International Journal of Islamic and Middle Eastern Finance and Management

The applicability of credit scoring models in emerging economies: An evidence from Jordan

Journal:	*International Journal of Islamic and Middle Eastern Finance and Management*
Manuscript ID	IMEFM-02-2017-0048.R5
Manuscript Type:	Research Paper
Keywords:	credit scoring techniques, Classification techniques, Jordanian banks, emerging market countries

The applicability of credit-scoring models in emerging economies: Evidence from Jordan

Abstract

Purpose: The main aim of this paper is to propose an objective and efficient method for assessing credit risk by introducing and investigating to a greater extent the applicability of credit-scoring models in the Jordanian banks and to what range they can be utilized to achieve their strategic and business objectives.

Design/methodology/approach: The research methodology comprises two phases. The first phase is the model development. Three modelling techniques are the utilized to build the scoring models namely logistic regression (LR), neural networks (NN) and support vector machines (SVM) and the best performing model will be selected for next stage. The second phase is two-fold: 1) linking the credit expert knowledge in a way that can enhance the outcomes of the scoring model, 2) a profitability test to explore if the selected model is efficient in meeting banks strategic and business objectives.

Findings: The findings showed that LR model outperformed both ANN and SVM across various performance indicators. LR model also fit best with achieving the bank's strategic and business objectives.

Originality/value: To the best of our knowledge, this study is the first that applied several modelling and classification techniques for Jordanian banks and calibrating the best model in terms of its strategic and business objectives. Furthermore, credit experts' knowledge was engaged with the scoring model to determine its efficiency and reliability against the sole use of an automated scoring model in the hope to encourage the application of credit scoring models as advisory tool for credit decisions.

Keywords: Credit scoring techniques, Classification techniques, Jordanian banks, Emerging market countries.

Paper type: Research paper.

1. Introduction

Loan granting is considered one of the main income sources for banks and financial institutions. Therefore, due to rapid growth of consumer credit and large amounts of financial data, careful assessment should be taken when deciding to grant credit to potential customers. However, emerging financial economies still utilize subjective, judgemental approaches, and have inefficient policies and guidelines for evaluating credit.

Credit-scoring has been a challenging and widely discussed topic in literature since the inspirational work of Altman (1968). Many credit-scoring models using statistical and artificial intelligence (AI) techniques have been proposed by researchers and financial risk modelers in industry (Thomas, 2000). Consequently, the Basel Committee on Banking Supervision (2000) required all financial institutions to have rigorous and complex credit-scoring systems in order to help them improve their credit risk levels and capital allocation.

Most of developed economies (e.g. US and UK) adopted credit-scoring systems as the main source of decision to admit and extend credit. However, these systems are not used in the emerging financial economies, where subjective and judgmental analyses are more commonly utilized. Nevertheless, several related studies, which will be discussed later on, show how there is a movement towards developing automated scoring models for banking systems in such countries. Given these new developments, the motivation of this paper is to investigate the effectiveness of three well-known classification methods, namely, LR, ANN and SVM in the Jordanian retail banking.

The Jordanian banking sector is considered as one of the fundamental pillars of the Jordanian economy. It provides several financial credit facilities to customers and various industry sectors, while handling different types of risk. Retail banking systems in Jordan, however, commonly use subjective methods to evaluate loans (Bekhet and Eletter, 2014). Additionally, they use the opinion of credit officers and other factors (e.g. reputation and financial capabilities of loan applicants). The process of loan assessment in Jordanian retail banking usually begins with an interview with the loan applicant, then gathering of the essential information needed to decide whether he/she is eligible for a loan. Subsequently, there is processing of the application form by the credit analyst to decide, whether or not the loan will be granted. These forms of loan evaluation are not efficient as they are subject to human error and bias along with being time-consuming (Handzic et al., 2003). It is, therefore, necessary to investigate the application of a credit-scoring modelling based on several classification techniques and to choose the best model that will satisfy the bank's strategic objectives. To the best of our knowledge, this is the first study that will apply several classification techniques to Jordanian retail banking and calibrate the best model in terms of the business objectives of the banks.

Hence, this research will address three research questions:

1. Is there a need for credit-scoring models in the Jordanian banking system?
2. Are these models applicable in the Jordanian banking system and what is their role in achieving the banks' strategic and business objectives in terms of profitability?
3. What is the role of credit expertise and how can this enhance the credit-scoring model decision?

This paper is organized as follows: section 2 highlight the related work, section 3 gives an overview of the Jordanian banking sector, section 4 describes the methodology adopted in the study, section 5 demonstrates results of the proposed models, section 6 calibrates the best model to meet banking objectives and finally, section 7 draws conclusions as well as providing guidelines for future research directions.

2. Related work

Researchers have been considering developing credit-scoring methods for decades. A lot of complex credit-scoring models were developed using classifiers categorized as traditional and advanced techniques, for example LR in Baesens et al. (2003), West (2000) applied several ANNs, whereas, Bellotti and Crook (2009) applied SVM. As discussed above, credit-scoring systems are widely used in developed financial industries, where they became a main source of decision of granting and extending credits. However, in emerging countries, there are several studies which show the trend in developing scoring models for their banking systems.

Abdou and Pointon (2009) carried out an investigation on the existing methods in credit decision making within the Egyptian banking sector and proposed using credit-scoring techniques to examine to what extent decision making can be improved. Both studies considered the use of LR and ANN, their results are shown to provide more efficient classification results than the judgmental techniques. Furthermore, ANN give better accuracy rates than LR.

Bazmara and Donighi (2014) used fuzzy rules expert systems to evaluate bank customer credit in one bank in Iran. They benchmarked their results among others such as LR and ANN showing better performance. Dinh and Kleimeier (2007) proposed a multi-purpose credit-scoring model for the Vietnamese retail banking market using LR. Firstly, they show how to identify borrower characteristics that should be part of a credit-scoring model. Then they focused on achieving the strategic objectives of the bank. Finally, they assess the use of credit-scoring models in the context of transactional versus relationship lending. In Triki (2016) they compared LR and ANN against a sample collected from a Tunisian bank, their results reveal that LR outperformed ANN. Sharma et al. (2012) examined the challenges and complexities relating to credit-scoring model development within the Nepalese banking sector. The developed model was deployed using LR.

Akkoc (2012) proposed a model based on statistical techniques and Neuro Fuzzy on credit card data in Turkey and compared its results to LR and ANN. Bekhet and Eletter (2014) and Ala'raj et al (2015) explored the use of credit-scoring techniques in Jordanian banking sector. In their studies LR, ANN and SVM were investigated but what distinguishes our study from theirs is that we are considering extra benefits of the developed credit-scoring model rather than just raw scores.

3. Jordanian banking sector: an overview

This section provides the status of present banking and credit risk analysis characteristics in Jordan. According to Zeitun and Benjelloun (2013) the banking industry in Jordan is considered to be one of the most important sectors.

Figure 1 illustrates the significant role of the Jordanian banking sector in financing the major economic activities in the country, which in return reflects on the growth and the prosperity of the economy. World Bank data indicate that approximately 50% of businesses surveyed receive financing from banks in Jordan. Figure 1 provides examples of industries supported by banks in Jordan.

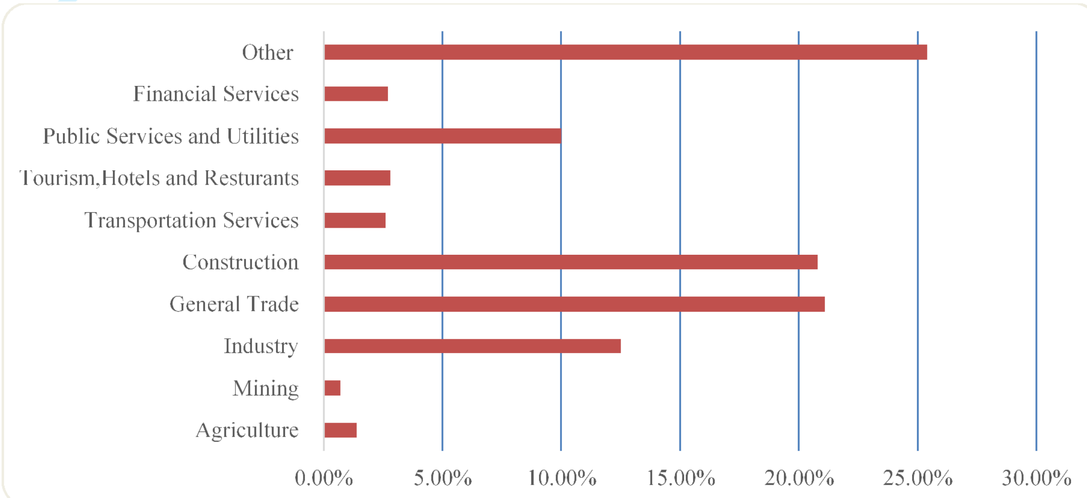

Figure 1. Distribution of credit facilities according to economic sectors (2003– 2016) (CBJ)

The Association of Banks in Jordan (ABJ, 2015) stated in a recent report the number of banks operating in Jordan reached 25 banks at the end of 2015. The services of these banks cover most parts of Jordan through a network of branches that consist of 786 branches. These statistics clearly demonstrate the rapid growth of the banking industry in Jordan, which results in the growth of the credit environment.

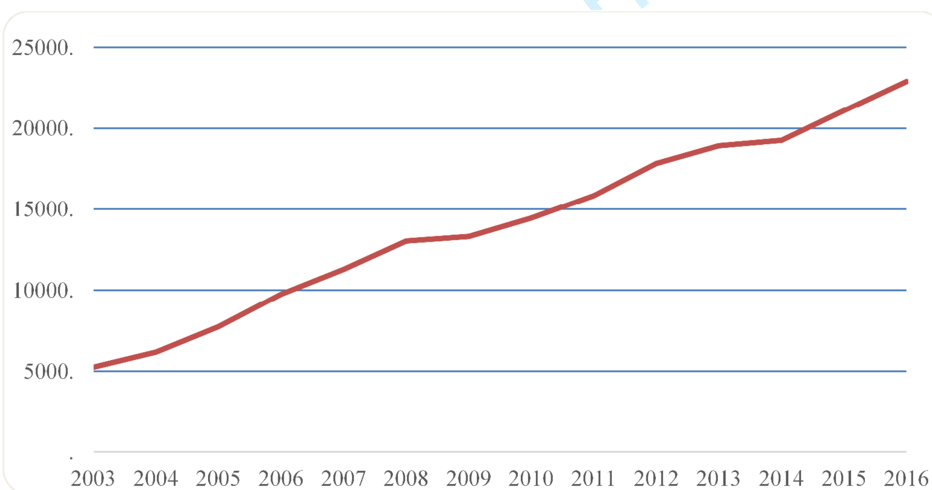

Figure 2. Total credit facilities by licensed banks in Jordan (CBJ)

Figure 2 demonstrates the growth of total credit facilities extended by banks operating in Jordan to individuals, companies and governments which, in turn, contributes to the national economy. Credit facilities by banks extended from 5.3 billion Jordanian Dinars (JDs) in 2003 to 22.9 billion JDs in 2016. This 17.6 billion JDs increase represents 335% growth with a 14% annual growth rate indicating that economy is developing and banks are generating profits but this should be backed with strict credit assessment to avoid any discarded events to occur. According to (ABJ, 2013,p.42) total credit facilities as a percentage of GDP ranged between 73-93% during the period of 2003-2012, reflecting the vital role of these institutions in financing and supporting all sectors of the Jordanian economy.

Retail banking in Jordan is growing. Figure 3 below provides an overview of overdrafts, discounted bills, loans and advances distributed by year from 2003 through 2016. Figure 3 indicates that the volume of overdrafts gradually increased annually, loans and advances grew at a rate of 450%, and discounted bills were reduced by 28.4%. Clearly, Jordan is experiencing rapid growth and changes within the banking sector. These changes indicate that a more effective method of analyzing credit risk will be needed in order to maintain positive economic growth and to avoid negative consequences of irresponsible lending.

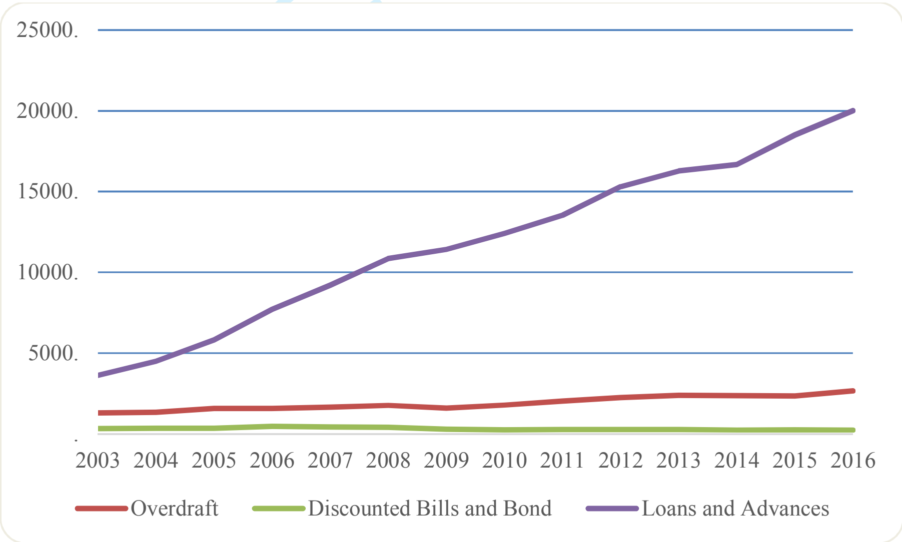

Figure 3. Credit Facilities According to Type (CBJ)

Figure 4 indicates that an increase in lending of 10 billion JDs in 2003 to approximately 33 billion JDs in 2016 occurred. This increase equates to 230% and averages to an annual growth rate of 13%. Jordanian banks participate in market competition by providing numerous financial services to individuals and companies, which causes them to need to manage different forms of risk (ABJ, 2015), thus contributing to additional need for implementing more effective credit risk analyses.

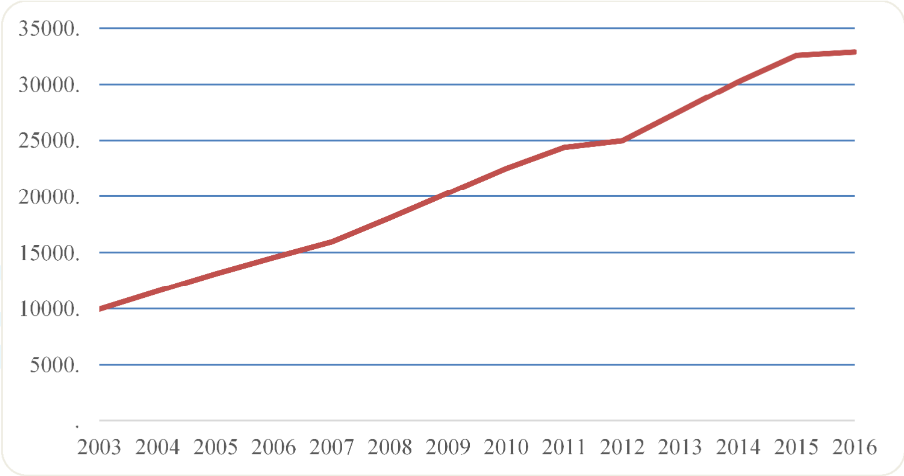

Figure 4. Total Deposits by licensed banks in Jordan (CBJ)

The World Bank (2014) reported that approximately 5.6% of businesses applying for credit were rejected; however, the percentage of non-performing loans in Jordan remains high. According to ABJ (2015) that the percentage of non-performing loans to total loans reached 8.4% in the middle of 2012, then it regressed to 7% in 2014 before landing to around 5%. Despite the fact non-performing loans rate is decreasing year after year history can still repeat itself. According to the Jordan Economics report (2015) the ratio of non-performing loans reached 7% at the end of June 2014. Regardless of the decrease, ratio of default loans is still higher than the international average for default rates which is lower than 5% (Al-Shawabkah and Tambyrajah, 2009), this indicates that there are problems remain in assessing risk in Jordan.

Regardless of the risk associated with issuing credit, the nature of loan evaluation and assessment in Jordanian banks remains subjective and is based on the analyst's experience and some minor guidelines set by the bank (Handzic et al., 2003), resulting in an approach that is inefficient and inconsistent. Accordingly, lenders should search for more computerized ways of completing credit evaluation and decision. The goal of this paper is to provide evidence to encourage Jordanian banks to consider implementing an effective and effcient computerized credit-scoring systems.

4. Research methodology

This section introduces and describes three classification techniques in the field of credit-scoring, namely, LR, ANN, and SVM. LR is the industry standard for building credit-scoring models due to its flexibility and the fact that it has a binary outcome (Akkoc, 2012). ANN and SVM are advanced tools for creating credit-scoring models that also allow for non-linear relationships among variables; whereas LR is restricted by the assumption of linearity (Huang et al., 2007).

The methodology framework and how the classification techniques will be used is illustrated in Figure 5. Looking at the figure the workflow is divided in to 2 stages, in the first stage is the model development which comprises the aforementioned techniques. The best performing classifier will qualify to the next stage. Stage 2 covers 2 approaches that involve expert judgments in a way that can enhance outcomes of the model and then a profitability test is carried out to explore if credit-scoring models are efficient approaches to be considered to be applied within Jordanian banks. This stage will be discussed in the coming sections. Experimental analyses were performed using Matlab 2013b software.

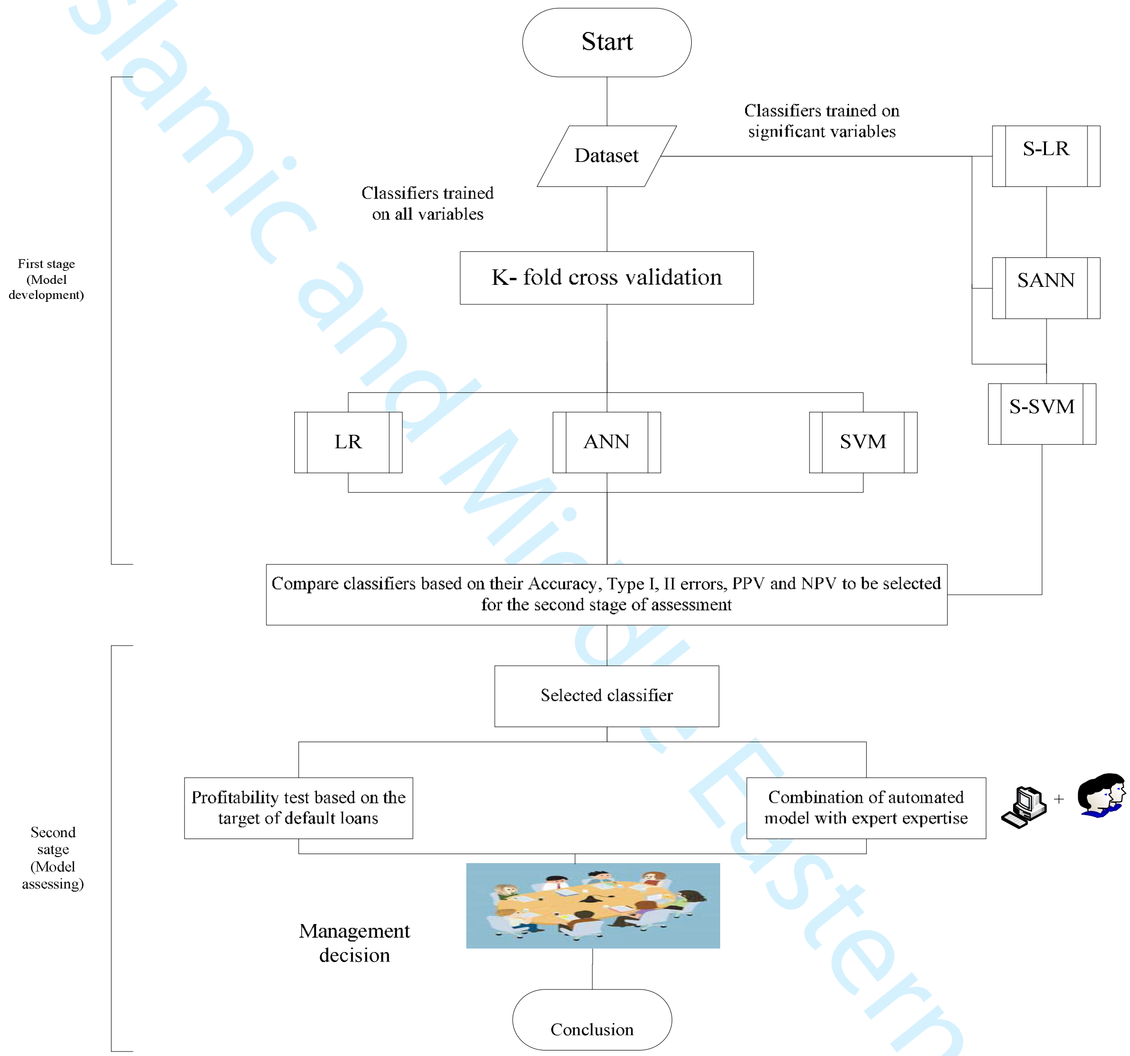

Figure 5. Research workflow

4.1. LR

In credit-scoring, the classification is a binary problem in which the decision is characterized by 0 (grant/good loan) or 1 (reject/ bad loan) (Thomas, 2000). Therefore, LR was developed to address this issue by reducing the output to either 0 or 1. LR studies the relationship between several independent variables and the probability of a loan being granted by fitting them to a logistic curve (Sweet & Martin, 1999). In this context, LR is expressed as:

$$log[p(1-p)] = \beta^0 + \beta 1X\,1 + \beta 2X2 + \cdots + \beta nXn$$

Where p is the probability of outcome of interest (0/1), β_o is the intercept term, and βi represent the coefficient related to the independent variables *Xi_(i=1…n), and log[p(1-p)]* is the dependent variable which is the logarithm of ratio of two probabilities outcome of interest. The objective of LR in credit-scoring is to determine the conditional probability of a specific input (customer's characteristics) belonging to a certain class. Despite the wide application of LR, its accuracy decreases when the relationship between variables is non-linear (Akkoc, 2012). ANN and SVM were introduced to solve the problem of linearity (Huang et al., 2007).

4.2. ANN

ANN is an AI technique that was originally inspired by how the human brain processes information. ANN mimics this process by allowing for concurrent complex processing of inputs in order to achieve an output (Bhattacharyya and Maulik, 2013).

One of the most common architectures for ANNs is the multi-layer perceptron (MLP), which consists of one input layer, one or more hidden layers and one output layer. One of the key issues needing to be addressed in building ANNs is their topology, structure and learning algorithm Angelini et al. (2008). The most commonly utilized topology of ANN model in credit-scoring is the three-layer feed-forward back propagation.

As shown in Figure 6. The ANN model is made up of three layers: input, hidden and output layers. The "nodes" represent neurons, much like the human brain. For credit-scoring, the structure of the ANN model is to enter the attributes of each applicant in the input layer to process them, then they are transferred to the hidden layer, which contains an algorithm for processing all attributes based on weights, then the values are sent to the output layer that provides the final answer, which is to give or not to give a loan. The output is calculated by using weights. In summary ANN tend to establish the relationship between a customer's probability of default and their given characteristics. To have more insight about the pros, cons and various technicalities of ANN please refer to (West, 2000).

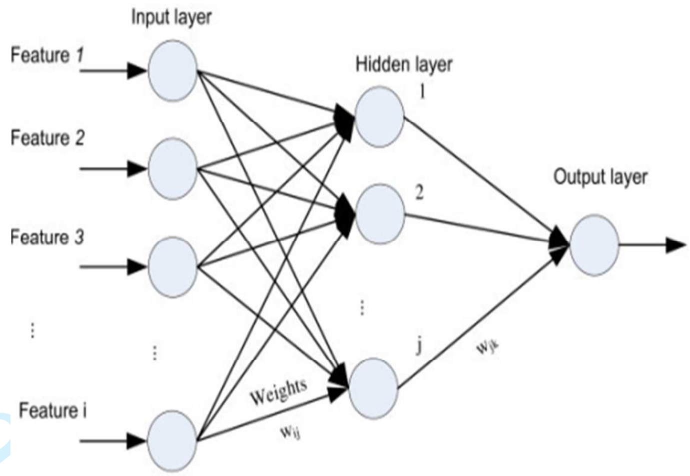

Figure 6. ANN topology

4.3. SVM

SVMs are a machine learning technique used in classification and credit-scoring problems (Huang et al., 2007). The SVM was first proposed by Cortes and Vapnik (1995) and is used to find an optimal hyperplane that categorizes the training input data in two classes (good/ bad loans). Figure 7 shows the hyperplane characteristics and the SVM.

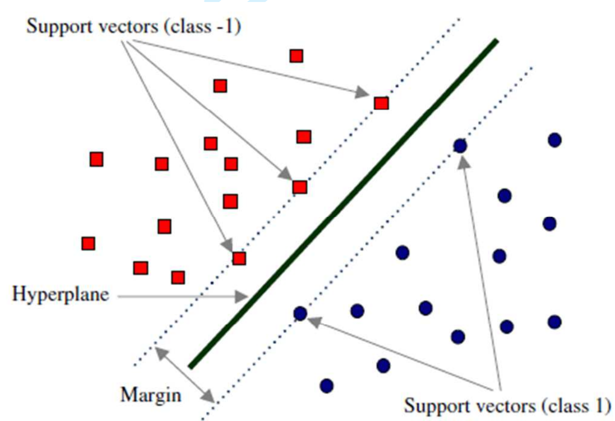

Figure 7. The basics of SVMs

As it can be seen in Figure 7, the dashed lines parallel to the optimal hyperplane measure the distance to the solid line. The dashed lines are called margins, and the training data that lies on the margins is called support vectors, however, the SVM attempts to find the best optimal hyperplane that separates the data correctly, so the margin width between the optimal hyperplane and the support vectors are maximized to fit the data. Based on the features of the support vectors, this form of statistical modelling can be used to predict whether an applicant has good or bad credit. To have more insight about the pros, cons and technicality of SVMs please see (Cortes and Vapnik, 1995).

4.4. Data collection

This study presents analysis based on a historical loan dataset gathered from one public commercial bank in Jordan. These data are confidential and sensitive; hence, acquiring it was detailed and time-consuming process. The dataset consists of 500 loans, of which 400 are good loans and 100 are bad loans. Bad loan cases were more difficult to obtain due to manual storage at the banking institution. Therefore, bad loan data also include current cases as opposed to historical cases. These are loans that are currently 90 consecutive days past due, which is considered to be in default status in Jordanian banking policy. It is clear that the dataset is biased towards good loans due to the low default rates occurred at that time in the bank. Each loan in the dataset is made up of 12 independent variables with various types and the dependent variable, which indicates the status of the loan as either good or bad. A summary of the dataset and variables are described in Table 1.

No.	Description	Definition	Type	Code
X_1	Age	Applicant's Age	Continuous	AGE
X_2	Gender	Male/ Female	Categorical	GEN
X_3	Marital Status	Married/ Single	Categorical	MAR STA
X_4	Job	Type of Job	Nominal	JOB
X_5	Monthly Income	Applicant's monthly salary	Continuous	MON INC
X_6	Years at Job	Years at present job	Continuous	YAJ
X_7	Loan Purpose	Personal/Auto/Housing loan	Nominal	LOANP
X_8	Loan Amount	Loan amount in JD	Continuous	LOANA
X_9	Loan Duration	Duration of the loan in months	Continuous	LOAND
X_{10}	Loans from other Banks	Does applicant have loans from other banks	Categorical	LFOB
X_{11}	Liabilities	Obligations to his bank	Categorical	LIAB
X_{12}	Property	Applicant owns a house or not	Categorical	PROP
Y	Class	0 (Good applicant)/ 1(Bad applicant)	Continuous	LOANS

Table 1. Description of the Jordanian dataset attributes

According to Abdou and Pointon (2011) in the credit-scoring process, there are certain and specific criteria that are looked at. The main determinants of whether a default will take place or not can be classified into the following four areas:

1. Financial Indicators (MON INC, LIAB).
2. Demographic Indicators (AGE, GEN, MAR STA, PROP).
3. Employment Indicators (YAJ, JOB).
4. Behavioural Indicators (LFOB).

Therefore, regarding our collected data, the majority of the features fall in the above categories.

4.5. Data splitting

Accoridng to Garcia et al. (2015) data splitting is a fundamental step in model evaluation, where the data is partitioned into a training set for building and learning of the model and a testing set for evaluating and assessment of the model. Both a large training and test set produces more accurate performance estimates. Data splitting is essential in credit-scoring modelling because the data are usually limited and these partitioning strategies have great impact on consistent model evaluation. As data is limited, it is necessary to ensure adequate balance of training set and testing set size.

Several splitting techniques are available. This paper presents data that adopted the K–fold cross validation technique. This technique involves dividing the datasets into K subsets (or folds) of equal size (K = 1, 2…, K) but K cannot exceed the size of the dataset. Therefore, the model training is based K-1 folds, and the left K folds will be saved for model evaluation or testing. The process continues until all K folds are used for evaluation. All the tested K fold predictions are used to estimate the model accuracy (by taking the average).

4.6. Data normalization

Each variable in the dataset comprises values that vary in range. Data were standardized to avoid bias. Standardization was accomplished by normalizing variable values to the range of 0 to 1 This transformation is done by taking the maximum value within each attribute and dividing all the values in the attributes with its maximum value (Alaraj et al., 2014).

4.7. Performance measures indicator

To measure the accuracy and discriminate ability of a developed model, many performance evaluation measures can be used. Accuracy, error rate, Type I and II errors and area under the curve (AUC) (Abdou and Pointon, 2011) are commonly used as performance measure indicators in the field credit-scoring.

The most applied evaluation measurements in credit-scoring literature are accuracy, Type I and Type II errors (Nanni and Lumini, 2009). A combination of these measurements is used, in order to measure the performance of the proposed credit-scoring models. Accuracy measures the percentage of the correctly classified loans (both good/ bad):

- *Accuracy = TP+TN / TP+FN+TN+FP*

While the Type I and II error evaluations measure the percentages of the misclassified bad loans and good loans respectively, which are not evaluated by the accuracy measure.

- *Type I error = FP/ TN+FP (Bad as Good, lose money)*
- *Type II error = FN/ TP+FN (Good as Bad, lose potential income)*

Type I error is related to financial loss, and Type II is related to the opportunity of profit loss. In the financial point of view, risks associated with Type I error are more costly than Type II (West, 2000), in other words it can be seen as a 'Financial vs. profit loss'.TP represents good loans that are classified as good, TN represents bad loans that are classified as bad, FN (Type II) represents a good loan misclassified as bad loan and FP (Type I) represents a bad loan misclassified as a good loan. All these measures can be derived from a 2x2 confusion matrix (see Table 2).

Some additional measures can be included in the modelling. These items which are limited in the literature, such as positive prediction values (PPV) and negative prediction value (NPV) measures. PPVs are the percentage of how many loans were predicted by the model as good loans, in other words, it also shows how many good loans were correctly predicted by the model. NPVs are the percentages of how many loans were predicted by the model as a bad loan, indicating how many bad loans were correctly predicted by the model. PPV and NPV can also help in calibrating the model to meet its strategic objectives on the percentages of FP and FN.

- $PPV = TP/\ TP+FP$
- $NPV = TN/\ TN+TN$

	Predicted (%)		
Actual	**Good loans**	**Bad loans**	**Accuracy (%)**
Good applicant	TP	FN (Type II error)	Acc_{good}
Bad applicant	FP (Type I error)	TN	Acc_{bad}
	PPV	NPV	Acc_{total}

Table 2. *Confusion Matrix for Credit-scoring*

5. Results and discussion

The 5-fold cross validation sampling technique was used for training and testing of the developed models for the dataset. To decrease the effect of the variability of the training set and to achieve reliable results, each trial was repeated ten times; and, hereafter the test results are the average of 50 trials. This section will demonstrate the development and results of several credit-scoring models using LR, ANN, and SVM.

5.1. LR

Table 3 summarizes the classification results of the LR credit-scoring model. The average total accuracy of the model is 88% with 93.75% accuracy for good loans and 65% accuracy for bad loans. The Type I and II error rates were 35% and 6.25% respectively. 91.46% of the data was predicted as good loans (PPV) and 72.2% of data were predicted as bad loans (NPV).

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	75	5	0.9375
Bad loans	7	13	0.65
	0.9146	0.722	0.88

Table 3. *LR classification results*

5.2. ANN

Hidden nodes need to be accurately determined in the ANN model to avoid model over fit. With ANN, it is also important to choose the optimal model or topology. To achieve this, a trial and error process can be carried out using various ranges of parameters until choosing the topology with the best accuracy. First a range of 20 to 40 neurons in the hidden layer was tested. The ideal topology was selected based on the lowest Mean Square Error, and the highest accuracy, mainly for bad loans. The activation function used was 'logistic sigmoid'. The Results of the ANN model is described in Table 4, the model achieved a total accuracy of 85% with ability to classify good loans by 92.53% and 55% for correctly classifying bad loans. The Type I error was worse than Type II error with 45% and 7.5% respectively. 89.15% was predicted by the model as good loans (PPV) and 64.7% as bad loans (NPV).

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	74	6	0.9250
Bad loans	9	11	0.55
	0.8915	0.647	0.85

Table 4. *ANN classification result*

5.3. SVM

Selecting the appropriate kernel function is important for the model determination of SVM. The proposed SVM credit-scoring model was developed using a linear kernel. Table 5 shows the SVM classification results where the model recorded 86% accuracy, on classifying good loans, of 92.5% and 60% accuracy on classifying bad loans. Regarding Type I and II error rates, misclassification occurred of good and bad loans by 40% and 7.5%, respectively. 90.24% of the data was assigned as good loans (PPV), 66.66% of the data to be bad loans (NPV).

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	74	6	0.9250
Bad loans	8	12	0.60
	0.9024	0.666	0.86

Table 5. *SVM classification results*

5.4. Models results summary

When comparing all models, LR achieved the best accuracy results. LR attained an accuracy of 88% while ANN and SVM scored 85% and 86% respectively. LR's superior performance lies in its ability in recognizing good and bad loans where it achieved lower Type I and II error rates. For the PPV and NPV rates, LR scored higher than ANN and SVM, which also indicates the predictive power of its classifying good and bad loans in the LR credit-scoring model. Regarding ANN and SVM, SVM was better in classifying bad loans as it did produce lower a Type I error than ANN, while in the Type II error, both models performed the same. The SVM model PPV and NPV rates were superior to ANN. However, in conclusion, the LR credit-scoring model achieved the best model classification and predictive ability, followed by the SVM, and finally, the ANN credit-scoring models.

All variables were assessed in these analyses; however, in many research fields such as pattern recognition, it may be more accurate to select a group set of representative variables with more predictive information (Tsai, 2009). Variables that do not fit the model well or are redundant may affect the performance of the model. Additionally, decreasing the number of irrelevant or redundant variables significantly reduces the running time of the models and gives more generalizable ability (Guyon and Elisseeff, 2003). For this reason, three additional models were assessed. These models incorporated the inclusion of variables in a step-by-step fashion, providing the ability to determine model fit and accuracy with each variable addition. This form of modeling includes initiating the model with no variables, then entering the variables one by one to test the improvement of model performance with the addition of each variable. To determine the most effective model, an additive process is used until the addition of a variable does not improve the model fit.

5.5. Simplified LR (SLR)

Table 6 reports the results of the simplified LR credit-scoring model with selected features. Model accuracy did not improve with the addition of an eighth variable; therefore, seven variables out of twelve showed significance and were selected for building the model. These included AGE, MAT STA, JOB, YAT, LOANP, LOANA and LFOB. The accuracy result was 89% with good loan classification accuracy of 93.75% and bad loans classification accuracy of 70%. Type I and II error rates attained 30% and 6.25% respectively. The predictive power of good loans reached 92.6% and for bad loans it reached 73.7%.

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	75	5	0.9375
Bad loans	6	14	0.70
	0.926	0.737	0.89

Table 6. S*LR classification results*

5.6. Simplified ANN (SANN)

Five variables were selected for the final model using ANN. These variables included AGE, MAR STA, JOB, LOANP, LFOB. Table 7 demonstrates that accuracy results achieved 88% with 90% accuracy of good loan classification and 60% accuracy of bad loan classification. Misclassifications of good and bad loans scored 5% and 40% respectively. 90.5% of good loans predicted by the model were true while 75% of the bad loans predicted were true.

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	76	4	0.95
Bad loans	8	12	0.60
	0.905	0.75	0.88

Table 7. S*ANN classification results*

5.7. Simplified SVM (SSVM)

Table 8 summarizes the performance of the simplified SVM credit-scoring model with seven variables selected into the model, which included AGE, GEN, MAR STA, JOB LOANP, LFOB. Results are 89% for total accuracy, where 95% was of good loans classification accuracy and 65% classification accuracy for bad loans. The Type I and II error rates show 35% and 5% loans misclassification respectively. Finally, from all loans predicted as good (PPV), 91.6% were true good loans and from all loans predicted as bad loans (NPV) 76.5% were true bad loans.

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	76	4	0.95
Bad loans	7	13	0.65
	0.916	0.765	0.89

Table 8. S*SVM classification results*

5.8. Simplified Models results summary

When using the step forward fashion to assess variable accuracy, the SLR model is superior to SANN and SSVM. SLR and SSVM models both scored 89% but the SLR ability in classifying bad loans was better than the SSVM model. However, misclassifying a bad loan is more expensive and the resulting cost is higher than misclassifying a good loan (West, 2000; Abdou, 2009). SANN performance was the lowest in both cases, whether using selected variables or all of them.

Results of the three credit-scoring models with selected variables achieved classification results better than that of the models that were initially created which included all variables. This indicated that some variables affected the model performance, and it is best to leave those variables out of the model. In comparison to the models (LR, NN and SVM) that used all variables. The SLR credit-scoring model outperformed the LR model by 1%; The SANN model was significantly better than the ANN model by 3%. Finally, the SSVM model also was superior to SVM model by 3%.

Table 9 presents all model results for comparison. Accuracy in classifying bad loans, low Type I error rate and high PPV are all considered the essential indicators for banks when assessing credit worthiness of loan applicants. The best model among all the developed models that can be chosen as the best credit-scoring model to be applied in Jordanian retail banking is the SLR model.

	Acc_{total}	Acc_{good}	Acc_{bad}	Type I error	Type II error	PPV	NPV
LR	0.88	0.9375	0.65	0.35	0.0625	0.9146	0.722
SLR	**0.89**	0.9375	**0.7**	**0.3**	0.0625	**0.926**	0.737
ANN	0.85	0.925	0.55	0.45	0.075	0.8195	0.647
SANN	0.88	0.95	0.6	0.4	0.05	0.905	0.75
SVM	0.86	0.925	0.6	0.4	0.075	0.9024	0.666
SSVM	0.89	0.95	0.65	0.35	0.05	0.916	0.765

Table 9. *Models results comparison*

According to Garcia et al (2015) it is not applicable to verify that one model achieves results better than another, because of the different performance measures or splitting techniques used. For a complete performance evaluation, it would seem proper to implement some hypothesis testing to emphasize that the experimental differences in performance are statistically significant, and not just due to random splitting effects. Choosing the right test for specific experiments depends on factors such as the number of datasets and the number of classifiers to be compared.

	SLR	SSVM	LR	SANN	SVM	ANN
SLR	1	0.082	0.046	0.027	0.008	0.003
SSVM	0.082	1	0.039	0.031	0.015	0.009
LR	0.046	0.039	1	0.281	0.026	0.019
SANN	0.027	0.031	0.281	1	0.042	0.037
SVM	0.008	0.015	0.026	0.042	1	0.452
ANN	0.003	0.009	0.019	0.037	0.452	1

Table 10. *p-values for pairwise model comparsion using McNemar method*

For this purpose, a McNemar statistical test (West, 2000) is conducted to determine whether SLR performance was not a matter of random process or due to data splitting effect. As we can see from Table 10, SLR is statistically better than other models except SSVM at 5% significance level, whereas it is better than SSVM at 10% significance level. The large values reflected in Table 10 indicate that the corresponding row and classifiers perform at the same level and there is little difference in the prediction vectors of these classifiers.

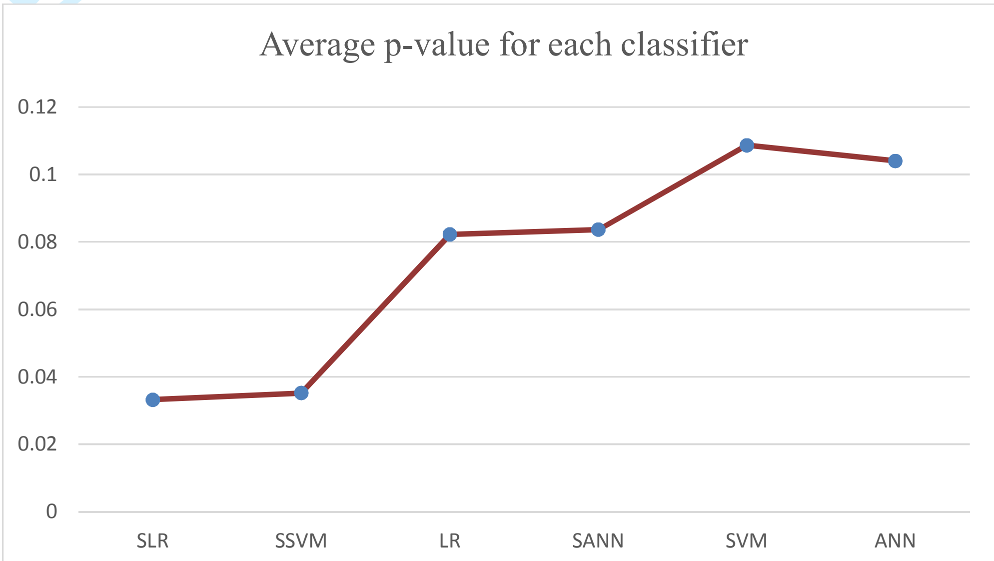

Figure 8. Average p-value for all classifiers

Figure 8 is constructed as an average of Table 10 for each row (excluding the 1's on the diagonal). This is seen in Figure 8 and when taking Table 8 and 9 into consideration, we can see that the lowest average p-value is the SLR classifier. This is a strong sign that the prediction vector of SLR is statistically closer to the vector of actual values than the prediction vectors of other classifiers.

6. Banks' business and strategic objectives

Implementing a model that can capture a potential bad loan is of a great importance to banking institutions as it allows these institutions to avoid financial loss by granting loans that are not paid back. However, at the same time, it is important to accurately assess potential good loan applications due to higher generation of income. When managing a loan portfolio it is to also good to have a low rate of default loans. This rate varies from bank to another.

There is a threshold that must be achieved when assigning good or bad categories. Basically, when loan applicant characteristics are above the threshold, they are rated as good, and when below, they are rated as bad. The common choice of threshold values in credit-scoring literature is 0.5; All previous models in this study used 0.5 to distinguish between good and bad loans. According to Bellotti and Crook (2009) the choice of the threshold value depends on the earlier assumption of the misclassification costs of good and bad loans. For example if a banks' focus is on decreasing the misclassification of bad loans, the bank should tend to reduce the threshold in order to have more bad loans classified correctly. However, misclassifying good loans will exist as well. This is called a trade-off between bad and good loans. Therefore, as the cost of misclassifying bad loans is associated with more losses than misclassifying a good loan, we aim to calibrate the chosen model (SLR) by determining an optimal cut-off score that meet the banks business and strategic objectives.

6.1. Model calibration

In general, banks in Jordan don't use credit-scoring systems in their loan evaluation processes, but the credit policies and guidelines they follow state that there should be a specific rate for the non-performing or defaulted loans within their portfolio. However, technically setting an optimal cut-off score that can lead to better bad loans classification is associated with banking behaviour. Based on the credit expert's knowledge from the bank we collected the data from, the target rate of non-performing loans was roughly 1.6% (It means that from all good loans accepted by the model 1.6% of good loans are predicted to be the limit for misclassification), in other words the desired PPV = 100% - 1.6% = 98.4% .

If we consider this target default rate of 1.6% as a cut-off score, it means that any loan that falls below 1.6% is considered good; hence we obtained the results below:

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	46	34	0.575
Bad loans	0	20	0.1
	0.1	0.3703	0.66

Table 10. SLR classification results with 1.6% cutt-off score.

Table 10 summarizes the results of the SLR credit-scoring model with a cut-off score of 1.6%. The results reveal that using the 1.6% cut-off which is also the target rate of non-performing loans has no bad loans, thus resulting in the PPV leading to 100%. Conversely, the rate of correctly classified good loans is 57.5% with a total accuracy of 66%. Therefore despite having 0% misclassified bad loans (Type I error), there is 42.5% misclassified good loans (Type II error). Instead, this cut-off can be calibrated via trade-off between good and bad loans in order to increase the rate of correctly accepted good loans and sacrifice with a bad loan, while still achieving the bank's target of having 1.6% of non-performing loans. As a result the best cut-off that leads to the anticipated results is 17.21%. These results are presented in Table 11.

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	64	16	0.80
Bad loans	1	19	0.95
	0.9846	0.5428	0.83

Table 11. SLR classification results with 17.21% cutt-off score

Table 12, presents the results of SLR credit-scoring model with a cut-off score of 17.21%. Although, the PPV rate is lower than the PPV rate with 1.6% cut-off score, the target rate of non-performing loans is better because there are more good loans classified correctly (22.5% increment of 18 good loans). The results of good and bad loans classified is 80% and 95%, respectively, with 83% total classification accuracy. The Type I error rate achieved is 5% and Type II error rate is of 20%. The NPV rate is higher than the NPV rate of 1.6% cut-off by 17.25%. A superior balance between good and bad loan misclassification is achieved with the cut-off of 17.25% because there is a higher rate of correctly classified good loans, thus increasing income.

6.2. Model profitability

According to Schreiner (2003), banks can estimate the effects of a scoring model in terms of profitability even before it is applied. According to the bank's presumed target on the non-performing loans, a cut-off score of 17.21% reached this target. To check the reliability of the proposed model on the banks' strategic objectives, we aimed to test the effectiveness of the model on the bank's profitability. To estimate the effects of the proposed credit-scoring model on profit, Schreiner (2003) proposed to measure the profitability a model can make by considering the cost of losing a good loan to avoid misclassifying a bad loan. The net profit of a credit-scoring model can be calculated as:

(The financial cost of each bad loan* TN) – (The opportunity cost of each good loan * FN)

Where TN is the number of bad loans avoided or correctly classified, and FN is the number of lost or misclassified good loans. Figure 8 illustrates the number of good loans lost against each bad loan avoided over several cut-off scores.

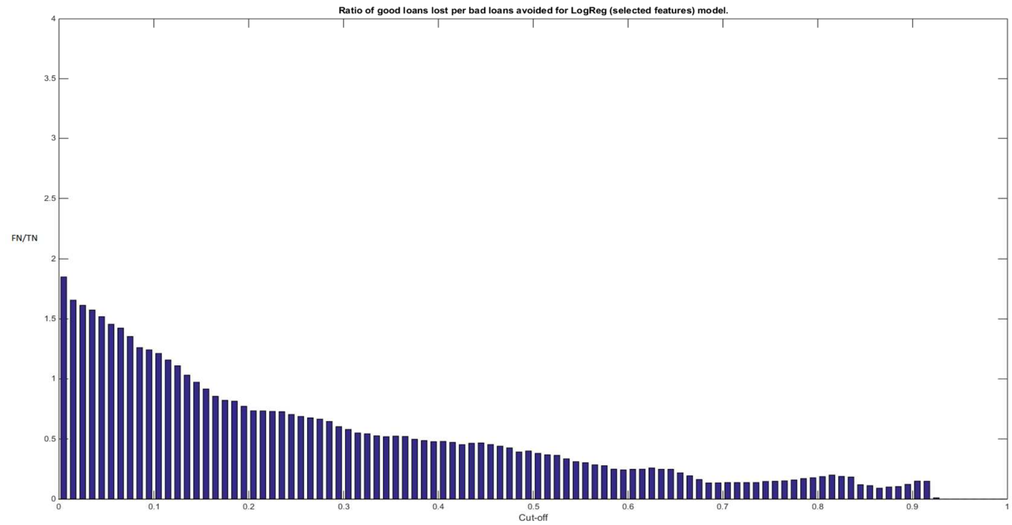

Figure 8. The ratio of FN loans per TN avoided

The figure displays the trade-off between good loans and bad loans, the cut-off score of 17.21% shows that 0.8% of good loans need to be sacrificed to save each bad loan, while at the cut-off of 1.6% the loss is around 1.6% good loans to avoid misclassification of a bad loan.

To estimate the effects of this model on bank profits, prior knowledge about the costs of loans should be available because the costs are associated with future loan applicants. Intuitively, banks believe that the costs related to bad loans are higher than that of losing good loans. However, it is a complicated task to deliver consistent estimates of the misclassification costs associated with loans, consequently valid expectations might not be available mainly in the environment of Jordanian banking. In the case of our bank, no such information or assumptions is available. Schreiner (2003) stated that banks in rich and developed countries assume that it takes 10 good loans to pay off the losses of one bad loan. Moreover, in the credit-scoring literature, the widely public German credit dataset showed a ratio of misclassification costs of 5:1 (West, 2000). Here we will consider different ratios as a benchmark to measure the profit in our bank over different cut-off scores. Four different scenarios will be carried out using ratios of 10:1, 10:5, 10:10; 10:12 and 5:1. These rations are tested to measure the effect on the profit change of the model. Figure. 7 illustrate the results.

As indicated in Figure 9, all ratios lead to an increase of profits at the 17.21% cut-off score, while at the cut-off score of 1.6% one ratio makes a profit loss and three ratios make profits, but these profits are less than the profits at the 17.21% cut-off score. These results indicate that the bank's target rate of non-performing loans of 1.6% is well identified. At the ratio 10:1, the cut-off score 17.21% achieved the highest increase in profit among other ratios across all cut-off points. One more finding is that all four ratios achieve profits at cut-off values around 15% or higher.

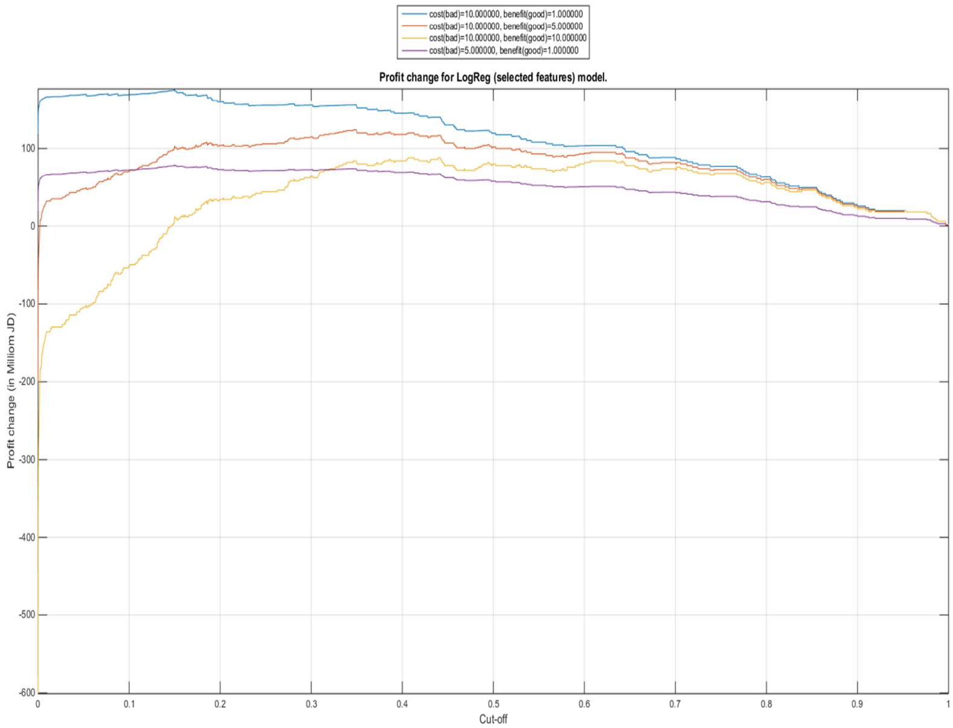

Figure 9. The profit change of SLR model over different ratios

6.3. Quantitative and qualitative lending approaches

Based on the previous experimental findings, and the effects of the proposed credit-scoring model, the question that arises is to what extent these scoring systems can be applied or adopted by the banking system in Jordan? As discussed earlier in this paper, the lending policies of Jordanian banks are subjective based on analyses of conducted by credit officers and analysts. These analysts provide expertise in addition to minimal guidelines established by management. Moreover, Jordanian banks lend to customers based on collaterals and relationships (Al-Shawabkah and Tambyrajah, 2009). Credit-scoring systems give a single quantitative measure based on a determined cut-off score which determines the customer's eligibility for a loan or not. In practice, a scoring model could accept a customer who is known to the bank of his bad history and could reject a customer who is known to the bank for his good reputation and financial capabilities. Anderson (2007) states that judgmental techniques are still used with lending decisions, based on little or unstructured data and experience.

Thus, this portion of the paper presents analyses that consider that there can be a qualitative risk that the quantitative scoring models do not discover or capture. Additionally, another question can arise for those customers who sit at the border of the cut-off score, either good or bad, and to what extent they are eligible for a loan or if they need another chance or further assessment. A bit of human assessment may be needed in situations where a cut-off based on a quantitative evaluation is not enough to make an appropriate determination.

Mester (1997) and Schreiner (2003) provided some examples of countries' experience on using scoring systems, for example Mester referred to a bank that uses scoring systems up to certain amounts and Schreiner provided an example of a bank that filed bankruptcy because it solely trusted the scoring systems. These examples raise the concern about relying totally on a credit-scoring system while limiting the expertise of the credit officer. In such cases, credit officers can identify and discover risks that cannot be directly or indirectly recognised by an automated scoring system. These analysts and credit experts provide expertise in certain situations based on their experience, knowledge and relationship with the customer.

To overcome such issues, and to take advantage of the expertise and the knowledge of credit officers and analysts, a combination of a quantitative and qualitative credit-scoring system is proposed to determine its efficiency and reliability against the sole use of an automated credit-scoring model. A Model of 2 cut-off points is assessed and presented in this section of the paper, where 2 automatic cut-offs are predetermined. These cut-off points have customers who scored less than the first cut-off and are automatically accepted and where customers greater than the second cut-off point are automatically rejected. In this scenario, the customer predictions which fall between the 2 cut-off points will be further examined and evaluated by the credit officer or analysts.

According to Jensen (1992), the determination of the cut-off points is considered a complex task because the predictions of good and bad customers' loans frequently overlap. One strategy to overcome this complexity is to assign these cut-off points as stated in the Credit Risk: Estimation Techniques report (www.crisil.com), where the two cut-off scores are selected, such that it minimizes the cost of evaluation ($0<\text{Cut-off}_1<\text{Cut-off}_2<1$). Usually, the first cut-off and the second cut-off points are selected as follows:

- The maximum rate of the misclassification of bad loans (bad loan → good loan) should not be more than 5%.
- At least 75% of bad loans should be correctly classified by the model (bad loan → bad loan), this stress that bad loans are associated with its high costs of misclassifications that have occurred

The following are the rules for predicting the default:

- If the model prediction is < Cut-off_1 then loan is accepted.
- If the model predictions is > Cut-off_2 then the loan is rejected.
- If the model predictions are between the 2 cut-offs the loans are further evaluated and examined further ($\text{Cut-off}_1<\text{Predictions}<\text{Cut-off}_2$).

Based on the above discussion, the results of the model are demonstrated in Table 12.

Actual class	Predicted class		
	Good loans	**(0.1522<loans to Review<0.3961)**	**Bad loans**
Good loans	62	11	8
Bad loans	1	4	15
Misclassified bad loans	5		
Correctly classified bad loans			75

Table 12. SLR model with 2 cut-off scores

The model was trained and tested on cut-off points of 0.1522 and 0.3961. The model correctly classified 62 good loans and misclassified 8 good loans. While it classified 15 bad loans correctly and missed only 1 bad loan. The loans that needed further review and assessment were 11 good loans and 4 bad loans. Let's assume that that the credit officer accepted all the 11 good and rejected all the 4 bad loans, the total accepted good loans will be 73 and the rejected bad loans will be 19. However, if the credit analyst rejected the good loans and accepted the bad loans, the total will be 19 misclassified good loans and 5 misclassified bad loans.

In comparison with the SLR model with cut-off scores of 0.5 and 0.1721, the accepted good loans were 75 and 64 loans respectively, and the rejected bad loans were 14 and 19 loans respectively. It can, thus, be concluded that the SLR credit-scoring model of 0.5 cut-off score recognizes good loans better than the combined scoring model, while the combined model was superior to recognise bad loans if the credit officer evaluated the loans precisely. For the SLR credit-scoring model of 0.1721 cut-off score, the combined model performed better assuming that all reviewed loans were correctly evaluated. In the case that the credit officer falsely evaluated the loans, the combined model will be worse than SLR with 0.5 and 0.1721 cut-off points respectively.

In summary, due to common lending nature in Jordanian banks the integration of the credit analysts' knowledge and expertise in credit-scoring models shows that it could enhance the model outcome, but it depends on the efficiency and precision of the credit analyst in evaluating the loans that need further assessment.

7. Conclusion

This paper presented a set of automated credit-scoring models to choose among the best fit model that can stand as an effective and supportive decision tool in the process of credit evaluation. The credit-scoring models were designed using well-known classification techniques, namely LR, ANN and SVM.

In the light of the study's results and discussion, adopting credit-scoring models in Jordanian banking sector could be a great and profitable tool and hand for decision making process. The existence of a consistent credit-scoring model helps in reducing processing and analysis costs and efforts, allow faster decisions, guarantying credit collectively and reduce any exposure to risks, hence having greater impact on credit and portfolio risk management policies and practices.

One limitation in this study is the use of a relatively small dataset. Finally, there is no overall optimal model because the choice of the classification technique depends on the nature of the problem, dataset size, and economic conditions of the market. Nevertheless, the bank's management will need to decide on the best model to be applied. This paper can help provide guidance in the transformation to automated systems in future. Ultimately, future research could be extended by:

1) Collecting a larger dataset to validate the selected model.

2) Testing different classification techniques such as Decision Trees and Naïve Bayes.

3) Developing behavioural scoring systems which focus on the payment behaviour of existing loan customers.

References

Abdou, HA., and Pointon, J. (2011), "Credit-scoring, statistical techniques and evaluation criteria: A review of the literature", Intelligent Systems in Accounting, Finance and Management, Vol.18 No.2-3, pp.59-88.

Abdou, HA., Pointon, J. (2009), "Credit-scoring and decision making in Egyptian public sector banks", *International journal of managerial finance*, Vol 5 No.4, pp.391-406.

Akkoc, S. (2012), "An empirical comparison of conventional techniques, neural networks and the three-stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit-scoring analysis: The case of Turkish credit card data", *European Journal of Operational Research,* Vol.222 No.2, pp. 168-178.

Al-Shawabkah, A. and Tambyrajah, A. (2009), "Impact of knowledge management on credit risk management in Jordanian banks", *In Proceedings of 10th European Conference on Knowledge Management*. Vicenza, Italy.

Alaraj, M., Abbod, M., and Hunaiti, Z. "(2014) Evaluating Consumer Loans Using Neural Networks Ensembles", *Int. Conf. on Machine Learning, Electrical and Mechanical Engineering (ICMLEME'2014)*, Dubai, UAE pp.55-61.

Alaraj, M., Abbod, M., and Al-Hnaity, B. (2015), "Evaluation of Consumer Credit in Jordanian Banks: A Credit-scoring Approach", *Proceedings of the 2015 17th UKSIM-AMSS International Conference on Modelling and Simulation. IEEE Computer Society, Cambridge, UK, pp.127-130.*

Altman, EI. (1968), "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", *The journal of finance,* Vol.23 No.4, pp.589-509.

Anderson, R. (2007), *The credit-scoring toolkit*. Oxford University Press.

Angelini, E., di Tollo, G., and Roli, A. (2008). "A neural network approach for credit risk evaluation", *The Quarterly Review of Economics and Finance*, Vol. 48 No.4, pp.733-755.

Association of Banks in Jordan (2013), "Development of Banking Sector in Jordan", available at: http://www.abj.org.jo/enus/developmentofthejordanianbankingsector.aspx (accessed 20 Nov 2016).

Association of Banks in Jordan (2015), "Development of Banking Sector in Jordan", available at: http://www.abj.org.jo/enus/developmentofthejordanianbankingsector.aspx (accessed 20 Nov 2016).

Baesens, B. Van Gestel, T., Viaene, S., M. Stepanova, M., Suykens, J. and Vanthienen, J., "Benchmarking state-of-the-art classification algorithms for credit scoring," *J. Oper. Res. Soc*., Vol. 54 No.6, pp. 627-635.

Bank Audi (2015)," Between a large international support and adverse regional turmoil spillover effects", *Jordan Economic Report*, available at: http://www.bankaudigroup.com/GroupWebsite/openAudiFile.aspx?id=2552 (accessed 20 Nov 2016).

Basel committee on banking supervision (2000), "Credit Ratings and Complementary Sources of Credit Quality Information. Bank for International Settlements. Basel Committee on Banking Supervision", Working Paper 3, available at: http://www.bis.org/publ/bcbs72a.pdf (accessed 23 Nov 2016).

Bazmara, A. and Donighi, SS (2014), "Bank Customer Credit-scoring by Using Fuzzy Expert System", *International Journal of Intelligent Systems and Applications (IJISA),* Vol.6 No.11, pp.29-35.

Bekhet, HA and Eletter, SFK. (2014), "Credit risk assessment model for Jordanian commercial banks: Neural scoring approach", *Review of Development Finance*, Vol.4 No.1, pp.20-28.

Bellotti, T. and Crook, J. (2009), "Support vector machines for credit-scoring and discovery of significant features", *Expert Systems with Applications,* Vol.36 No.2, pp.3302-3308.

Bhattacharyya, S., and Maulik, U. (2013), *Soft computing for image and multimedia data processing*. Springer.

Central Bank of Jordan (2013), *Publications and Statistic*. Available at: http://statisticaldb.cbj.gov.jo/index?lang=en (accessed 22 Nov 2016).

Cortes, C. and Vapnik, V (1995), "Support-vector networks", *Machine Learning*, Vol.20 No.3, pp.273-297.

"Credit Risk: Estimation techniques report", available at: http://www.crisil.com/global-research-analytics/gra-publications.html (accessed 25 Nov 2016).

Dinh, THT and Kleimeier, S. (2007), "A credit-scoring model for Vietnam's retail banking market", *International Review of Financial Analysis,* Vol.16 No.5, pp.471-495.

García, V., Marqués, AI. and Sánchez, JS. (2015), "An insight into the experimental design for credit risk and corporate bankruptcy prediction systems", *Journal of Intelligent Information Systems,* Vol.44 No.1, pp.159-189.

Guyon, I. and Elisseeff, A. (2003), "An introduction to variable and feature selection", *The Journal of Machine Learning Research*, Vol.3, pp.1157-1182.

Handzic, M., Tjandrawibawa, F. and Yeo, J. (2003) How neural networks can help loan officers to make better informed application decisions. *Informing Science,* Vol. 6, pp. 97-109.

Huang, C., Chen, M. and Wang, C. (2007), "Credit-scoring with a data mining approach based on support vector machines", *Expert Systems with Applications,* Vol.33 No.4, pp.847-856.

Jensen, HL. (1992), "Using neural networks for credit-scoring", *Managerial Finance,* Vol.18 No.6, pp.15-26.

Mester, LJ. (1997), "What's the point of credit-scoring?" *Business review*, Vol. 3, pp.3-16.

Nanni, L. and Lumini, A. (2009), "An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit-scoring", *Expert Systems with Applications,* Vol.36 No.2, pp.3028-3033.

Schreiner, M. (2003), "Scoring: the next breakthrough in microcredit", Consultative group to assist the poorest (CGAP).

Sharma, S., Harvey, J. and Robson, A. (2012), "Automated credit decision process–an insight into developing a credit-scoring model within the Nepalese banking sector", *International Journal of Decision Sciences, Risk and Management,* Vol.4 No. (3-4), pp. 233-253.

Sweet, S. A. and Grace-Martin, K. (1999), *Data analysis with SPSS*, Allyn & Bacon USA.

Thomas, LC. (2000), "Survey of credit and behavioral scoring: forecasting financial risk of lending to consumers", *International Journal of Forecasting.*Vol.16 No.2, pp.149-172.

Triki, I. (2016), "Credit-scoring Models for a Tunisian Microfinance Institution: Comparison between Artificial Neural Network and Logistic Regression", *Review of Economics & Finance*, Vol. 6, pp. 61-78.

Tsai, C. F. (2009), "Feature selection in bankruptcy prediction", *Knowledge-Based Systems*, Vol.22 No.2, pp.120-127.

West, D. (2000), "Neural network credit-scoring models", *Computers & Operations Research,* Vol. 27 No.11, pp.1131-1152.

World Bank, European Bank for Reconstruction and Development, European Investment (2014), "Jordan Enterprise Survey 2013", available at: http://microdata.worldbank.org/index.php/catalog/2065/study-description (accessed 27 Nov 2016).

Zeitun, R. and Benjelloun, H. (2013), "The efficiency of banks and the financial crisis in a developing economy: the case of Jordan. Journal of Finance", *Accounting and Management,* Vol.4 No.1, pp.1-20.

No.	Description	Definition	Type	Code
X_1	Age	Applicant's Age	Continuous	AGE
X_2	Gender	Male/ Female	Categorical	GEN
X_3	Marital Status	Married/ Single	Categorical	MAR STA
X_4	Job	Type of Job	Nominal	JOB
X_5	Monthly Income	Applicant's monthly salary	Continuous	MON INC
X_6	Years at Job	Years at present job	Continuous	YAJ
X_7	Loan Purpose	Personal/Auto/Housing loan	Nominal	LOANP
X_8	Loan Amount	Loan amount in JD	Continuous	LOANA
X_9	Loan Duration	Duration of the loan in months	Continuous	LOAND
X_{10}	Loans from other Banks	Does applicant have loans from other banks	Categorical	LFOB
X_{11}	Liabilities	Obligations to his bank	Categorical	LIAB
X_{12}	Property	Applicant owns a house or not	Categorical	PROP
Y	Class	0 (Good applicant)/ 1(Bad applicant)	Continuous	LOANS

Table 1. Description of the Jordanian dataset attributes

	Predicted (%)		
Actual	**Good loans**	**Bad loans**	**Accuracy (%)**
Good applicant	TP	FN (Type II error)	Acc_{good}
Bad applicant	FP (Type I error)	TN	Acc_{bad}
	PPV	NPV	Acc_{total}

Table 2. Confusion Matrix for Credit-scoring

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	75	5	0.9375
Bad loans	7	13	0.65
	0.9146	0.722	0.88

Table 3. LR classification results

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	74	6	0.9250
Bad loans	9	11	0.55
	0.8915	0.647	0.85

Table 4. ANN classification result

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	74	6	0.9250
Bad loans	8	12	0.60
	0.9024	0.666	0.86

Table 5. SVM classification results

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	75	5	0.9375
Bad loans	6	14	0.70
	0.926	0.737	0.89

Table 6. SLR classification results

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	76	4	0.95
Bad loans	8	12	0.60
	0.905	0.75	0.88

Table 7. SANN classification results

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	76	4	0.95
Bad loans	7	13	0.65
	0.916	0.765	0.89

Table 8. SSVM classification results

	Acc_{total}	Acc_{good}	Acc_{bad}	Type I error	Type II error	PPV	NPV
LR	0.88	0.9375	0.65	0.35	0.0625	0.9146	0.722
SLR	**0.89**	0.9375	**0.7**	**0.3**	0.0625	**0.926**	0.737
ANN	0.85	0.925	0.55	0.45	0.075	0.8195	0.647
SANN	0.88	0.95	0.6	0.4	0.05	0.905	0.75
SVM	0.86	0.925	0.6	0.4	0.075	0.9024	0.666
SSVM	0.89	0.95	0.65	0.35	0.05	0.916	0.765

Table 9. Models results comparison

	SLR	SSVM	LR	SANN	SVM	ANN
SLR	1	0.082	0.046	0.027	0.008	0.003
SSVM	0.082	1	0.039	0.031	0.015	0.009
LR	0.046	0.039	1	0.281	0.026	0.019
SANN	0.027	0.031	0.281	1	0.042	0.037
SVM	0.008	0.015	0.026	0.042	1	0.452
ANN	0.003	0.009	0.019	0.037	0.452	1

Table 10. p-values for pairwise model comparison using McNemar method

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	46	34	0.575
Bad loans	0	20	0.1
	0.1	0.3703	0.66

Table 10. SLR classification results with 1.6% cutt-off score.

	Predicted class		
Actual class	**Good loans**	**Bad loans**	**Accuracy**
Good loans	64	16	0.80
Bad loans	1	19	0.95
	0.9846	0.5428	0.83

Table 11. SLR classification results with 17.21% cutt-off score

Actual class	Predicted class		
	Good loans	**(0.1522<loans to Review<0.3961)**	**Bad loans**
Good loans	62	11	8
Bad loans	1	4	15
Misclassified bad loans	5		
Correctly classified bad loans			75

Table 12. SLR credit-scoring model with 2 cut-off scores

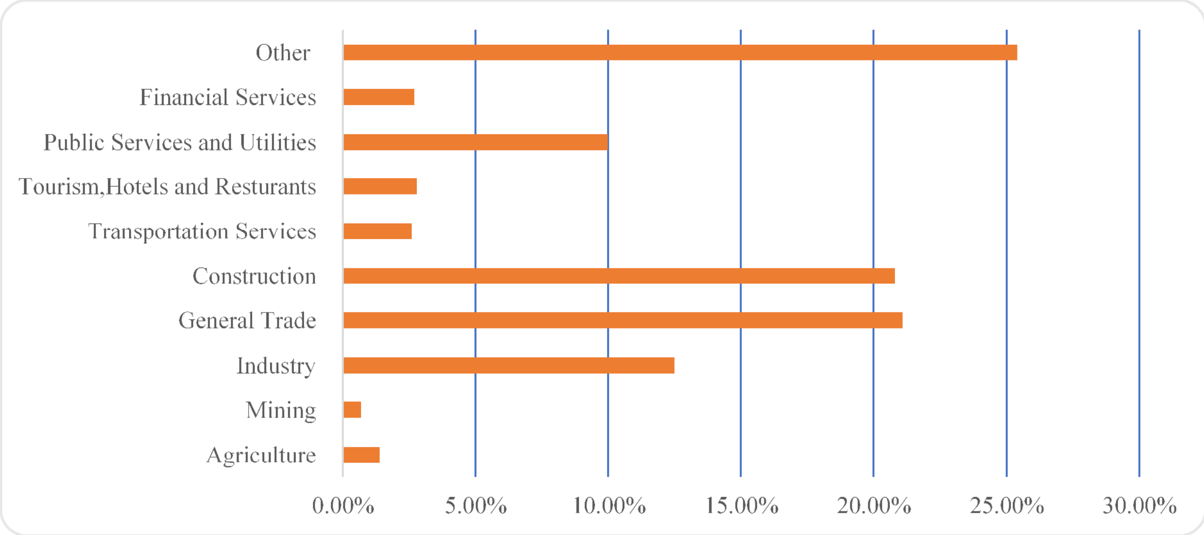

Figure 1. Distribution of credit facilities according to economic sectors (2003– 2016) (CBJ)

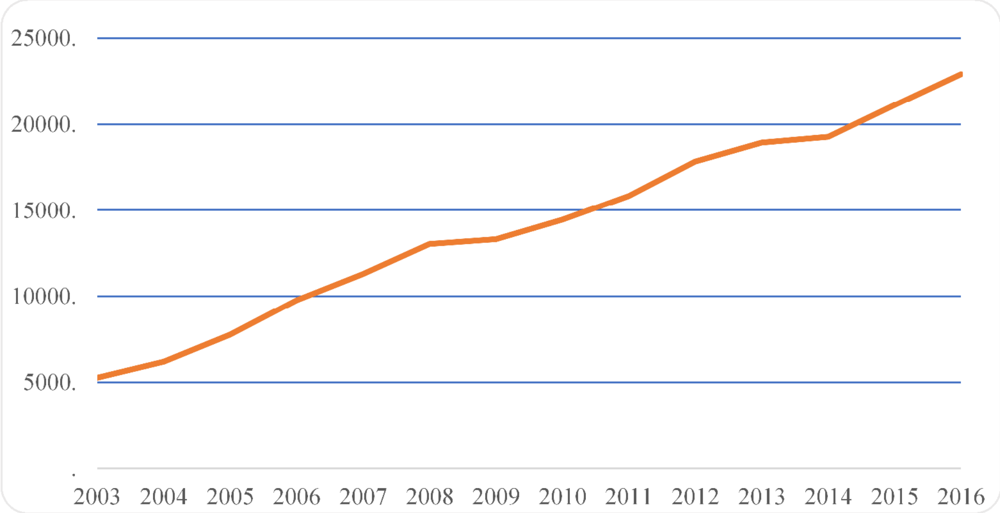

Figure 2. Total credit facilities by licensed banks in Jordan (CBJ)

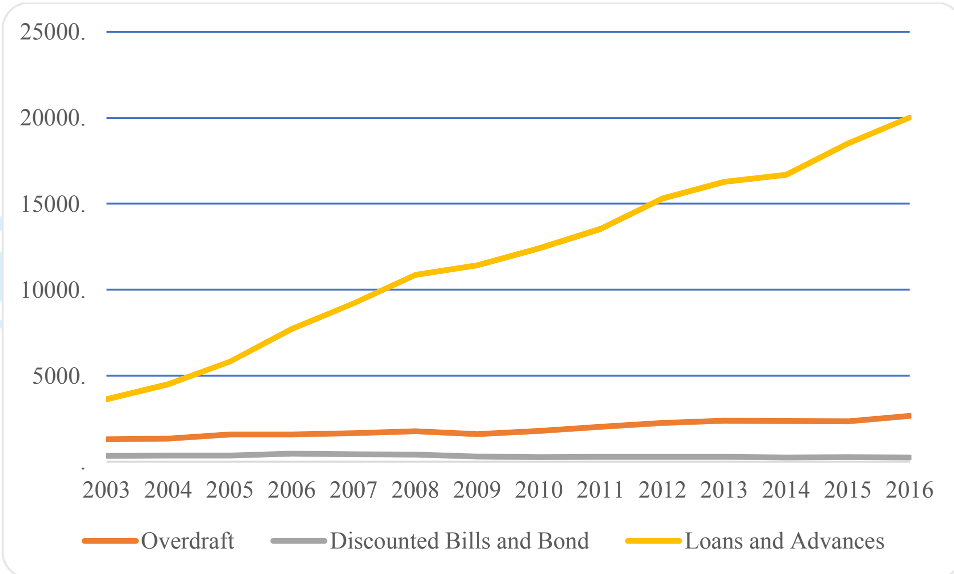

Figure 3. Credit Facilities According to Type (CBJ)

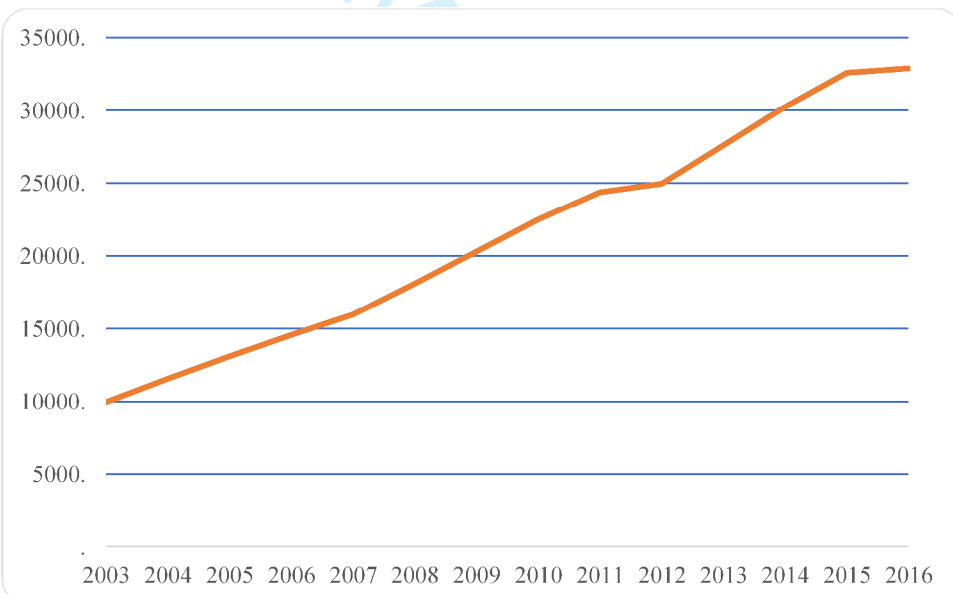

Figure 4. Total Deposits by licensed banks in Jordan (CBJ)

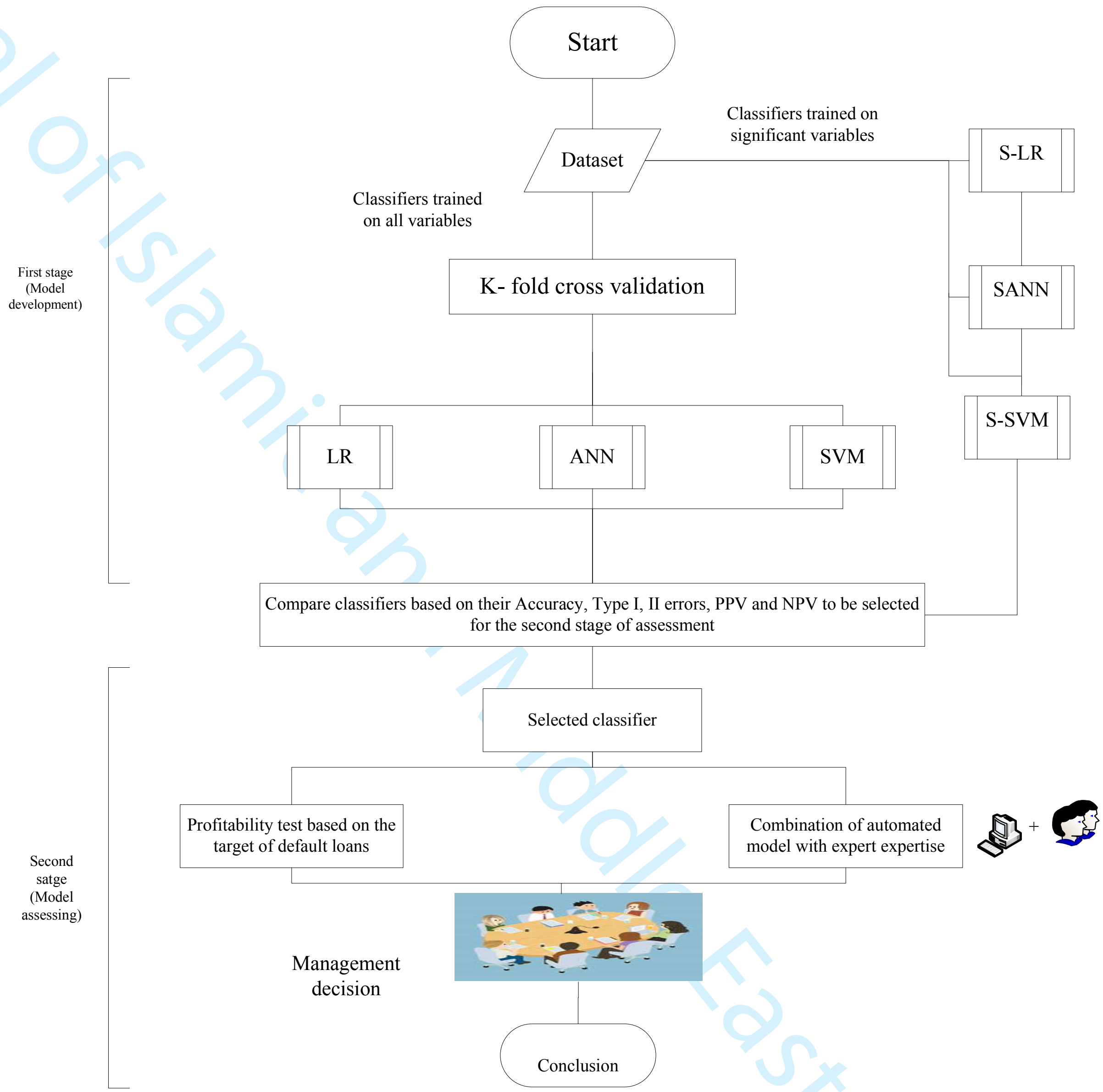

Figure 5. Research workflow

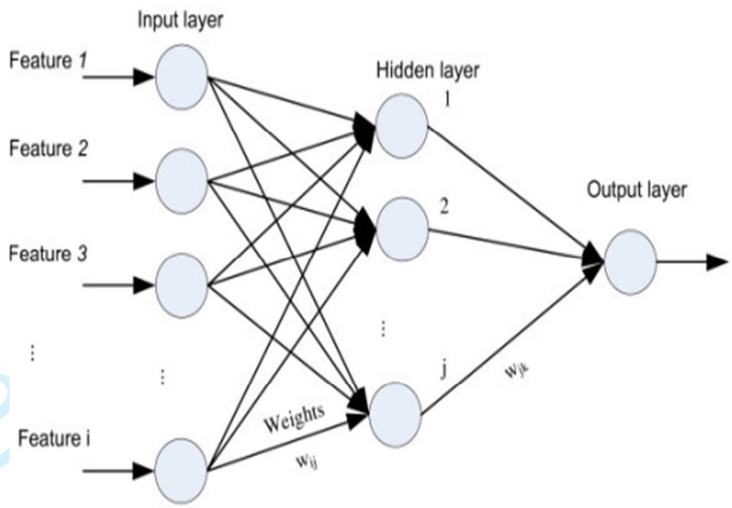

Figure 6. ANN topology

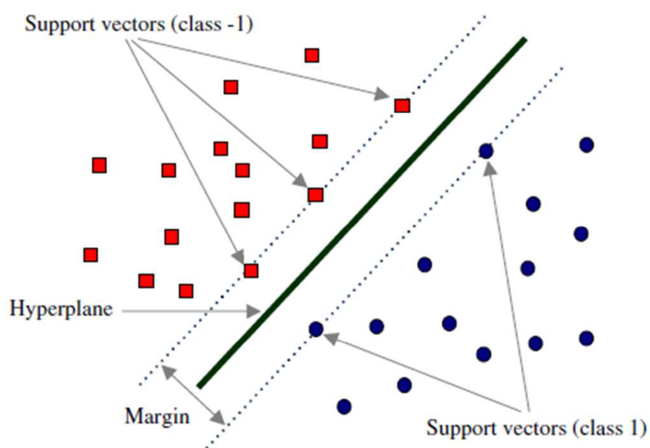

Figure 7. The basis of support vector machines

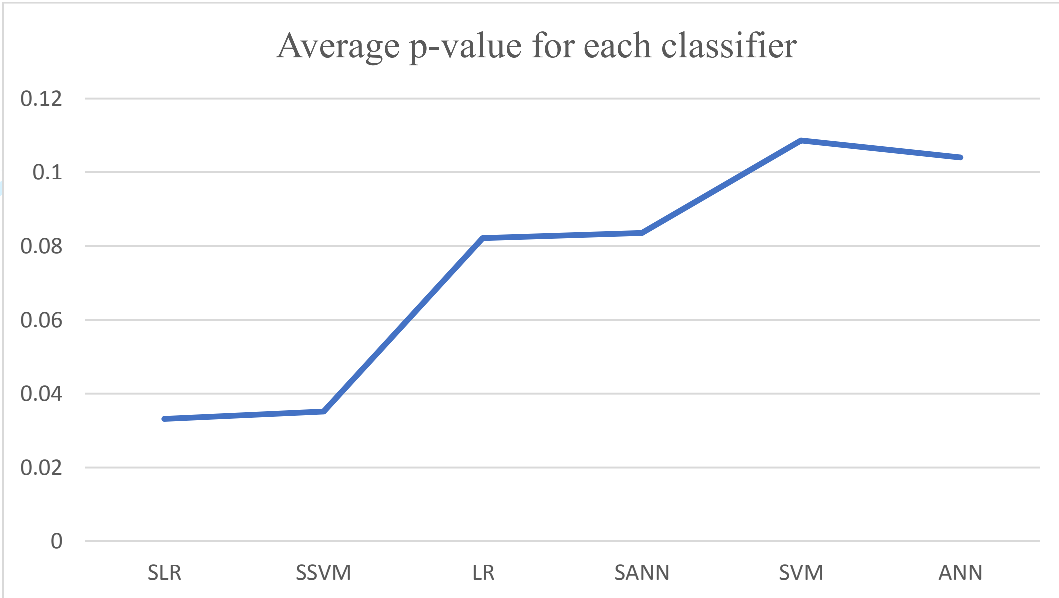

Figure 8. Average p-value for all classifiers

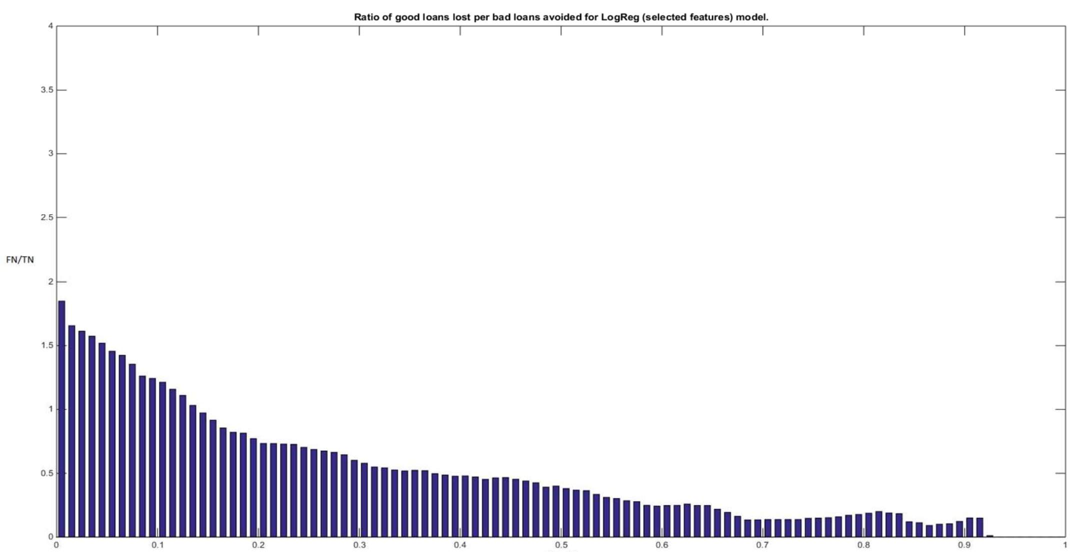

Figure 9. The ratio of FN loans per TN avoided

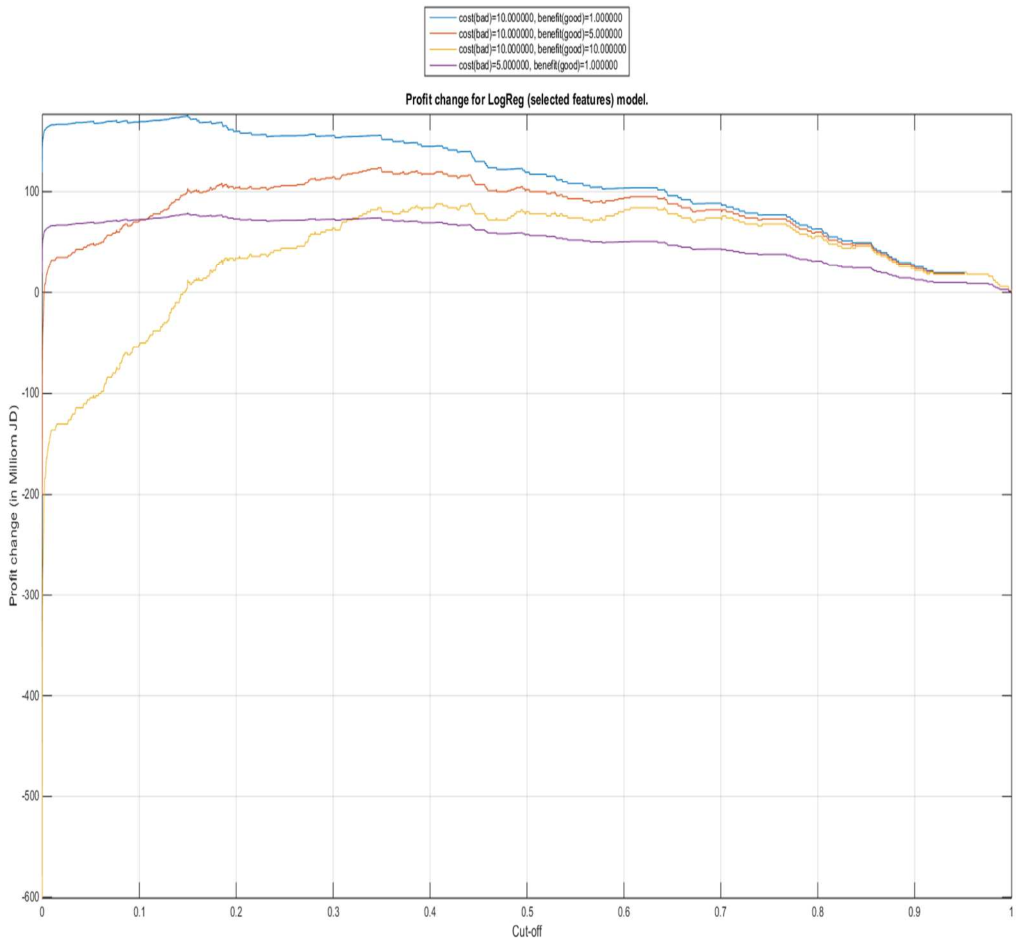

Figure 10. The profit change of SLR model over different ratios