinars	200-500 Dinars	500-1000 Dinars	> 1000Dinars
	0,52686506	0,19542708	0,2682677	0,00410318	0,00533699

Source: Author's calculation by using Matlab

Table-8. The most likely profiles given X_9= 1.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	1	1	1	1	1	1	1	1
1	1	2	1	1	1	1	1	1
1	1	2	1	1	1	2	1	1
1	1	3	1	1	1	1	1	1
1	1	3	1	1	1	2	1	1
1	1	4	1	1	1	1	1	1
1	1	4	1	1	1	2	1	1
1	2	4	1	1	1	1	1	1
2	1	1	2	1	2	1	2	1
2	1	1	2	1	3	1	2	1
2	1	2	2	1	3	1	2	1
2	1	3	2	1	3	1	2	1
2	1	4	1	1	1	1	1	1
2	1	4	2	1	2	1	2	1
2	1	4	2	1	2	2	2	1
2	1	4	2	1	3	1	2	1
2	1	4	2	1	3	2	2	1
2	1	4	2	2	2	1	2	1
2	1	4	2	2	3	1	2	1
2	2	4	2	1	3	1	2	1

Funding: This study received no specific financial support.
Competing Interests: The authors declare that they have no competing interests.
Contributors/Acknowledgement: All authors contributed equally to the conception and design of the study.

REFERENCES

Acid, C.L.S. and J. Huete, 2001. The search of causal orderings: A short cut for learning belief networks. Proc. 8th Conference on Uncertainty in Artificial Intelligence.

Arminger, G., D. Enache and T. Bonne, 1997. Analyzing credit risk data: A comparison of logistic discrimination, classification tree analysis, and feed-forward network. Computational Statistics, 12(2): 293-310. *View at Google Scholar*

Bhatt, N. and S.Y. Tang, 2002. Determinants of repayment in microcredit: Evidence from programs in the United States. International Journal of Urban and Regional Research, 26(6): 360-376. *View at Google Scholar | View at Publisher*

Boyle, M., J. Crook, R. Hamilton and L. Thomas, 1992. Methods for credit scoring applied to slow payers. In L. Thomas, J. Crook and D. Edelman (Eds), Credit scoring and credit control. Oxford: Oxford University Press. pp: 75-90.

Brigham, E.F., 1992. Fundamentals of financial management. 6th Edn., Forth Worth: Dryden Scan.

Butz, C., S. Hua, J. Chen and H. Yao, 2009. A simple graphical approach for understanding probabilistic inference in bayesian networks. Information Sciences, 179(6): 699–716. *View at Google Scholar | View at Publisher*

Castillo, E., J.M. Gutierrez and A.S. Hadi, 1997. Expert systems and probabilistic network models. New York: Springer. pp: 605.

Cheng, J., R. Greiner, J. Kelly, D.A. Bell and W. Liu, 2002. Learning bayesian networks from data: An information theory based approach. Artificial Intelligence, 137: 43- 90.

Chickering, D.M., 2002. Optimal structure identification with greedy search. Journal of Machine Learning Research, 3: 507–554. *View at Google Scholar*

Claeskens, G. and N.L. Hjort, 2008. Model selection and model averaging. Cambridge: Cambridge University Press.

Cooper, G.F., 1990. The computational complexity of probabilistic inference using bayesian belief networks. Artificial Intelligence, 42(2-3): 393- 405. *View at Google Scholar | View at Publisher*

Cooper, G.F. and E. Herskovits, 1992. A bayesian method for the induction of probabilistic networks from data. Machine Learning, 9(4): 309-347. *View at Google Scholar*

Cowell, R.G., 1999. Probabilistic networks and expert systems. New York: Springer.

Croson, R. and U. Gneezy, 2009. Gender differences in preferences. Journal of Economic Literature, 47(2): 448-474.

Davis, D.B., 1987. Artificial intelligence goes to work. High Technologies, 7(4): 16-27. View at Google Scholar

Deschaine, L. and F. Francone, 2008. Comparison of discipulus TM linear genetic programming software with support vector machines, classification trees, neural networks and human experts, white paper. Retrieved from http://www.rmltech.com/ [Accessed 10 June 2008].

Dinh, T.H.T. and D.S. Kleimeier, 2007. A credit scoring model for Vietnam's retail banking market. International Review of Financial Analysis, 16(5): 471-495. View at Google Scholar | View at Publisher

Fan, L., H. Zhongsheng and L. Andrew, 2015. Identifying future defaulters: A hierarchical bayesian method. European Journal of Operational Research, 241(1): 202-211. View at Google Scholar | View at Publisher

Fatemeh, 2015. A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. Journal of Retailing and Consumer Services, 27: 11-23. View at Google Scholar | View at Publisher

Friedman, N. and D. Koller, 2003. Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. Machine Learning, 50(1): 95- 125. View at Google Scholar

Heckerman, D. and D. Geiger, 1995. Learning Bayesian networks: A unification for discrete and Gaussian domains. In P. Besnard and S. Hanks (Eds.), Proc. Eleventh Conference on Uncertainty in Artificial Intelligence (UAI '95). San Francisco, Calif: Morgan Kaufmann. pp: 274–284.

Heckerman, D., D. Geiger and D. Chickering, 1995. Learning bayesian networks: The combination of knowledge and statitical data. Machine Learning, 20(3): 194–243. View at Google Scholar | View at Publisher

Henrion, M., 1988. Propagating uncertainty in bayesian networks by probabilistic logic sampling. Proceedings of the Fourth Conference on Uncertainty in Artificial Intelligence. pp: 149-163.

Hojsgaard, S., D. Edwards and S. Lauritzen, 2012. Graphical models with R. New York: Springer.

Koller, D. and N. Friedman, 2010. Probabilistic graphical models: Principles and techniques. Cambridge, MA/London, England: The MIT Press.

Korb, K.B. and A. Nicholson, 2010. Bayesian artificial intelligence. 2nd Edn., Boca Raton, FL: Chapman, Hall/CRC.

Lobna, A., M. Afif and Z.G. Sonia, 2016. The consumer loans payment default predictive model: An application in a Tunisian commercial bank. Asian Economic and Financial Review, 6(1): 27. View at Google Scholar | View at Publisher

Maja, P., B. Mateja and Sandra, 2004. Comparison of logistic regression and linear discriminant analysis: A simulation study. Metodološkizvezki, 1(1): 143-161. View at Google Scholar

Margaritis, D., 2003. Learning bayesian network model structure from data. PhD Thesis of CMU-CS-03-153.

Mouley, S., 2014. Réformes et restructuration du système bancaire et financier en Tunisie: Quelle vision et quel plan stratégique prioritaire? In Cahier du Cercle des Economistes de Tunisie.

Neapolitan, R.E., 2003. Learning bayesian networks. Upper Saddle River, NJ, USA: Prentice Hall, Inc.

Oreski, S. and G. Oreski, 2014. Genetic algorithm-based heuristic for feature selection in credit risk assessment. Expert Systems with Applications, 41(4): 2052- 2064. View at Google Scholar | View at Publisher

Pearl, J., 1988. Probabilistic reasoning in intelligent systems. Networks of Plausible Inference, Morgan Kaufmann.

Pitt, M.M. and S.R. Khandker, 1998. The impact of group-based credit programs on the poor households in Bangladesh: Does the gender of participants matter? Journal of Political Economy, 106(5): 958- 996. View at Google Scholar | View at Publisher

Robinson, R.W., 1977. Counting unlabeled acyclic digraphs. In C. H. C. Little, Ed., Combinatorial mathematics V, of lecture notes in mathematics. Berlin: Springer, 28: 43, 622.

Roslan, A.H. and A.K. MohdZaini, 2009. Determinants of microcredit repayment in Malaysia: The case of Agrobank. Humanity and Social Sciences Journal, 4(1): 45-52. View at Google Scholar

Salazar, G.L., 2008. An analysis of repayment among clients of the microfinance institution Esperanza international, Dominic an Republic. Journal of Agricultural Economics, 90(5): 1366- 1391.

Scheines, R., P. Spirtes, C. Glymour, C. Meek and T. Richardson, 1998. The tetrad project: Constraint based aids to causal model specification. Multivariate Behavioural Research, 33(1): 65-117. View at Google Scholar | View at Publisher

Schreiner, M., 2004. Benefits and pitfalls of statistical credit scoring for microfinance. Savings and Development, 28(1): 63-86. *View at Google Scholar*

Spirtes, P., C. Glymour and R. Scheines, 2001. Causation, prediction and search, adaptive computation and machine learning. 2nd Edn.: The MIT Press.

Sucar, L.E. and M. Martinez-Arroyo, 1998. Interactive structural learning of bayesian networks. Expert Systems with Applications, 15(3-4): 325-332. *View at Google Scholar* | *View at Publisher*

Thomas, L., 2000. A survey of credit and behavioural scoring; forecasting financial risk of lending to consumers. International Journal of Forecasting, 16(2): 149- 172. *View at Google Scholar* | *View at Publisher*

Vaclav, K., 2015. Genetic algorithms for credit scoring: Alternative fitness function performance comparison. Expert Systems with Applications, 42(6): 2998- 3004. *View at Google Scholar* | *View at Publisher*

Viganò, L., 1993. A credit-scoring model for development banks: An African case study. Savings and Development, 17(4): 441-482. *View at Google Scholar*