Khemakhem, Sihem and Boujelbene, Younes "Artificial Intelligence for Credit Risk Assessment: Artificial Neural Network and Support Vector Machines" *ACRN Oxford Journal of Finance and Risk Perspectives 6.2 (2017): 1-17.*

ARTIFICIAL INTELLIGENCE FOR CREDIT RISK ASSESSMENT: ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINES

SIHEM KHEMAKHEM[1], YOUNES BOUJELBENE[2]
[1] University of Sfax, Faculty of Economics and Management of Sfax
[2] University of Sfax, Faculty of Economics and Management of Sfax

***Abstract**: This work tries to determine the probability of default as a tool to measure credit risk in a Tunisian commercial bank. A scoring model was built according to the traditional technique of logistic regression (LR), and artificial intelligence techniques i.e. artificial neural networks (ANN) and support vector machines (SVM). Then a comparison was made between these models using performance metrics such as the confusion matrix and the area under the ROC curve (AUC) in order to identify the most efficient model. Our results show that the Radial Basis Function kernel SVM was the most performing method in terms of accuracy, sensitivity and specificity with the least error rates. Thus, in the Tunisia context, this model is worth implementing in banking institutions in order to improve their credit risk management measures to monitor and control credit.*

***Keywords**: Credit risk; Support Vector Machines; Artificial Neural Network; Logistic Regression; Performance metrics*

Introduction

Credit risk refers to the possibility that debtors to a bank do not meet their commitments by failing to meet loan repayment deadlines or cease to pay off their credit (Apostolik et al., 2009). In fact, credit risk is the most critical and the greatest challenge facing the management of banks.

Several works have dealt with bank failures and concluded that the quality of assets was a statistically significant leading indicator of insolvency (Dermirgue-Kunt & Detragiache, 1989; Barr & Siems, 1994), and that financial institutions around the world still have a high level of non-performing loans.

In Tunisia, according to the financial statistics of the central bank[1], the credit risk is constantly growing due to the increase of credit to the economy. Indeed, the overall amount of credit increased from 36352 MTD in 2009 to 61464 MTD in 2014, accounting for 69.08%. In addition, a general problem witnessed by all Tunisian banks is the constant increase of the number of non-performing loans. Due to the impact of the regression in economic activity in the country after the revolution, Tunisian banks credit risk has drastically increased with a raise of non-performing loans of around 14.5% from 2009 to 2014. In 2009, the rate was 13.2% but it reached 16.2% in 2014. It is worth noting that this problem is getting more and more serious during the last years, which shows the failure of the management authorities of the Tunisian banking institutions to overcome this growing problem by applying more accurate and reliable measures to avoid risk and minimize the rate of insolvent credit. Today, Credit risk prediction has become more and more important to save the Tunisian banking industry. This situation

[1] The Tunisian Central Bank. Available at: http://www.bct.gov.tn (last accessed 10/24/2015).

could cause a local financial crisis. Finally, the Tunisian commercial banks are considered financial institutions of vital importance seeking profit by providing various financial services to individuals and businesses while managing different types of risk. Therefore, risk-taking is often regarded as the basic drive of financial performance and profitability (Bekhet & Eletter, 2014).

The inability of creditors to accurately assess the credit risk of potential borrowers has a catastrophic impact on the financial system and economic activity in general. During the last decade, the credit risk analysis has attracted significant attention of decision makers in financial institutions worldwide. This was partly due to the global economic crisis and recent changes in financial legal systems (e.g. Basel III). In addition, the increased competition in the banking sector has led many institutions to find innovative ways of risk prevention rules in order to maintain their competitiveness (Harris, 2013). Consequently, in the current economic and business environment, financial institutions face a higher risk of losses associated with the inappropriate credit authorization decisions (Yu et al., 2008). To manage the increased default risk facing financial institutions, more efficient credit assessment techniques are constantly being developed. In this context, the development of new risk analysis tools aims to improve the ability of banks to predict the risk class of companies that ask for credit.

Credit scoring is a key instrument for financial institutions to assess the credit risk and make management decisions. In practice, credit scoring refers to a classification problem where a new credit applicant must be classified in one of the predefined classes (usually "good" and "poor" customers, depending on whether they are likely to default on their payments).

Most conventional approaches for credit scoring are based on statistical parametric models such as linear regression, discriminate analysis, and logistic regression (LR). However, the modern credit scoring has been oriented to implementing non parametric methods and artificial intelligence techniques like decision trees, artificial neural networks (ANN) and support vector machines (SVM). In contrast to the parametric statistical methods, these alternative models do not require any specific prior knowledge. They are used to automatically extract knowledge from learning observations.

In this study, we are motivated by the application of modern methods of credit risk assessment to estimate the default probability of the borrower. In this sense, our research problem is to consider other techniques than traditional methods to explain and predict the counterparty risk. To do so, the following question is addressed: To what extent can modeling and prediction artificial intelligence methods reduce the credit risk estimation errors?

To address this problem, the literature provides some replies. Traditional credit risk assessment methods are usually based on the experience and judgment of banking staff. However, with the increasing number of credit applicants, these conventional approaches have become obsolete because they cannot meet the need for effective credit risk evaluation needs. It has been shown that artificial intelligence techniques are effective and efficient compared with conventional statistical methods in the field of finance (Lin et al., 2009). In addition, researchers have attempted to model the non-refund, and relied on other techniques to explain and predict the counterparty risk. They proposed the use of data mining techniques such as ANN and SVM (Belotti & Crook, 2009; Huang et al., 2004; Wang et al., 2005; Harris, 2013), to find the hidden information in the database using advanced algorithms (Han & Kamber, 2001).

This research leads to a comparative analysis of the predictive capabilities of three data mining techniques: LR based in part on the traditional scoring method; and SVM and ANN on the other hand, which are artificial intelligence methods. Eventually, the most robust model is considered to form the desired score function that allows us to understand such risk and improve the performance of the Tunisian banking system. Our research seeks to improve the credit scoring system and therefore increase the deployment of intelligent technologies that substitute human thought and cope with the complexity of real-world problems. Therefore, any improvement in financial prediction systems can be translated into huge savings (West, 2000). The remainder of this work is structured as follows: Section 2 provides a detailed literature review. Section 3 describes the main methodological steps used in our research. Section 4 presents empirical results. Section 5 provides conclusions and perspectives to future research.

Literature Review

Recent empirical work was oriented to checking the predictive power of artificial intelligence methods for credit risk estimation problems such as ANN and SVM (Liu et al., 2011; Malley et al., 2012; Wu & Liu, 2007). They offer a different view on the credit scoring issue. They can be promising alternatives to standard statistical methods (Kruppa et al., 2013).

In contrast to traditional statistical techniques, artificial intelligence techniques do not involve data distributions. These techniques automatically extract knowledge from the training samples (Wang et al., 2012). Previous studies have shown that artificial intelligence techniques are better than the statistical techniques in dealing with credit scoring problems, particularly for the classification of non-linear models (Huang et al., 2004). Li et al. (2006) developed a credit assessment model using SVM to identify potential loan applicants. The results revealed that the SVM model outperforms the model of ANN in terms of generalization. Huang et al. (2004) studied the performance of the SVM approach in attributing credit scores. They compared the results generated by SVM with the back propagation neural networks. However, SVM presented a slight improvement as compared to neural networks. Hung & Chen (2009) suggested a selective ensemble of three classifiers: the decision tree, ANN and SVM to assess the credit risk. Based on predicted probabilities of credit risk, this comprehensive method provides an approach that inherits the advantages and avoids the disadvantages of different classification methods. Research by Karaa & Krichen (2012) aimed to compare the predictive power of two prediction methods of credit risk: ANN and SVM. In their study, they focused on the operating loans granted to Tunisian industrial companies. They used a database composed of 1435 credit records covering the periods 2002, 2003, 2004, 2005 and 2006. The results showed the superiority of ANN to SVM in terms of accuracy and the reduction of type I error. In fact, the entire sample accuracy is 90.2%, (ANN) and 70.13% (SVM) and type I error is approximately 18.55% (ANN) and 29.91% (SVM). Nnamdi & Shola (2011) compared the precision of SVM and ANN when applied to credit risk evaluation. They concluded in the light of experimental results that the ANN system outperforms the SVM in terms of accuracy (85.305% for ANN and 84% for SVM) and the minimum learning time. Salehi & Mansoury (2011) explored the effectiveness of ANN and LR in the prediction of credit risk. They claimed that both models are similarly efficient. Tsai & Wu (2008) and Burger & Hofinger (2005) reported that the neural network model formed by the gradient retro-propagation algorithm is a popular financial decision-making tool. The precision of this model outperforms that of other models in terms of forecasting, such as LR, discriminant analysis, k-nearest neighbors and decision trees. Belloti & Crook (2009) carried out a comparative analysis between SVM, LR, discriminant analysis and k-nearest neighbors. They found that the SVM was the most successful in classifying credit applicants. They concluded that the SVM technique can be used successfully as a method for determining the risk of default. Yeh & Lien (2009) compared the predictive efficiency of the probability of default among six data mining techniques: discriminant analysis, LR, ANN, k-nearest neighbors, Naïve Bayesian classifier and the decision tree. The results showed that the forecast of the default probability by the ANN is the only method that could be used to represent the real probability of default. Therefore, this technique should be used to determine client scores rather than other data mining techniques.

Research methodology

The objective of our study is to examine the performance of two advanced data mining methods; ANN, SVM with the well-known LR to predict credit risk. This section describes the data used for learning and testing models. It is also important to understand the characteristics of the data through the study of the relationship that may exist between the variables.

Description of the database: sample and variables

We opted in this study for the operating loans granted to Tunisian companies. All credit records were collected from a Tunisian commercial bank covering the years 2011 and 2012. Out of the total of 408 observations, 300 observations successfully completed their credit obligations and were classified as creditworthy borrowers and 108 observations were delayed in the execution of their duties and were classified in a group of non-creditworthy borrowers. This research used a binary variable - Credit (Solvent = 1, Non solvent = 0) as a response variable. This study used the 25 following variables as independent variables where 03 are binary qualitative variables and 22 are quantitative variables:

- V1: Financial Profitability (Net income / Net Equity)
- V2: Operating profitability (Gross operating surplus / Turnover)
- V3: Economic profitability (Operating income / Economic Assets)
- V4: Net profitability (Net income / Sales)
- V5: Financial autonomy (Shareholders' equity / Permanent Capital)
- V6: Structural Balance (Permanent Capital / Fixed assets)
- V7: WC coverage (Working capital / working capital requirement)
- V8: Solvency (Net capital / Total assets)
- V9: Asset coverage (Net capital / Fixed assets)
- V10: Long/Medium Term Debt / Fixed assets
- V11: Financial dependence (Long & Medium Term Debt / Permanent Equity)
- V12: Repayment capacity (Long & Medium Term / Cash Flow Net debt)
- V13: Debt ratio (Financial charges / turnover)
- V14: Financial burden (Financial expenses / Gross operating surplus)
- V15: Working capital ratio (Turnover / total fixed assets)
- V16: Inventory turnover ratio (Turnover / net stocks)
- V17: Devoted Turnover (Movement / Sales)
- V18: Share of funding (Bank commitment / banking system commitment)
- V19: Study duration of a credit report (Log (study period))
- V20: Corporate banking relationship Duration (1 if the relationship length ≥15 months; 0 otherwise)
- V21: Guarantees (Log (guarantees))
- V22: Size of the company (Log (turnover))
- V23: Score: credit line number (Log (Score))
- V24: Ownership structure (1 if the officer holds more than 50% of the capital; otherwise)
- V25: Legal form (1 = SARL; 0 otherwise)

Study of linear dependencies between variables

In this part, we study the linear dependencies between variables by the correlation matrix. In this development, we present the level and signs of correlation between the variables in our study. To do so, we used a Pearson test for the crossing of two quantitative variables. Pearson's correlation coefficient is used to characterize a positive or negative linear relationship. The threshold of the correlation coefficient is equal to 0.8 (Hai et al., 2013). In other words, if the absolute value of the correlation coefficient of two indicators is larger than the threshold of 0.8, it is stated that the two indicators reflect the repeated information and one of them can be deleted. Secondly, to study the relationship between a qualitative variable and a quantitative variable, we calculated a parameter called the correlation ratio noted η^2 using the function eta2 ( ) of R software. η^2 is the ratio of the intergroup variation i.e. the square of the differences between the group average and the global average and the total variation i.e. the sum of squared deviations from the mean. The correlation ratio varies between 0 and 1. If the ratio is close to 0, the two variables are not correlated. If the ratio is equal to 1, the variables are correlated. Finally, we used the Chi test (χ^2) for crossing two variables. The χ^2 test is used to determine if the link between these two variables is significant or not. The statistical significance was set at $p < 0.005$.

The correlation matrix presented in Table 1 shows that the level of correlation is very low, except for the variables (V2 / V3), (V2 / V4), (V3 / V4), (V5 / V11), (V6 / V15) and (V24 / V25). Thus, the V20 and V24 variables are highly correlated with the dependent variable "Credit". This matrix shows that all variables in our study are positively correlated with the dependent variable "Credit".

Table 1 : Confusion matrix

	Credit	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25
Credit	1																									
V1	0.004	1																								
V2	0.018	0.007	1																							
V3	0.021	0.047	**0.892**	1																						
V4	0.016	0.043	**0.933**	**0.959**	1																					
V5	0.023	-0.051	0.081	0.106	0.118	1																				
V6	0.000	-0.013	0.003	-0.001	0.005	-0.015	1																			
V7	0.005	-0.023	0.043	0.057	0.039	0.169	-0.015	1																		
V8	0.017	-0.121	0.124	0.054	0.054	0.464	0.002	0.255	1																	
V9	0.015	-0.018	0.013	0.055	0.036	0.172	0.100	0.030	0.122	1																
V10	0.037	0.000	-0.031	0.011	-0.014	-0.325	0.030	-0.021	-0.248	0.198	1															
V11	0.016	0.035	-0.078	-0.111	-0.119	**-0.936**	0.017	-0.166	-0.447	-0.176	0.316	1														
V12	0.027	0.020	-0.012	-0.016	-0.003	-0.326	-0.003	-0.010	-0.145	-0.058	0.355	0.320	1													
V13	0.008	-0.052	-0.746	-0.746	-0.790	-0.146	-0.016	-0.053	-0.055	-0.048	0.052	0.125	0.006	1												
V14	0.002	-0.059	-0.021	0.002	0.007	-0.086	-0.008	-0.043	-0.136	-0.028	0.052	0.082	0.067	0.003	1											
V15	0.005	-0.006	-0.016	0.015	0.010	0.029	**0.888**	-0.013	-0.072	0.323	0.209	-0.029	-0.017	-0.035	0.004	1										
V16	0.010	-0.023	0.057	0.027	0.019	-0.012	-0.020	0.098	0.078	-0.076	-0.026	0.008	0.001	-0.022	-0.027	-0.033	1									
V17	0.004	0.023	-0.005	-0.020	-0.013	0.009	-0.001	0.045	-0.025	-0.051	-0.013	-0.024	-0.025	0.102	-0.123	-0.030	-0.029	1								
V18	0.001	0.000	0.004	0.017	0.010	0.027	-0.007	0.001	-0.017	-0.013	0.007	-0.028	-0.007	-0.012	-0.014	-0.010	-0.007	0.043	1							
V19	0.017	-0.096	-0.019	0.034	0.010	-0.090	-0.036	-0.024	-0.152	-0.097	0.130	0.078	0.022	-0.018	0.081	-0.015	-0.040	-0.035	-0.030	1						
V20	**0.000**	0.001	0.000	0.000	0.000	0.000	0.000	0.002	0.003	0.001	0.002	0.001	0.001	0.003	0.000	0.000	0.000	0.015	0.001	0.001	1					
V21	0.030	-0.067	0.078	0.025	0.045	-0.052	-0.015	-0.122	-0.116	-0.010	0.021	0.040	-0.035	-0.004	0.120	-0.005	-0.026	-0.105	0.053	-0.021	0.005	1				
V22	0.115	-0.055	0.085	0.099	0.047	0.010	0.003	-0.069	-0.076	-0.008	0.011	0.042	-0.013	-0.005	0.071	0.025	0.088	-0.024	0.026	0.012	0.122	0.357	1			
V23	0.075	-0.023	0.049	0.117	0.080	0.012	-0.099	-0.078	-0.066	-0.074	-0.013	0.014	-0.063	0.000	0.096	-0.098	-0.082	-0.014	0.050	0.078	0.037	0.395	0.341	1		
V24	**0.000**	0.006	0.025	0.017	0.015	0.000	0.001	0.002	0.021	0.002	0.000	0.000	0.008	0.005	0.004	0.000	0.011	0.002	0.001	0.001	0.427	0.000	0.001	0.028	1	
V25	0.067	0.004	0.006	0.017	0.013	0.001	0.002	0.003	0.014	0.019	0.007	0.004	0.001	0.007	0.000	0.009	0.005	0.000	0.001	0.006	0.325	0.012	0.007	0.013	**0.000**	1

Data Preparation

Automatic selection of the independent variables

In bank credit data, more variables can be collected and some of them are redundant or irrelevant. The discarding of such variables can improve the performance of prediction models (Fukunaga, 1972).

In order to select the most relevant variables, we perform an automatic selection of variables by the stepwise method. It is a mixed method which presents a combination of the bottom-up and the top-down selection methods. The procedure starts with a blank template, and then at each step, a new variable among the variables selected by the above procedure is added. The new model is then evaluated using an optimization criterion. Then all variables in the model are evaluated accordingly. The procedure stops when all the variables outside the model provide no improvement to the studied model. To implement this method, we used the Stepwise procedure of R software. The following variables thus selected are used later in this work: Operating profitability (V2), net profitability (V4), solvency (V8), asset coverage (V9), long & medium term debt / fixed assets (V10), financial dependence (V11), repayment capacity (V12), inventory turnover ratio (V16), share of funding (V18), study duration of a credit report (V19), corporate banking relationship duration (V20), size of the company (V22), score: credit line number (V23) and capital structure (V24).

Training and testing samples

The subsequent preparation is to divide the available applicants into a training set and a testing set. The first set (70% of the sample) is used to design various models and build assignment rules of an individual based on these characteristics. The second set (30% of the sample) serves to check whether the model based on the training sample is statistically reliable.

The data mining techniques in credit scoring

The credit scoring methods used to estimate the credit risk are listed below.

Logistic Regression

The origin of this approach dates back to the work of Wiginton (1980). In LR models, the dependent variable is generally binary and the independent variables can be continuous or categorical. According to Agresti (2002), the binary LR is used to estimate the effect of an explanatory variable (X) in the response variable (Y) that is whether binary or dichotomous.
The aim of LR modeling is to estimate credit risk and extract variables found important in credit risk prediction.

In the model of LR, Z is the linear combination of k ($k = 1,2,\ldots,K$) weighted independent variables by logistic factors:

$$\boldsymbol{Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k} \tag{1}$$

$$\boldsymbol{\hat{y} = \frac{1}{1+e^{-Z}}} \tag{2}$$

To estimate the parameters of LR, we use the maximum likelihood method to find β that maximizes $L(\beta)$. The log likelihood function is defined as:
$\beta = (\beta_0, \beta_1, \ldots, \beta_k)$

$$\boldsymbol{L(\beta) = \sum_{t=1}^{n}(y_t \ln \hat{y}_t + (1 - y_t)\ln(1 - \hat{y}_t)) = log - likelihood = LL(\beta)} \tag{3}$$

In order to check the importance of the estimate parameters as a whole, we use the likelihood ratio test.

$$\boldsymbol{G^2 = -2\ln\left[\frac{L(\text{reduced model})}{L(\text{complete model})}\right]}$$

$$\boldsymbol{G^2 = [-2 \times LL(\text{reduced model})] - [-2 \times LL(\text{complete model})]} \tag{4}$$

The purpose of this test is to check the difference between complete and reduced models in terms of significance.

The hypothesis for the likelihood ratio test is $H_0: \ \hat{\beta}_0 = \hat{\beta}_1 = \cdots = \hat{\beta}_k = 0 \ (k = 0,1,\dots,K)$ with the alternative $H_1: \ \exists\hat{\beta}_0 \neq \hat{\beta}_1 \neq \cdots \neq \hat{\beta}_k \neq 0$. The G^2 statistic is an asymptotically distributed Chi-Square (χ^2) at $(1 - m)$ degrees of freedom (df), with m the number of mode parameters (Bolton, 2009). The test criteria are: discard H_0 if the statistic G^2 is higher than the χ^2 in the table for a specified threshold, generally 5%. The estimated model has at least a significant explanatory variable. Otherwise, accept H_0.

Artificial Neural Networks

To develop new information processing tools, an interesting approach is to mimic the behavior of the human nervous system and its evolution mechanisms. The ANN is composed of interconnected biological neuron models, elementary processes, operating in a coordinated manner and each contributing to producing a result. Indeed, each artificial neuron performs simple local decision-making operations resulting in a score depending on all of these decisions.

ANN is generally composed of an input layer representing the input neurons (input variables), an output layer that consists of the output variables vector that allow to transfer information outside the network and one or a group of hidden layers representing the set of hidden nodes with incoming connections emanating from the input neurons.

The weights on arcs of ANN need to be estimated before the model is used. These weights correspond to the learning process. To overcome classification problems, supervised learning is applied with the desired input-output combinations known. Supervised learning is much faster than the other existing algorithms (unsupervised or reinforced) since the weight adjustment is made directly from the error or the difference between the output obtained by the ANN and the desired output. A popular supervised learning algorithm is the error back propagation.

Support Vector Machine (SVM)

The SVM method is a set of supervised learning techniques aimed at solving discrimination and regression problems. SVM was developed in the 1990s based on the Vapnik-Chervonenkis theory on developing a statistical learning theory. SVM was quickly spread thanks to its ability to process large data. The main aim of SVM is finding the boundary hyperplane parameters by maximizing the distance between the hyperplane and the support vectors from the training data X. An individual i is then assigned to one of the classes of the variable Y depending on the sign of the separating hyperplane equation given by the function f.

SVM is a recent classification alternative. This method relies on the existence of a linear classifier in a suitable space. As a two-class classification method, SVM uses a training data set to learn the model parameters. It is based on the use of a kernel function which allows an optimum separation of data.

The credit scoring models published in the literature are usually based on multivariate statistical methods and ANN. Recently, SVM began to integrate the credit scoring field thanks to works by Huang et al. (2004), Chen & Shih (2006), Huang et al. (2007), Xu et al. (2009), Belotti & Crook (2009), Yu et al. (2010), Hens & Tiwari (2012) and Harris (2013). This is a very efficient method when applied to limited samples. It does not require prior assumptions on the data.

Empirical results

Logistic Regression

Initially, it is interesting to evaluate the logistics (logical) relationship between each independent variable and the dependent variable "Credit". We present in the following table the results of the estimation of logistic coefficients, standard error values of each parameter estimator and the level of significance of each independent variable.

Table 2 : The LR model

Variable Parameter coefficient ($\boldsymbol{V_i}$)	***Parameter estimator*** ($\boldsymbol{\widehat{\beta}_i}$)	***Standard Error*** $\boldsymbol{SE(\widehat{\beta}_i)}$	***Ratio*** $\boldsymbol{\widehat{Z}}$ $\boldsymbol{\frac{\widehat{\beta}_i}{SE(\widehat{\beta}_i)}}$	***P>|z|***
Constant	-16.155762	3.2172	-5.022	0 ***
V2	-5.61248	2.1842	-2.57	0.0102 *
V4	7.221608	2.6034	2.774	0.0055 **
V8	3.326893	1.1149	2.984	0.0028 **
V9	-0.062885	0.0319	-1.969	0.0489 *
V10	-0.510175	0.2782	-1.834	0.0666 .
V11	0.075664	1.0453	0.072	0.9423
V12	-0.123148	0.0462	-2.663	0.0077 **
V16	0.006017	0.0093	0.649	0.5166
V18	0.519098	0.6988	0.743	0.4576
V19	0.620536	0.2268	2.737	0.0062 **
V20	1.708235	0.8767	1.949	0.0514 .
V22	0.810224	0.1884	4.3	0 ***
V23	0.1605	0.3433	0.468	0.6401
V24	1.86096	0.4849	3.838	0.0001 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Moreover, it appears after the parameter estimation by maximum likelihood that 10 variables are actually significant in our model, considering the probability values (column "P> | z |"). These are the variables related to operating profitability (V2), net profitability (V4), solvency (V8), asset coverage (V9), long and medium term debt / fixed assets (V10), repayment capacity (V12), study duration of a credit report (V19), corporate banking relationship duration (V20), size of the company (V22) and ownership structure (V24). These variables are meaningfully involved in learning the logistic function and the highest values of the Wald statistics with the lowest risk levels. The company size (V22) correlates positively and significantly with the credit risk. In our study, all sample loan applicants are small and medium sized businesses. This feature urges banks not to attribute loans without sufficient collateral and adopt a stringent credit policy. This strategy may reduce the credit risk level.

The LR model generates a likelihood ratio statistic $G^2 = 124.91$, critical probability is associated 6.856552e-20 (table 3). The likelihood ratio statistic G^2 is distributed as a χ^2 with degrees of freedom equal to 14. In our case, the G^2 statistic is greater than χ^2 in the table and therefore we reject the hypothesis H_0 . In this case, the $\hat{\beta}$ estimator has a significant effect on the credit variable. The model is globally significant; there is indeed a relationship between the explanatory variables and the dependent variable.

Table 3 : Model overall significance test

$(A): -2LL(0)$	331.71
$(B): -2LL(14\,Variables)$	206.26
G^2: (A) - (B)	124.91
dl	14
$P(> Chi-2)$	6.856552e-20

Artificial Neural Network

To process results obtained by the ANN method, we used the Matlab software R2011.b. This includes a "toolbox neural network" application that allows the modeling of ANN. Different architectures were simulated and optimized by varying the number of neurons in the hidden layer to select the best architecture with a minimum error rate since there is no law, no rule, and no theorem that determines the number of neurons in the hidden layer to get an optimum ANN.

The model implemented is the multilayer perceptron with gradient back-propagation algorithm comprising an input layer, a hidden layer, whose determination of the appropriate number of neurons will be optimized, and an output layer (Figure 1).

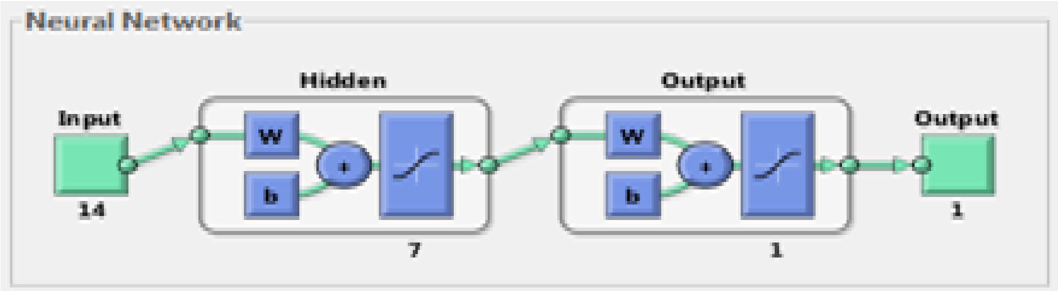

Figure 1: The selected architecture

The optimal network architecture is as follows:
• the input layer comprises 14 neurons representing the 14 explanatory variables;
• the hidden layer includes 7 neurons;
• the output layer includes a one neuron indicating whether the borrower is classified as good or poor.

This architecture allowed us to determine the least mean squared error which stands at 0.0490 (Figure 2).

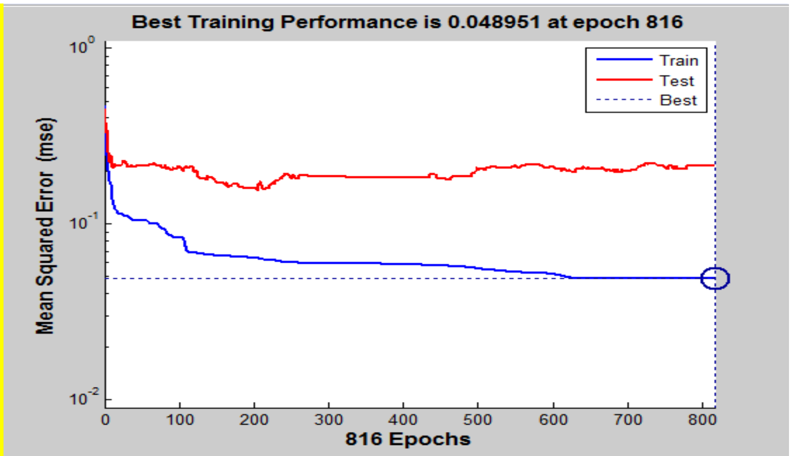

Figure 2: Training and testing curves relative to the optimal network with single hidden and output layers

Support Vector Machine

The quality of a decision model obtained by the SVM method involves selecting several parameters, namely the kernel type and parameters ($\gamma \ldots.$), and the regularization parameter C used to avoid the misclassification error. The research for nonlinear separating surfaces is obtained by the introduction of a kernel function in the scalar product implicitly inducing non-linear transformation of data to a greater intermediate space (feature space). For classification tasks, there are no theoretical guidelines on the type and parameters of the kernel function.
We consider four models based on a linear, polynomial, radial basis function or sigmoid kernel. Parameters to be estimated for these models are given in Table 4:

Table 4 : Kernel functions and parameters

Kernel function	Functional Form	parameters	Default values
Linear	$K(x_i.x_j) = x_i^T.x_j$		
Polynomial	$K(x_i.x_j) = (\gamma x_i.x_j)^d$	$\gamma \in R$ et $d \in N$	$\gamma = 1$ et $d = 3$
Radial basis function	$K(x_i.x_j) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right)$	$\gamma \in R$	$\gamma = 1$
Sigmoid	$K(x_i.x_j) = \tanh(\gamma x_i.x_j)$	$\gamma \in R$	$\gamma = 1$

The choice of appropriate kernel parameters allows improving the SVM classification quality. We seek to change the values of the various kernel parameters, and the parameter C in order to determine the best combination. To do so, we used the "Research Grid" algorithm mentioned by Huang et al. (2007).

We consider the values of Log (C) $\in \{-4, -3, \ldots, 2\}$ and Log (γ) $\in \{-4, -3, \ldots, 2\}$ for the RBF kernel, Log (C) $\in \{-4, -3, \ldots.7\}$ and Log (γ) $\in \{-4, -3, \ldots, 6\}$ for the sigmoid kernel, the degree $d \in \{-4, -3, \ldots, 2\}$ for the polynomial kernel and Log (C) $\in \{-8, -7, \ldots, 4\}$ for the linear Kernel. The implementation of this algorithm is provided by the application tune.svm within the e1071 package in R.

The best performance values according to this approach are: γ = 0.125 and C = 1 for the radial basis function Kernel ; γ = 0.0625, C = 1 for the sigmoid Kernel*;* d = 1, C = 2 and γ = 0.25 for polynomial kernel and C =1 for the linear kernel.

Comparing several performance evaluation metrics

The confusion matrix

Various performance metrics have been proposed in the literature to evaluate the performance of classification models. The confusion matrix is one of the most widely used tools in the field of accounting and finance.

Table 5 : Confusion matrix for credit scoring

		predicted $\hat{Y}$	
		Good payer $\hat{Y}$ *=1*	*Poor payer* $\hat{Y}$ *=0*
actual Y	*Not risky* Y *=1*	True Positive (TP)	False Negative (FN) (error I)
	Risky Y *=0*	False Positive (FP) (error II)	True Negative (TN)

Table 6 illustrates the confusion matrix for the three credit scoring models for training and test samples.

Table 6 : Confusion matrix of different models

Model Description		**sample**	**TP**	**FP**	**TN**	**FN**
LR		Training	202	33	43	8
		Test	83	15	17	7
ANN		Training	206	34	42	4
		Test	84	15	17	6
SVM	Radial basis function kernel	Training	207	32	44	3
		Test	86	14	18	4
	polynomial kernel	Training	207	39	37	3
		Test	83	19	13	7
	Linear kernel	Training	206	37	39	4
		Test	81	19	13	9
	Sigmoid kernel	Training	199	51	25	11
		Test	86	23	9	4

All performance indicators presented in this study were calculated on training and test sample.

ACRN Oxford Journal of Finance and Risk Perspectives
Vol.6 Issue 2, November 2017, p.1-17
ISSN 2305-7394

Table 7 : performance measures according to different models

Model		*Sample*	*Accuracy*	*Sensitivity*	*Specificity*	*Type I Error*	*Type II Error*
LR		Training	85.664%	0.9619	0.5658	0.0381	0.4342
		Test	81.967%	0.9222	0.5313	0.0778	0.4688
ANN		Training	86.713%	0.9810	0.5526	0.0190	0.4474
		Test	82.787%	0.9333	0.5313	0.0667	0.4688
SVM	**Radial basis function *kernel***	**Training**	**87.762%**	**0.9857**	**0.5789**	**0.0143**	**0.4211**
		Test	**85.246%**	**0.9556**	**0.5625**	**0.0444**	**0.4375**
	polynomial *kernel*	Training	85.315%	0.9857	0.4868	0.0143	0.5132
		Test	78.689%	0.9222	0.4063	0.0778	0.5938
	Linear *kernel*	Training	85.664%	0.9810	0.5132	0.0190	0.4868
		Test	77.049%	0.9000	0.4063	0.1000	0.5938
	Sigmoid *kernel*	Training	78.322%	0.9476	0.3289	0.0524	0.6711
		Test	77.869%	0.9556	0.2813	0.0444	0.7188

Table 7 shows the accuracy, sensitivity, specificity and classification error rates for each model. The accuracy measures the fraction of correctly classified borrowers (true positives and true negatives) out of their total number. This indicator is an important criterion in the evaluation of the classification capacity (Abdou & Pointon, 2011) and performance (Paliwal & Kumar, 2009) of the scoring models. Sensitivity is the true positive rate (the fraction of credit-worthy borrowers correctly classified by the classification model), While specificity is the true negative rate (the fraction of non-credit-worthy borrowers correctly classified by the classification model). By comparing the predictive models, we consider type I error (the fraction of borrowers wrongly classified as insolvent) and type II error (the fraction of borrowers wrongly classified as solvent) models. The misclassification costs associated with type II error are significantly higher than those associated with type I error (West, 2000; Lee & Chen, 2005).

According to Table 7, we notice that the radial basis function kernel SVM has the highest accuracy, sensitivity and specificity rates and the lowest type I and II errors. These indicate that the RBF is better and more powerful than the LR and ANN in screening good and poor payers. The ANN model score is slightly more accurate in terms of the predictive power than the LR (86.713% vs. 85.664%) for the training set and (82.787% vs. 81.967%) for the testing set. However, type I error is low for the three credit scoring models, while type II error is high. This high error II arose from the fact that the number of non-creditworthy borrowers is much less than those of creditworthy ones, and thus, models are over learned from creditworthy borrowers. In fact, the three models achieved better results in the classification of creditworthy borrowers (sensitivity) than the classification of non-creditworthy borrowers (specificity) for the training and testing sets.

The Receiver Operating Characteristic (ROC)

The ROC curve is used to estimate the classification system performance. To deal with bank credit, the ROC curve connects the fraction of true positives (sensitivity) to the false positives $(1 - specificity)$, for the same group when the application acceptance score threshold is varied.

To numerically evaluate the ROC curve, we use the area under the ROC curve (AUC: Area Under the Curve) as an evaluation index. AUC measures the model discrimination capacity by showing the probability that a good borrower will have a score higher than the score of a bad borrower. Since credit data are commonly unbalanced, AUC has been suggested as a measure of the discrimination capacity of a classifier regardless of the class distribution or the classification error cost (Baesens et al., 2003).

In order to compare the different methods in accordance with the AUC values, ROC curves are estimated on a test sample. Figure 3 illustrates ROC curves of the three methods studied to compare the predictive power of the different scoring models.

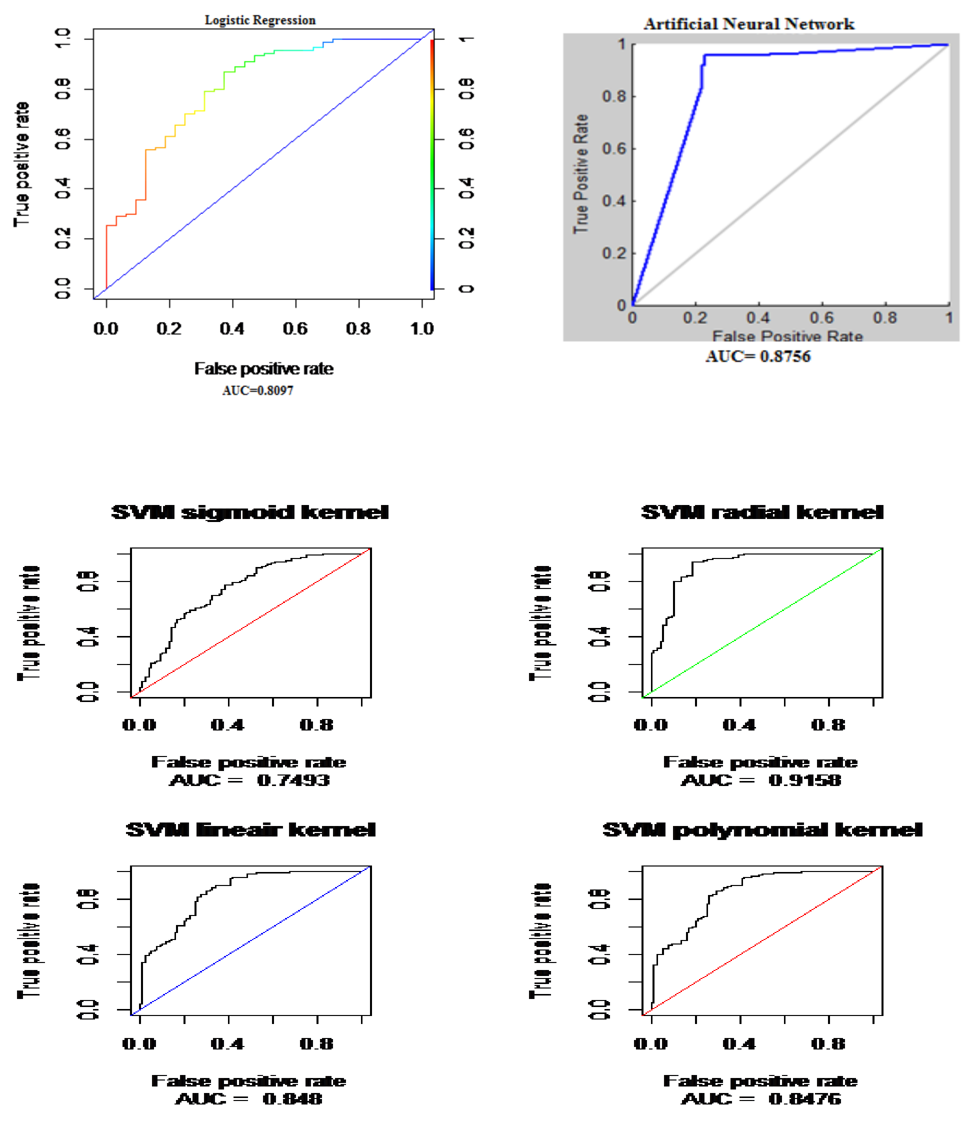

Figure 3: ROC curves of the three models based on LR, ANN and SVM

In practice, the AUC value varies between 0.5 and 1. A score of 1 corresponds to the classifier achieving perfect accuracy; while a score of 0.5 means that the classifier has no discriminative power (Harris, 2013).

We note from Figure 3 that the RBF kernel SVM has the best performance with an AUC equal to 0.9158. This value shows that our model has an excellent discriminating power, i.e. it is 91.58% likely for a positive event to be classified as positive by the test on the range of possible threshold values. The ANN (AUC = 0.8756), linear kernel SVM (AUC = 0.8480) and polynomial kernel SVM (AUC = 0.8476) methods are less efficient than the RBF kernel SVM. Methods that predict classes the worst are those of LR (AUC = 0.8097 and Sigmoid kernel SVM (AUC = 0.7493). We can conclude that these two score models are not discriminating.

Our results have the following implications. They would contribute to the literature by providing a useful method that would serve as a decision making tool. The proposed method provides accuracy and reliability in classification and is believed to have promising practical potential for financial institutions. In fact, credit scoring is one of the main areas of accounting and finance with intelligent technologies applied. With the recognition of the huge technological developments, credit scoring will play an increasingly important role for business loans in the future. Given that a slight improvement in performance and accuracy of forecasting models can have important business implications, reduce evaluation errors and lead to significant future savings, a more precise model can change the relationship between borrowers and lenders. Consequently, our results imply that the application of the SVM method in the credit risk assessment could be adopted to analyze in an optimal way the financial situation of the companies and to strengthen the competitiveness of the financial institutions in the market of the loans.

Conclusions

The objective of this paper was to check the ability of artificial intelligence models to predict the credit risk as regards loan applicants to a Tunisian bank in order to achieve the probability of default of each applicant.

The prediction techniques used in this study are LR, SVM and ANN. The SVM technique looks the most efficient with RBF kernel as compared to other techniques; the correct classification rate and associated AUC are the most significant. The use of artificial intelligence techniques in the evaluation of credit risk improves the effectiveness of credit decision, and minimizes the application processing cost and time as well. The current paper provides insights into the potential of using the RBF model for credit scoring applications in the Tunisian commercial banks. We consider that our model can provide a way to ensure a competitive advantage over other banks that fail to implement such a methodology.

We emphasize the possibility of hybrid models describing the financial and stock aspects of companies following the unexpected exchange of the economic cycle by combining the scoring models to structural models. We currently focus on the development of scoring models by studying the corporate accounting data and compare them to determine the most efficient model.

References

Abdou H., & Pointon J. (2011). Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature. Intelligent Systems in Accounting, Finance and Management, 18 (2/3), 59-88.

Agresti A. (2002). Categorical Data Analysis. New York: John Wiley & Sons, Inc.

Apostolik R, Donohue C., & Went, P. (2009). Foundations of Banking Risk: An Overview of Banking, Banking Risks, and Risk-Based Banking Regulation. New Jersey: John Wiley& Sons, Inc.

Baesens, B., Van Gestel T, Viaene S, Stepanova M, Suykens J, & Vanthienen J. (2003). Benchmarking state-of-the-art Classification Algorithms for Credit Scoring. Journal of the Operational Research Society, 54 (6), 627-635.

Barr, R., & Siems, T. (1994). Predicting Bank Failure Using DEA to Quantify Management Quality. Federal Reserve Bank of Dallas, Financial industry Studies, Working papers N°94-1.

Bekhet, H. A., & Eletter S.F.K. (2014). Credit Risk Assessment Model for Jordanian Commercial Banks: Neural Scoring Approach. Review of Development Finance, 4 (1), 20-28.

Bellotti, T., & Crook J. (2009). Support Vector Machines for Credit Scoring and Discovery of Significant Features, Expert System with Applications, 36 (2), 3302-3308.

Bolton, C. (2009). Logistic Regression and Its Application in Credit Scoring, Disertation, Universiteit of Pretoria.

Burger, M., & Hofinger, A. (2005). Regularized Greedy Algorithms for Network Training with Data Noise. Computing, 2-30.

Chen, W.H., & Shih JY. (2006). A study of Taiwan's Issuer Credit Rating Systems Using Support Vector Machines, Expert Systems with Applications, 30 (3), 427-435.

Demirgüç-Kunt A., & Detragiache E. (1998). Financial Liberalization and Financial Fragility. IMF Working Paper 98/83, (Washington: International Monetary Fund).

Fukunaga K. (1972). Introduction to statistical pattern recognition. New York: Academic Press, 369 p.

Hai, L., Shi, B., & Peng, G. 2013. A Credit Risk Evaluation Index System Establishment of Petty Loans for Farmers Based on Correlation Analysis and Significant Discriminant. Journal of Software, 8 (9), 2344-2351.

Han, J. & Kamber, M. (2001). Data Mining: Concepts and Techniques, San Francisco, CA, USA: Morgan Kaufmann.

Harris, T. (2013). Quantitative Credit Risk Assessment Using Support Vector Machines: Broad Versus Narrow Default Definitions. Expert Systems with Applications, 40 (11), 4404-4413.

Henley, W. E., & Hand, D. J. (1996). A k-nearest Neighbour Classifier for Assessing Consumer Risk. Statistician, 45 (1), 77-95.

Hens, A.B., & Tiwari, M.K. (2012). Computational Time Reduction for Credit Scoring: An Integrated Approach Based on Support Vector Machine and Stratified Sampling Method. Expert Systems with Applications, 39 (8), 6774-6781.

Huang, C.L., Chen, M.C., & Wang, C.J. (2007). Credit Scoring with a Data Mining Approach Based on Support Vector Machines. Expert System with applications, 33 (4), 847-856.

Huang, Z., Chen, H., Hsu, C.J., Chen, W.H., & Wu, S. (2004). Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study. Decision Support Systems, 37 (4), 543-558.

Hung, C., & Chen, J.H. (2009). A Selective Ensemble Based on Expected Probabilities for Bankruptcy Prediction. Expert Systems with Applications, 36 (3), 5297-5303.

Karaa, A., & Krichène A. (2012). Credit–Risk Assessment Using Support Vectors Machine and Multilayer Neural Network Models: A Comparative Study Case of a Tunisian Bank. Journal of Accounting and Management Information Systems, 11 (4), 587-620.

Khashman, A. (2011). Credit Risk Evaluation Using Neural Networks: Emotional Versus Conventional Models. Applied Soft Computing, 11 (8), 5477-5484.

Kruppaa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer Credit Risk: Individual Probability Estimates Using Machine Learning. Expert Systems with Applications, 40 (13), 5125-5131.

Lee, T., & Chen, I. (2005). A Two-Stage Hybrid Credit Scoring Model Using Artificial Neural Networks and Multivariate Adaptive Regression Splines. Expert Systems with Applications, 28 (4), 743-752.

Li, S.T., Shiueb, W., Huang, M.H. (2006). The Evaluation of Consumer Loans using Support Vector Machines. Expert Systems with Applications, 30 (4), 772-782.

Lin, S.W., Shiue, Y.R, Chen, S.C., & Cheng, H.M. 2009. Applying Enhanced Data Mining Approaches in Predicting Bank Performance: A Case of Taiwanese Commercial Banks. Expert Systems with Applications, 36 (6), 11543-11551.

Liu, Y., Zhang, H.H., & Wu, Y. (2011). Soft or Hard Classification? Large Margin Unified Machines. Journal of the American Statistical Association, 106 (493), 166-177.

Malley, J.D., Kruppa, J., Dasgupta, A., Malley, K.G., & Ziegler A. (2012). Probability Machines: Consistent Probability Estimation Using Non Parametric Learning Machines. Methods of Information in Medicine, 51 (1), 74-81.

Nnamdi, I.N., & Shola, G.O. (2011). A Comparison of Different Soft Computing Models for Credit Scoring. International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering, 5 (6), 883-888.

Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid System with Genetic Algorithm and Artificial Neural Networks and Its Application to Retail Credit Risk Assessment. Expert Systems with Applications, 39 (16), 12605-12617.

Paliwal, M., & Kumar, U.A. (2009). Neural networks and statistical techniques: A review of applications. Expert Systems with Applications, 36 (1), 2-17.

Salehi, M., & Mansoury A. (2011). An Evaluation of Iranian Banking System Credit Risk: Neural Network and Logistic Regression Approach. International Journal of the Physical Sciences, 6, 6082-6090.

Tsai, C., & Wu, J. (2008). Using Neural Network Ensembles for Bankruptcy Prediction and Credit Scoring. Expert System with Applications, 34 (4), 2639-2640.

Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two Credit Scoring Models Based on Dual Strategy Ensemble Trees. Knowledge-Based Systems, 26, 61-68. DOI: 10.1016/j.knosys.2011.06.020

Wang, Y., Wang, S., Lai K.K. (2005). A New Fuzzy Support Vector Machine to Evaluate Credit Risk. IEEE Transactions on Fuzzy Systems, 13 (6), 820-831.

West, D. (2000). Neural Network Credit Scoring Models. Computers and Operations Research, 27 (11/12), 1131-1152.

Wiginton, J.C. (1980). A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior. The Journal of Finance and Quantitative Analysis 15 (3), 757-770.

Wu, Y., & Liu, Y. (2007). Robust Truncated-Hinge-Loss Support Vector Machines. Journal of the American Statistical Association, 102 (479), 974-983.

Xu, X., Zhou, C., & Wang, Z. (2009). Credit Scoring Algorithm Based on Link Analysis Ranking with Support Vector Machine. Expert Systems with Applications, 36 (2), 2625-2632.

Yeh, I.C., & Lien, C.H. (2009). The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients. Expert Systems with Applications, 36 (2), 2473-2480.

Yu, L., Wang, S., & Lai, K.K. (2008). Credit Risk Assessment with a Multistage Neural Network Ensemble Learning Approach. Expert Systems with Applications, 34 (2), 1434-1444.

Yu, L., Yue, W., Wang, S., & Lai, K.K. (2010). Support Vector Machine Based Multiagent Ensemble Learning for Credit Risk Evaluation, Expert Systems with Applications 37 (2), 1351-1360.