



Bank Credit Risk: Evidence from Tunisia using Bayesian Networks

Mohamed Wajdi Triki^a

Younes Boujelbene^b

^a Corresponding Author, Faculty of Economics and Management of Sfax, University of Sfax, Tunisia, mohamedwajditriki@yahoo.com

^b Faculty of Economics and Management of Sfax, Higher Institute of Business Administration, University of Sfax, Tunisia, younes.boujelbene@fsegs.rnu.tn

Keywords

Credit scoring, Bank, Consumer credit.

Jel Classification

C11, E4.

Abstract

In this article, a problem of measurement of credit risk in bank is studied. The approach suggested to solve it uses a Bayesian networks. After the data-gathering characterizing of the customers requiring of the loans, this approach consists initially with the samples collected, then the setting in works about it of various network architectures and combinations of functions of activation and training and comparison between the results got and the results of the current methods used.

To address this problem we will try to create a graph that will be used to develop our credit scoring using Bayesian networks as a method. After, we will bring out the variables that affect the credit worthiness of the beneficiaries of credit. Therefore this article will be divided so the first part is the theoretical side of the key variables that affect the rate of reimbursement and the second part a description of the variables, the research methodology and the main results.

The findings of this paper serve to provide an effective decision support system for banks to detect and alleviate the rate of bad borrowers through the use of a Bayesian Network model. This paper contributes to the existing literature on customers' default payment and risk associated to allocating loans.

1- Introduction

The encouraging 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. In this sense, a great deal of non-performing households has increased in recent years (Mouley, S 2014). This increase in non-performing households is mainly due to the rise in unpaid loans induced by an encouraging loan policy to individuals, notably real estate loans.

In fact, when attempting to evaluate customers' credit risk, financial institutions observed that high risks are associated with inappropriate credit granting-policies. On the one hand, banks are likely to accurately predict the future performance of loans applicants. On the other hand, based on customers' repayment behavior and information about current use of credit cardholders, financial institutions tend to identify some criteria related to loans allocation. Among these, we cite the following: credit limit and annual percentage rate of allocated loans to customers. In addition, banks are concerned with improving the efficiency of their services through automating credit-granting decisions.

In this respect, financial institutions attempt 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. Financial institutions are concerned with assessing the default risk associated with sales on credit. To reach this end, they assign customers certain risk classes with regard to their individual propensities to default payment. In this sense, the probability of default can be obtained either on the basis of an external or an internal scoring model. It is worth noting that the Tunisian commercial bank is the basic internal source of information on creditworthiness since it can provide data on customers' previous payment behaviors and also information related to customers, such age, profession, amount of loan, etc.

In the current study, we attempt to construct a credit scoring model that serves the prediction of the probability of default payment among recently allocated credits.

Then, the operational research techniques consist of variants of linear programming. In fact, a major part of scorecard builders adopt one of these techniques or a combination of both of them. Furthermore, credit scoring is also subject to a number of different non-parametric statistical and Artificial Intelligence modeling approaches.

These approaches include the ubiquitous neural networks (Fan *et al.*, 2015), expert systems (Davis, 1987), genetic algorithms (Vaclav K, 2015 ; Fatemeh *et al.*, 2015) and nearest neighbour methods (Oreski and Oreski, 2014). It is noteworthy that a variety of approaches can be used on the same classification of problems.

In the current paper, in order to evaluate customer repayment ability, taking into account variables such as demographic variables, amount of credit etc., we apply a Bayesian Network

scoring model. Bayesian Network is a directed acyclic graph that encodes conditional probability distribution.

Indeed, Bayesian networks are considered as one of the most complete and consistent formalism for the acquisition, representation and modeling of complex systems. Bayesian networks are acyclic directed graphs of nodes and arcs, where nodes represent variables (Degree of solvability, type of credit, credit amount...) and arcs have conditional dependencies between variables.

Thus, they are the result of a convergence between statistical methods ensuring the transition from observation to description and those of artificial intelligence. They are useful for classification problems when interactions between variables can be modeled by conditional probability relationships.

In the present study, we developed a decision support system for borrowers' default payments. To reach this end, we initially applied parametric learning in Bayesian Networks. However, we do not need to specify any information a priori that was proven to be highly performing, more robust, and sensible compared to the standard methods (logistic regression, discriminant analysis ...). Hence, we focused on simulation results.

Real data on 6240 customers collected during 2013-2014 were provided by one of the most important commercial banks in Tunisia and were used in the empirical analysis. Therefore, with reference to these data, we took into consideration the structural learning and network settings that link different variables. The structural learning is based on the K2 algorithm implemented in Matlab (R2015) with the Bayesian network.

The learning of parameters is also performed by using Matlab (R2015). Finally, a decision support interface is developed. Based on customer's profile as well as on various determinants, the system allows calculating a default score for each applicant. This score function denotes the joint probability distribution for each profile of the Bayesian network by using the parameters estimated in the learning data set.

On the basis of credit attributes and demographic variables, the results of the empirical analysis show that Bayesian Network credit scoring model can be employed to assess customer repayment ability.

2- Summary of theoretical literature

We focus ourselves in what follows on three principal determinants of refunding in the specific cases of banks (Rhee, S, G., 2008; Roy, D., 2006 and Redis., 2005): factors related to its characteristics, those related to its environment, and finally those related to the characteristics of its borrowers.

The analysis of the literature having for objective to identify the causes of unpaid (Anderson, R., 2007; Honlonkou and al., 2006) shows that the insufficiency of the amounts of credit to finance the projects is a decisive cause of a bad performance of refunding. In the same way, found that the

coefficient of the amount of the loans is significant and negative. This result was also confirmed by M. Labie, and M. Mees., (2005). Indeed, the negative sign is theoretically explained by the fact that the amount of the loans increases the profit associated with the moral risk. However, V. Hartarska, and D. Nasdolnyak., (2007) showed that the majority of the not refunded loans at the maturity were completely refunded a year later. In this context, the moral risk is interpreted as the choice of a project with a longer maturity than that of the loan rather than the choice of a riskier project (Bellucci, A., Borisov, A., Zazzaro, A., 2010). The negative sign relating to the amount of the loan can also be associated with the obstacles which the borrower can face to refund a higher amount over a given period (usually a year) (Armingier et al ; 1997). It may be that for a given maturity, the loans of significant size do not go in par with the requirements of the borrowers and are not appropriate to the local economy (Basel Committee on Banking Supervision., 2010).

For a particular borrower and a given duration of loan, it is shown (Bhagavatula., and all., 2010; Bedecarrats, F., Angora, R.W., 2009; Lhériaux., 2005) that, the probability of refunding decrease with the size of the loan. The speed of the evolution of the probability of no refunding with the size of the loan changes according to the initial equipment's of the borrowers and the costs which they associate with the strategies of the moral risk and the strategic defect. Thus, banks cannot reach a rate of perfect refunding on the basis of the several inciting mechanism of its methodology of loan (Salazar,2008). Banks will have to lay down a new objective as regards the performance of refunding (Bhatt and Tang, 2002). With an aim of not exceeding the new target threshold of defect, institution will grant higher loans to the slightly risky borrowers (Brennan, J. M, and W. N. Torous., 2009).

The main objective of this work is to develop a statistical model that can allow distinguishing the good borrowers from bad. One of the first steps is therefore to define what we mean by good and bad borrowers.

A borrower is considered to be good if he repaid (or has always repaid) correctly its loan and has never been late in paying for thirty (30) days or more.

A bad borrower is a borrower who has experienced at least once a delay in the repayment of its loan for 30 days or more. It is worth mentioning that these definitions arising from the discussions with the credit officers and the team of the credit department of the institution.

3- Bayesian Networks

A network is Bayesian graphical model probabilise. It is defined by:

-A graph oriented acyclic $G = (X, E)$, where $X = \{X_1; X_2; \dots; X_n\}$ is a set of variables (the nodes of the graph) and E a set of arcs. We note $\theta = \{\theta_1; \theta_2; \dots; \theta_n\}$ the set of probability distributions such that:

$$\theta_i = P(X_i / Pa(X_i)) \quad (1)$$

Or $Pa(X_i)$ is the set of nodes, connected to X_i by arcs of end X_i (the parent nodes of X_i). Then we say that $B(G; \theta)$ is a Bayesian Network if and only if:

$$P(X_1, X_2, \dots, X_n) = \prod_1^n \theta_i \quad (2)$$

(Theorem of Bayes)

This decomposition of the joint law of probabilities in a product of local terms is at the origin of the attraction generated by the Bayesian Networks. It is of the "compaction" of this act of joint probabilities that is does a number of algorithms for the calculation in a complex system probabilise. These algorithms will allow a typical use of bayesian networks: the inference.

The distributions of probabilities associated with each of the variables in the model can be either continuous or discrete. In addition, a bayesian network can both contain variables continuous and discrete. The parameters of the discrete variables can be summarized and represented by tables of probabilities conditioned to all possible combinations of the states of the variables "parent".

Each variable is a node of the graph, and takes its values in a discrete set or continuous. The graph is always directed and acyclic. The directed arcs represent a link of direct dependence (most of the time it is causation) .

Thus an arc ranging from the variable X to the Y variable will express that Y depends directly on X . The parameters express the weights given to these relationships and are the conditional probabilities of variables knowing their parents (example: $P(Y|X)$). It is possible to achieve the classifiers thanks to bayesian networks.

Because of this, the probabilities or the scores that will appear by the following represents the scores of reimbursement or the score of creditworthiness of a borrower based on these characteristics. Subsequently, the scores obtained will classify the good borrowers from bad.

Therefore, the behavior of a BN is determined through two parameters: its structure (the nodes and the links among the nodes) and the probability tables associated with the nodes. The structure and the conditional probabilities that are essential for characterizing the network can be either provided externally by experts or obtained from an algorithm which automatically induces them (Cheng *et al*, 2002). According to Heckerman and Geiger (1995), to establish the Bayesian Network structure, an expert is required to design the network with reliance on his/her knowledge about the relations among the variables.

Research has proven that while modeling the expert knowledge is nearly unreliable and time-consuming job, structure learning algorithms have become an important research area. During the previous years, research was centered on the presentation of algorithms whose aim has been

to induce the structure of the Bayesian Network that better represents the conditional independence relationships underlying the data (Friedman and Koller, 2003).

The structural learning methods generally require two components: the learning algorithm and the evaluation metric that measures the goodness of the net during each learning step. According to Heckerman and Geiger (1995), most distinct approaches of the structural learning mentioned in the literature are related to multiple connected networks and are grouped according to the extent to which they are necessary for imposing order on the variables.

Setting an order among the variables implies that a variable X_i can have the variable X_j as a parent only if X_j proceeds X_i in the established order.

According to Robinson (1977), if the ordering between the nodes is not established, the cardinality of the search space becomes bigger, and the number of networks grows hyper-exponentially.

The first stage of construction of the Bayesian network is the only one for which the human intervention is absolutely essential. It is to determine the set of variables X_i , categorical or numeric, which characterize the system.

As in any modeling work, a compromise between the accuracy of the representation and the utility of the model must be found, by means of a discussion between the experts and the modeler.

When the variables are identified, it is then necessary to specify the space of states of each variable X_i , i.e. the set of its possible values.

4- Data and methodological issues

Data from a Tunisian bank contain information about household's credit holders whose allocations were approved between 2013 and 2014. This database is employed to illustrate the proposed Bayesian Network scoring model. The original data contain 6240 customers, where 84.49 % of them are good customers and 15.51% are bad customers. The collected data yielded 9 variables which are identified in Table 1.

The characteristics that are usually used in consumer application credit scoring include the type of credit, gender, Activity area , credit amount, Degree of solvability, Term of loan, credit amount, etc.

Data included 3 demographic variables and 5 credit variables from the Tunisian Central Bank such as DEGREE OF SOLVABILITY, outstanding credits, type of credit, credit amount and credit duration. Table 1 shows a description of the variables considered.

Table 1: list of variables

X ₁ : Age	X ₄ : Level of graduation	X ₇ : Amount
X ₂ : Area	X ₅ : Activity area	X ₈ : Term of loan
X ₃ : Gender	X ₆ : Civil status	X ₉ : Degree of solvability

This table shows different variables of our study based in 6240 consumers.

According to Jensen (1996), obtaining a Bayesian Network requires the fact of identifying both the structure, which defined by a directed acyclic graph, and the conditional probabilities attributed to each node of the directed acyclic graph.

In this sense, obtaining a Bayesian Network rest upon two criteria: the structural learning and the parametric learning.

The *former* refers to the identification of the Bayesian Network typology and the latter refers to the estimation of numerical parameters, known as conditional probabilities.

In the BN, the structure in the directed acyclic graph may be determined either by expert knowledge or by learning algorithms. In fact, discovering the structure from a data set is a difficult task because it is proven that many structures are consistent with the same set of independencies (Margaritis, 2003; Koller and Friedman, 2010). Therefore, as Cheng *et al.* (2002) have maintained, the problem of discovering the causal structure increases with the number of variables.

In *score-based* structure learning, a score is assigned to each BN structure on the basis of the extent to which the model fits the data, and a model structure with the highest score is applied. These methods require the following procedures ; a scoring metric to measure the quality of every candidate BN with respect to a dataset, and a search procedure to move through the space of possible networks.

In *constraint-based* learning, the input is a set of conditional independence relations between subsets of variables. These learning algorithms are applied in order to build a BN that represents a large percentage (and, whenever possible, all) of the relations (Spirtes, Glymour and Schneines, 2001).

Thus, the adoption of this analysis yields a generated undirected graph. Then, when making an additional independence test, the network is transformed into a Bayesian Network.

With reference to Hojsgaard, Edwards, and Lauritzen's (2012) work, the model selection process was restricted by blacklisting arrows that point from a later block to an earlier block. So, obtaining the structure rests upon two options: we either select a single best model or else obtain some average model, which is known as model averaging (Claeskens and Hjort, 2008). The search was conducted using available learning algorithms that are included in the BNT package as a single

best model. Furthermore, to select the learning algorithm, we have to look for the plausibility of the model and the sparser graphs.

As Figure 1 indicates, some relationships between variables are easy to decode. For instance, the credit duration as well as degree of solvability has a direct influence on default payment. The parents of the variable « credit duration » are «the outstanding of credit» and «profession». The credit amount has a direct effect on the degree of solvability, which, in turn, acts on default payments. The types of credit, the outstanding credit and the credit amount indirectly act on default payments. Besides, profession has an indirect effect on payment default. This latter variable is affected by gender and age. Furthermore, the variable of credit amount 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 stays with the bank, the more the bank has information about his/her banking behavior, which serves lowering the probability of default. However, this variable needs to be updated regularly due to adverse and unexpected changes in the borrowers' situation.

Parameters were obtained again with the *BNT* package in MATLAB (2015a) by applying a Bayesian parameter estimation using the Dirichlet distribution (Neapolitan, 2003).

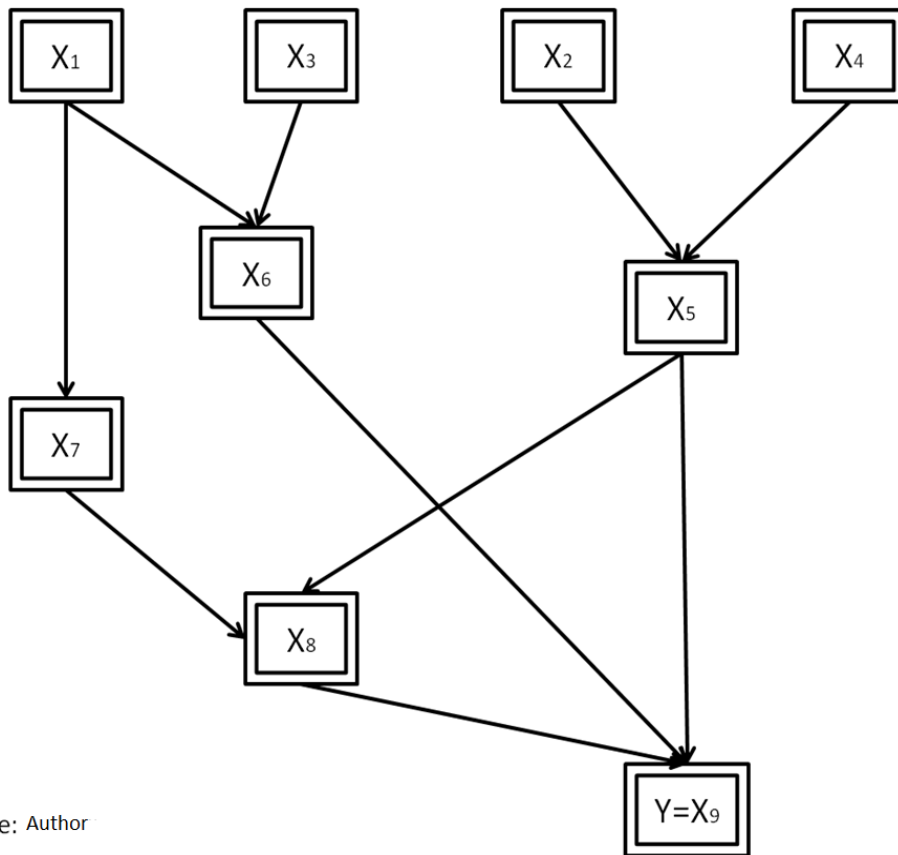
Let W be a dataset and let N_{ijk} be the number of cases in W in which the node i is in state k and its parents are in state j that is $X_i = x_i^k$ and $Pa(X_i) = x_i^j$.

The distribution of $(N_{ij1}, \dots, N_{ijr_i})$ is multinomial with parameters $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$ and $\theta_{ij} = (\theta_{ij1}, \dots, \theta_{ijr_i})$, where $\theta_{ijk} = P(X_i = x_i^k | Pa(X_i) = x_i^j)$ and the Bayesian estimation of θ_{ijk} is given by:

$$\hat{\theta}_{ijk} = \frac{N_{ijk} + \alpha_{ijk}}{N_{ij} + \alpha_{ij}} \quad (3)$$

Considering that in our case $\alpha_{ijk} = 1$. A conditional probability distribution is obtained for each node. The graph below illustrates an example of conditional probability distribution.

Figure1: Graph of study



Source: Author

The graph of the study is presented in Figure 1. It presents the different relations between the variables in the form of nodes and parents.

As graph indicates, 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 to the probabilistic model.

With regard to the parents X_5 : credit duration and degree of solvability variable (X_9) with reference to the analysis of the relationships between the variables, it is clear that most people who have payment default (30.22%) are those who obtain consumer credit whose repayment period is between (0-84 months) and their degree of solvability varies from 0 to 100 dinars.

Besides, 25% of respondents having default payment are those whose repayment period ranges from 0 to 84 months. So, they get a consumer credit and their DEGREE OF SOLVABILITY is greater than 1000 dinars.

Furthermore, 24.22% are those who were granted housing credits and the duration of which exceeds 84 months and that their degree of solvability is between 100 and 200 dinars.

Finally, 22.96% of applicants get housing credits whose repayment duration exceeds 84 months and whose degree of solvability varies between 100 and 200 dinars.

Therefore, the effects of various factors are instances of causal reasoning or prediction of variables related to non-payment. It is worth signaling that on the basis of parametric learning, consumer credit and degree of solvability are two of the most important indicators of default payments. This indicates that most Tunisian creditors belong to a middle class, which proves the increase in consumer credit rates in late 2007, reaching a volume of 2,559 MD (total credits of 6.395 MD).

5- Results and Discussions

In a Bayesian network, inference refers to a probability calculation afterwards (Buz *et al.*, 2009). Knowing the statements of several variables (called observation variables *e*), we determine the state probabilities of other variables (called target variables *X*) which rest upon conditional observations $p(X / e)$.

This sustains Koller and Friedman’s (2010) view that finding a high-probability joint assignment to some subset of features constitutes a very important task.

As Table 2 indicates, the probability of the credit type allocated to borrowers with default payments is slightly higher for those who were granted housing loans (58.14%) than those who obtained consumer credit.

Table 2: Table of frequencies of the 43 selected configurations

Loan Duration	0	1	
Frequency	58,14%	41,86%	
Gender	0	1	
Frequency	30,23%	69,77%	
localisation	0	1	
Frequency	32,56%	67,44%	
Age	0	1	2
Frequency	13,95%	83,72%	2,33%

Table 2 shows the frequency of the 43 configurations according to the main characteristics of our sample

Default payments result mainly from retired customers (52.62%) and from those who grant loans of less than 5000 dinars (Table 3). Then, 44, 88% of default payments are resulted from customers who were granted credits ranging from 5000 to 30000 dinars (Table 3). More importantly, our results indicate that younger households increase the creditworthiness since they have fewer commitments as compared to elder applicants (76, 88%) (Table 3). Finally, the majority of bad debtors (52.68%) have degree of solvability which does not exceed 100 Dinars.

The use of Bayesian network model has allowed us to identify the most probable states of borrowers who could not repay their debt. Based on applicants’ profiles, we can build a decision support system that serves to distinguish between good and bad borrowers.

There were only 10% female default borrowers as compared to 90 % of male bad debtors. So, females have less probability of default as compared to 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 Mohd Zaini, 2009). They make less default payment since they are commonly averse to risk, ethically stick to hard work, and maintain the culture of financial discipline (Pitt and Khandker, 1998; Bhatt and Tang, 2002; Croson and Gneezy, 2009). Then, our findings also indicate that those who obtain housing credit have more probability to default payment (60%) compared to consumer credit (40%). Thus, our results are fine-tuned with other authors' findings related to customers credit worthiness in connection with the risk associated to housing credit repayment (Schreiner, 2004; Dinh and Kleimeier, 2007).

Table 3: Classification default payment

	Gender	Branch of activity		Education level	
Male	4,87%	Farmer	19,19%	Illiterate	9,23%
Female	95,11%	Trade	61,81%	Bac	86,91%
		Services	12,91%	Superior	3,58%
		Others	6,06%		

Table 3 shows a classification of clients and their risk of non-reimbursement by age, sector of activity and level of education

As Table 3 illustrates, we also conclude that elder debtors are more likely default payment than less aged ones. This finding confirms the results of Arminger et al.'s (1997) which 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 Thomas (2000) and Boyle *et al.*'s (1992) 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 hesitant 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 could support the fact that they have the highest risk of falling into default payment.

This variable is significant as long as it is highly correlated with income. In this sense, an applicant's profession may indicate whether he/she has a high and stable income. Moreover, credit duration is considered as a predictive variable of default payment. Thus, most debtors who made default payments are those whose loan duration ranges from 0 to 84 months. This explains the fact that credits allocated to customers could be easily paid back when the duration of repayment is shorter.

In summary, BNs are a graph-based structure of a joint-multivariate probability distribution which captures the way an expert establishes the relationship between variables. Hence, the BN model provides a unified formalism for handling uncertainty and risk.

6- Conclusion

Recent research in credit scoring emphasizes the importance of not only distinguishing 'good' customers from 'bad' ones, but also predicting in advance when customers can make default payments. Such a prediction enables banks 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.

Bayesian Networks (BNs) have been chosen in order to produce an intuitive, transparent, graphical representation of the investigated interdependencies. The proposed decision model can be used to build bank decision support systems to deal with the issue of default payment in Tunisian commercial banks.

The proposed BN modeling involved a domain expert elicitation and a learning phase using an available dataset, which makes the decision model more robust and reliable. The BN structure was built by using K2 algorithm for learning structure which assumes apriori an ordering on the variables. On the basis of the results of our experimental data (6240 customers), the developed decision system shows that gender, as one of the most used socio-demographic variables, is a significant predictor since our findings indicated that those who have the greatest likelihood of default payment are male customers.

Equally important, the majority of those who make default payments 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, though our findings are effective in so far as they allow banks to adopt a sensible decision policy, 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 the distinction between ancient and new customers.

References

- Acid S., Campos L., Huete J., The Search of Causal Orderings: A Short Cut for Learning Belief Networks, In Proc. 8-th conference on Uncertainty in Artificial Intelligence, (2001)
- Arminger, G., Enache, D. & Bonne, T. (1997). « Analyzing Credit Risk Data: A Comparison of Logistic Discrimination, Classification Tree Analysis, and Feed-forward Network », *Computational Statistics*, Vol. 12, Issue 2, pp. 293-310.
- 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*, Vol. 26, No. 6, pp. 360-76.
- Boyle, M., Crook, J., Hamilton, R. and Thomas, L, 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, 1992, 75-90.
- Brigham, E. F. (1992). *Fundamentals of financial management (6th ed.)*. Forth Worth: Dryden Scan. J. Stat. 24, 1-13.
- Castillo E., Gutierrez J.M., Hadi A.S (1997). *Expert Systems and Probabilistic Network Models*. Springer, New York, 605 pp.
- Cooper and Herskovits, (1992) G.F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 309-347.
- F.V. Jensen (1996). *An introduction to Bayesian Networks*. Taylor and Francis, London, United Kingdom.
- G.F. Cooper, (1990) The computational complexity of probabilistic inference using Bayesian Belief Networks, *Artificial Intelligence*, Vol. 42 (23), pp. 393- 405.
- Croson, R. and U. Gneezy (2009), Gender Differences in Preferences, *Journal of Economic Literature*, Vol.47(2), pp.448-474.
- D. Heckerman, D. Geiger, and D. Chickering (1995). Learning bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, Vol. 20, pp. 194-243.
- Daly, R., Shen Q., Aitken S (2011), Learning Bayesian Networks: Approaches and Issues. *The Knowledge Engineering Review*, Vol.26 (2), pp. 99-157
- Davis D.B, *Artificial Intelligence goes to work* . High Technol, (1987), Apr 16-17.
- Deschaine L., FRANCONI F, Comparison of Discipulus TM Linear Genetic Programming Software with Support Vector Machines, Classification Trees, Neural Networks and Human Experts, White Paper. Available at: <http://www.rmltech.com/> (Accessed: 10 June 2008).
- Dinh, T.H.T. and S. Kleimeier, Dinh(2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis*, Vol.16(5), pp. 471-495.
- D. Margaritis, *Learning Bayesian network model structure from data*, (2003) (PhD Thesis of CMU-CS-03-153).

- D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques, The MIT Press, Cambridge, MA/London, England, (2010).
- E. Castillo, J.M. Gutierrez, A.S. Hadi, Expert Systems and Probabilistic Network Models, Springer-Verlag, (1997).
- Fan Liu, Zhongsheng Hua, Andrew Lim. (2015), Identifying future defaulters: A hierarchical Bayesian method, European Journal of Operational Research, Vol. 241, pp. 202-211.
- Fatemeh Jamaloo, et al. "Discriminative CSP Sub-Bands Weighting Based On DSLVQ Method In Motor Imagery Based BCI" International Journal of Online Engineering (IJOE), 5 (3), (2015), 156-161.
- G. Claeskens, N.L. Hjort, Model Selection and Model Averaging, Cambridge University Press, Cambridge, (2008).
- Hardy W. E., Jr., Adrian J. L. Jr. (1985), A linear programming alternative to discriminant analysis in credit scoring. Agribusiness, Vol. 1(4), pp. 285-292.
- S. Hojsgaard, D. Edwards, S. Lauritzen, Graphical Models with R, Springer, New York, 2012.
- J. Cheng, R. Greiner, J. Kelly, D. Bell, W. Liu (2002), Learning Bayesian networks from data: an information-theory based approach, Artificial Intelligence, Vol.137, pp.43-90.
- J. Pearl, Probabilistic Reasoning in Intelligent Systems, Networks of Plausible Inference, Morgan Kaufmann, (1988).
- Johnson R. W., Kallberg J. G., Management of accounts receivable and payable. New York: Wiley,(1988).
- Korb K. B., Nicholson A., Bayesian artificial intelligence. Chapman, Hall/CRC, Boca Raton, FL, 2nd edition, (2010).
- L.E. Sucar and M. Martinez-Arroyo (1998), Interactive structural learning of bayesian networks. Expert Systems with Applications, Vol.15, pp. 325-332.
- M. Henrion, Propagating uncertainty in Bayesian Networks by probabilistic logic sampling, (1988) Proceedings of the Fourth Conference on Uncertainty in Artificial Intelligence, pp. 149-163.
- Mouley S., 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, (2014) Number 4.
- N. Friedman, D. Koller, Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian Networks (2003), Machine Learning, Vol. 50, pp. 95-125.
- R.E. Neapolitan, Learning Bayesian Networks, Prentice Hall, Inc., Upper Saddle River, NJ, USA, 2003.
- Oreski, S., Oreski, G (2014), Genetic algorithm-based heuristic for feature selection in credit risk assessment. Expert Systems with Applications, Vol.41(4), 2052-2064.

- Pitt, M.M. & Khandker, S. R. (1998), The impact of group-based credit programs on the poor households in Bangladesh: does the gender of participants matter? *Journal of Political Economy*, Vol.106 (5), pp.958-996.
- P. Spirtes, C. Glymour, R. Scheines, *Causation, Prediction and Search*, Adaptive Computation and Machine Learning, 2nd ed., The MIT Press, (2001).
- Roslan, A. H. and A. K. Mohd Zaini (2009), « Determinants of Microcredit Repayment in Malaysia: The Case of Agrobank », *Humanity and Social Sciences Journal*, Vol.4(1), pp.45-52.
- Salazar, G. L. (2008), « An Analysis of Repayment among Clients of the Microfinance Institution Esperanza International, Dominican Republic », *Journal of Agricultural Economics*, Vol. 90 (5), pp.1366-1391.
- Schreiner, M., (2004), « Benefits and pitfalls of statistical credit scoring for microfinance ». *Savings and Development*, Vol.28(1), pp. 63-86.
- Tenenhaus M. (2000), *La Régression Logistique PLS*, Journées d'Etudes en Statistique, Modèles Statistiques pour données Qualitatives, 261-273.
- Thomas. L, *A Survey of Credit and Behavioural Scoring; Forecasting financial- risk of lending to consumers* (2000), *International Journal of Forecasting*, Vol.16(2), pp.149 -172.
- Vaclav K. (2015), *Genetic algorithms for credit scoring : Alternative fitness function performance comparison*, *Expert Systems with Applications*, Vol. 42, pp. 2998-3004.
- Van Gool, J., Verbeke, W., Sercu, P., Baesens, B. (2011), « Credit scoring for microfinance: is it worth it? » *International Journal of Finance & Economics*, Vol.17(2), pp.103-123.
- Viganò, L. (1993), « A Credit-scoring Model for Development Banks: An African Case Study », *Savings and Development*, Vol. 17(4), pp.441-482.