Expert Systems with Applications 40 (2013) 4404–4413

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions

Terry Harris *

*Credit Research Unit, Department of Management Studies, The University of the West Indies, Cave Hill Campus, Barbados*

ARTICLE INFO

*Keywords:*
Credit risk assessment
Credit scoring
Credit unions
Default definitions
Support vector machine

ABSTRACT

This paper compares support vector machine (SVM) based credit-scoring models built using Broad (less than 90 days past due) and Narrow (greater than 90 days past due) default definitions. When contrasting these two types of models, it was shown that models built using a Broad definition of default can outperform models developed using a Narrow default definition. In addition, this paper sought to create accurate credit-scoring models for a Barbados based credit union. Here, the results of empirical testing reveal that credit risk evaluation at the Barbados based institution can be improved if quantitative credit risk models are used as opposed to the current judgmental approach.



## 1. Introduction

Over the past decade, credit risk analysis has attracted significant attention from decision-makers at financial institutions around the world. This is in part due to the global economic crises and recent regulatory developments (e.g., Basel III). In addition, the increased competition within the financial services industry has led many firms to find innovative ways of leveraging risk in order to attain and/or maintain competitive advantage. As a result, in today's economic and business environment financial institutions face greater risk of losses associated with inappropriate credit approval decisions (Yu, Wang, & Lai, 2008).

To manage the increased risk of default (credit risk) facing financial institutions, evermore effective credit appraisal techniques are being developed. Traditional methods for evaluating customers' credit risks are based on the experience and judgement of staff. However, with increases in the number of applicants, these conventional approaches have become outdated, as they can no longer meet the demands for efficient and effective credit risk assessment.

In recent years, credit-scoring has emerged as a leading method used by financial institutions to assess credit risk (Huang, Chen, & Wang, 2007). The main idea behind credit-scoring involves the classification of potential customers into applicants with good credit and applicants with bad credit. This is done by evaluating the probability that the applicant will default based on a quantitative model built from historical data of past applicant/customer behaviour (Thomas, Oliver, & Hand, 2005).

When developing quantitative credit scorecards based on past customer behaviour, the criteria used to determine when a client is in default needs to be determined. Here, the Basle Committee on Bank Supervision has established two widely accepted criteria. According to the Basle Committee on Banking Supervision (Basel, 2006), a default is considered to have occurred when either or both of the two following events have taken place:

- The financial institution (Bank) considers that the obligor is unlikely to pay its obligations in full, and the financial institution is unable to realise (sell) security (if held) in order to satisfy the obligor's debts.
- The obligor is more than 90 days past due on any material credit obligation to the financial institution.

Possibly due to its ease of determination and less subjective nature, over the years many credit-scoring models have been developed using the 90 days past due rule as the default indicator. This practice raises two important questions. Firstly, can the performance of credit-scoring models be improved if more relaxed default definitions are used (e.g., 30 days past due, and 60 days past due)? In this paper, these more relaxed default definitions are referred to as "Broad" default definitions. The second question that emerges is whether the discriminatory properties of credit-scoring models can be improved if more severe default definitions are used to create quantitative credit scorecards (e.g., 120 days past due, and 150 days past due)? These more severe default definitions are referred to as "Narrow" definitions. To help shed light on these questions quantitative credit-scoring models are developed based on various default definitions, using data taken from a Barbados based credit union (the credit union).

* Tel.: +1 246 8246118.
*E-mail address:* terry.harris@cavehill.uwi.edu

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.01.044

Preliminary investigation suggests that financial institutions in Barbados—a Small Island Developing State (SIDS)—have been slow to adopt quantitative credit-scoring as a means of credit risk assessment. The local credit unions are no different. Interviews with senior officials at one of the country's largest credit unions, suggest that most (if not all) credit unions in Barbados use the traditional judgmental approach when evaluating a potential client's credit risk. Accordingly, this paper also investigates whether credit risk assessment in a Barbados based credit union could be improved if a modern approach is adopted. While the impact of quantitative credit-scoring models in the credit union environment has been previously investigated, this study represents the first of its kind in the Barbados and SIDS context (Desai, Crook, & Overstreet, 1996).

The remainder of this paper is organised as follows. In Section 2, a brief-background regarding the credit union movement is presented. In addition, the current situation at the Barbados based credit union is outlined. Finally, this section presents a brief discussion on quantitative credit-scoring. In Section 3 the model evaluation metric and the model performance metrics used in this study are briefly discussed. Section 4 presents the Support Vector Machine algorithm, the classification technique used to develop the credit scoring models for the credit union. The details of the loan datasets provided by the credit union are presented in Section 5. Described in Section 6, is the methodology of the study. Section 7, discusses the results of the study, and Section 8 highlights the conclusions and directions for future research.

## 2. Credit unions and credit-scoring

### 2.1. What is a credit union?

Credit unions are co-operative financial institutions that share a common collectivist philosophy. Notwithstanding this similarity, credit unions differ according to many qualitative and quantitative characteristics. One major distinction between credit unions is often the common bond requirement for membership. The common bond is that characteristic that ties members together (e.g., profession, religion, vocation).

Once individuals have joined a credit union as members they enjoy equal rights to vote and participate in the governance and management of their institution. This democratic approach to corporate governance, in which all members are seen as key stakeholders without regard to deposit or loan size, has added to the popularity of credit unions as financial institutions of choice for many individuals across the globe.

Over the decades, there has been tremendous growth in the membership of credit unions. In addition to the democratic underpinnings of credit union philosophy, this trend has been attributed to the increased services on offer from credit unions, and to more relaxed interpretations of common bond requirements for membership.

### 2.2. Credit unions in Barbados and the situation at the study institution

In Barbados, the first credit union was formed in 1947. Since this time there has been an exponential growth in the number of credit unions and credit union members on the Island. At the time of the writing of this paper, the largest Barbados based credit union had over BB $ 791 billion in total assets reported in its 2012 annual report.

Credit union expansion in Barbados has not been without its challenges. The recent "Financial Stability Report" published by the Central Bank of Barbados (CBB) disclosed that local credit unions are experiencing increased delinquency rates since 2008. Loans' in arrears three months and over now stand at around 7.8% of gross credit union loans (CBB 2011, 2012).

In the case of the study institution, during the financial year 2011 to 2012 the percentage of nonperforming loans in the total loan portfolio moved from 4.8% to 6.9%. Further analysis of the position of the credit union reveals a critical situation with regard to the percentage increase in nonperforming loans when comparing closing 2011 and 2012 figures (Table 1). Impaired consumer loans has rinsed by 69.31% in 2012 when compared to 2011 closing balances. There has been a triple digit (160.81%) increase in non-performing Business loans. Moreover, impaired mortgage loans have increase by 27.27%. This situation is unsustainable and all efforts to arrest it, including the development of quantitative credit scorecards should be considered.

### 2.3. Credit-scoring

The literature has demonstrated that quantitative credit risk assessment using credit-scoring is an accurate means of credit risk evaluation (Crook, Edelman, & Thomas, 2007; Hand & Henley, 1997; Thomas, Edelman, & Crook, 1987; Thomas et al., 2005). Indeed since Fisher's (1936) seminal paper, there have been numerous quantitative studies outlining techniques aimed at differentiating between "good" and "bad" credit applicants. Many of these classification models are based on classical statistical methods such as Discriminant Analysis.

Discriminant analysis was first proposed by Fisher (1936) and is a parametric technique that has been widely applied in credit-scoring applications to discriminate between the two groups of applicants. For instance, Durand (1941) used discriminant analysis to evaluate car loan applicants. In addition, Altman (1986) used discriminant analysis to examine corporate bankruptcy.

Another popular statistical technique used to predict the likelihood of applicant delinquency is Linear Regression. When used for credit-scoring this technique establishes a threshold credit-score. This threshold score is derived from the linear (or polynomial) relationships between historic client features and their associated weights. Eq. (1) below depicts the classical hypothesis function used when training a linear regression classifier.

$$Z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n \tag{1}$$

Here the variable $n$ represents the number of features collected from potential clients. The features themselves are represented by the $x$'s. The $\theta$'s represent the associated weights. The feature variables along with their weights are used to produce a credit-score, $Z \in \mathbb{R}$. When using linear regression, an applicant who scores below the threshold is rejected, while an applicant who scores above the predetermined threshold is granted credit. Orgler (1970) was one of the first researchers to use linear regression for credit-scoring. His work, on commercial loan analysis, demonstrated how this technique could be used in practical credit-scoring applications (Orgler, 1971).

Logistic Regression, as in (2), is another widely used classical statistical technique that has been applied to the field of credit-scoring. Logistic regression can be thought of as a special case of linear regression where $Z \in \{1,0\}$. To achieve this, the logistic/sigmoid function, $g(x) = \frac{1}{1+e^{-x}}$, is used to restrict the values assigned to $Z$.

$$Z = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \tag{2}$$

Wiginton (1980) was one of the first researchers to use logistic regression for credit-scoring. Although his results were not very impressive the simplicity of logistic regression has led to it

**Table 1**
Showing percentage increase in non-performing loans at study institution.

| Category | | 2012 (bal) $ | Difference $ | 2011 (bal) $ | Percentage increase (%) |
|---|---|---|---|---|---|
| Consumer | Performing | 414,032,806 | 20,705,687 | 392,327,119 | 5.28 |
| | Impaired | 29,691,068 | 12,154,729 | 17,536,339 | 69.31 |
| Business | Performing | 3,092,846 | 87,735 | 3,005,111 | 2.92 |
| | Impaired | 534,718 | 329,699 | 205,019 | 160.81 |
| Mortgage | Performing | 195,882,326 | 2,643,879 | 193,238,447 | 1.37 |
| | Impaired | 15,181,078 | 3,252,816 | 11,928,262 | 27.27 |

becoming a popular approach in many practical credit-scoring applications (Hosmer & Lemeshow, 1989).

In addition to these conventional statistical techniques, other classification methods have also been applied to the problem of quantitative credit risk assessment. Mathematical programming techniques such as Linear Programming, Quadratic Programming, and Integer Programming are widely used. More recently, machine learning techniques have been used to develop credit-scoring models. Some popular techniques from the literature include Artificial Neural Networks and Support Vector Machines (Bellotti & Crook, 2009; Huang, Chen, Hsu, Chen, & Wu, 2004; Wang, Wang, & Lai, 2005).

## 3. Model evaluation and model performance metrics

In this paper a distinction is made between the use of metrics during (i) the parameter selection/cross-validation phase and (ii) the reporting phase. The term evaluation-metric is used when referring to the metric used during the cross-validation (CV) phase. Alternatively, the term performance-metric is used when referring to the metric (s) used to report on the performance of the model during the reporting phase.

Many existing credit scoring models are built on samples of customer historical data where the primary objective is to avoid over-fitting while maximising generalisablity from the samples (Huang et al., 2004). To do this, improving the performance of the evaluation-metric during model development (CV phase) is of importance (Huang et al., 2004; Wang et al., 2005).

One popular evaluation-metric observed in the available literature is predictive accuracy, as in (3). This metric is the measure of how accurately the model classifies credit applicants on the cross-validation datasets. When credit-scoring models are developed using predictive accuracy as the evaluation-metric higher accuracy rates are achieved on this metric when it is used to report on performance (reporting phase). However, a problem emerges when the dataset used to build the model is skewed. This is because it becomes difficult to determine if higher percentage accuracy equates to better overall classification.

For example, suppose a financial institution develops a classifier which achieves accuracy of 95% when determining the creditworthiness of loan applicants. This system may seem to be a good means of classification. However, if the probability of a potential customer being un-creditworthy is only 1%, it becomes clear that predictive accuracy says very little about the quality of the classifier because 99% predictive accuracy can be achieved by classifying all applicants as creditworthy.

$$\text{Predictive Accuracy} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} + \frac{\text{True Negative}}{\text{False Negative} + \text{True Negative}} \tag{3}$$

Accordingly, predictive accuracy should not be used (as the evaluation-metric) when datasets are shewed. Other metrics that address this problem include the Precision, as in (4), and Recall, as in (5), evaluation metrics. Precision is the measure of how accurately the positive predictions have been classified (what fraction is correctly categorised), while Recall measures the proportion of the dataset, which was actually positive, that was predicted as positive. In the previous scenario, the algorithm that simply predicts that the applicant was creditworthy 100% of the time would continue to score 99% on predictive accuracy; however, it would score 0% on the Recall evaluation metric. As a result, tailoring classification models to improve Precision and/or Recall can help to improve classifier quality when the dataset is skewed.

$$\text{Precision} = \frac{\text{True Positive}}{\#\text{Predicted as Positive}} = \frac{\text{True Positive}}{\text{True positive} + \text{False Positive}} \tag{4}$$

$$\text{Recall} = \frac{\text{True Positive}}{\#\text{Actually Positive}} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{5}$$

In practical credit-scoring, the problem arises concerning the disproportionate cost of making Type I error and Type II error. If the null hypothesis on any credit approval decision is that the credit applicant is un-creditworthy, then a Type I error occurs when the null hypothesis is rejected when it should have been accepted. In this case the potential customer who is actually un-creditworthy is granted credit. Conversely, a Type II error occurs when the null hypothesis is accepted when it should have been rejected. This results in the financial institution denying credit to a creditworthy applicant. Clearly, models could be developed to minimise Type I and Type II errors separately and/or jointly. However, focusing solely on effectively minimising Type I and II errors or maximising Precision and Recall does not take into consideration the misclassification cost to the institution of making one type of error over another (Hand & Henley, 1997).

Accordingly, another important model evaluation metric in the field of credit scoring is the estimated misclassification cost (West, 2000). This measure takes into consideration the unequal costs associated with (i) granting credit to an ultimately un-creditworthy applicant (making type I error) and (ii) denying credit to an applicant who would have been creditworthy (making type II error). However, as noted by Lee and Chen (2005), the estimation of misclassification cost is a complicated and challenging endeavor; therefore, valid prediction might not be possible.

The Area under the Receiver Operating Characteristic (ROC) curve (AUC) model evaluation metric has emerged as one of the most promising and widely accepted model evaluation metric for credit scoring (Bellotti & Crook, 2009; Zhou, Lai, & Yen, 2009). The ROC curve is a two dimensional measure of classification performance where the sensitivity (the fraction of actual positives predicted as positive), as in (6) and the specificity (the fraction of actual negatives predicted as negative), as in (7) are plotted on the Y and X axis, respectively. The AUC metric as in (8), where $S_1$ represents the sum of the ranks of the creditworthy class, measures the area under the resulting curve. Here, 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.

$$\text{Sensitivity} = \text{Type I Accuracy} = \frac{\text{True Positive}}{\text{True Positive} + \text{False negative}} \quad (6)$$

$$\text{Specificity} = \text{Type II Accuracy} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}} \quad (7)$$

$$\text{AUC} = \frac{S_1 - \text{Sensitivity} * \frac{\text{Sensitivity}+1}{2}}{\text{Sensitivity} * \text{Specificity}} \quad (8)$$

The models presented in this paper will use the AUC as the model evaluation metric. As a result, this metric will be used to develop all models during the cross-validation phase. The performances of the models will be reported using all metrics mentioned in this section (except estimated misclassification cost).

## 4. The support vector machine

The Support Vector Machine (SVM) was first developed by Cortes and Vapnik (1995) for binary classification. To achieve this, the algorithm attempts to find the optimal separating hyperplane between classes by maximising the class margin (see Fig. 1). Points lying on the boundaries of the margin are called support vectors, while the middle of the margin is referred to as the optimal separating hyperplane. It is this margin maximisation characteristic of SVMs that is argued to improve the decision boundaries produced and hence leads to a better quality classifier. Accordingly, SVMs have been successfully used in many credit-scoring systems (Baesens et al., 2003; Bellotti & Crook, 2009; Lee, 2007; Wang et al., 2005).

### 4.1. SVMs for credit scoring

When a financial institution, such as the credit union, is presented with a new credit applicant, in order to make the credit approval decision the institution seeks to classify the applicant as either "good" or "bad" according to the SVM score. In the case of a linear SVM this score can be represented as the linear combination of the applicant's characteristics (features e.g., employment status, marital status, etc.) multiplied by some weights, as in (9).

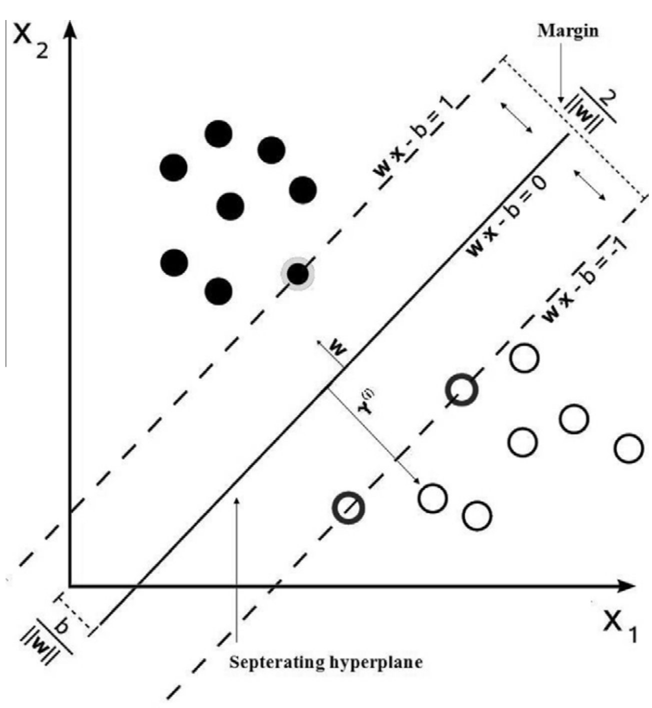


**Fig. 1.** Illustration of support vector machine (SVM).

$$z = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b \quad (9)$$

where $n$ represents the number of client features, the $w$'s and $b$ are learnt parameters, and the $x$'s are client features. Transforming the $w$'s and $x$'s into column vectors, (9) can be written more concisely as, $z = w^Tx + b$.

The SVM learns the parameters $w$ and $b$ from training examples of historic client data that the financial institution has collected over time. This training dataset will normally consist of a number of example clients; as a result, from a geometric perspective, calculating the value of $w$ and $b$ means looking for a hyperplane which best separates "good" clients from "bad". To do this, the SVM maximises the margin between the two clouds of data. As a result, when given a training example $(x^{(i)},y^{(i)})$, such that $y \in \{-1,1\}\ldots$, the functional margin $\hat{\gamma}$, of $(w,b)$ can be defined with respect to the training example as

$$\hat{\gamma} = y^{(i)}(w^Tx + b). \quad (10)$$

In order to confidently predict the class of the training example the functional margin needs to be large. Thus, if $y^{(i)} = 1$, then for the functional margin to be large $w^Tx + b$ must be a large positive number. As a result, if $y^{(i)} = -1$, then $w^Tx + b$ needs to be a large negative number. Accordingly, given a training set $S = \{(x^{(i)}, y^{(i)}); i = 1,\ldots,m\}$, the functional margin of $(w,b)$ with respect to $S$ is defined as the smallest of the functional margins of the training examples, as in (11).

$$\hat{\gamma} = \min_{i=1,\ldots,m}\hat{\gamma}^{(i)} \quad (11)$$

To find the geometric margin, $\gamma$, consider the case of a positive training example where $x^{(i)}$ corresponds to the label $y^{(i)} = 1$. The distance from this point to the decision boundary, $\gamma^{(i)}$, is a straight line (vector) orthogonal to the hyperplane (Fig. 1). To find the value of $\gamma^{(i)}$ the corresponding point on the decision boundary is found. This can be easily determined since $w/\|w\|$ is a unit-length vector pointing in the same direction as $w$. Therefore, the corresponding point on the hyperplane is given by the equation $x^{(i)} - \gamma^{(i)}.\ w/\|w\|$, and because this point lies on the decision boundary, it satisfies the equation $w^Tx + b = 0$ (Fig. 1), as in (12).

$$w^T\left(x^{(i)} - \gamma^{(i)}\frac{w}{\|w\|}\right) + b = 0 \quad (12)$$

The equation as in (12) can be simplified as follows:

$$w^Tx^{(i)} - \gamma^{(i)}\frac{w^Tw}{\|w\|} + b = 0. \quad (13)$$

Since, $w^Tw/\|w\| = \|w\|^2/\|w\| = \|w\|$, $\gamma^{(i)}$ can be solved for as is shown in (14);

$$\gamma^{(i)} = \left(\frac{w}{\|w\|}\right)^T x^{(i)} + \frac{b}{\|w\|}. \quad (14)$$

Generalising this representation to account for negative training examples, results in;

$$\gamma^{(i)} = y^{(i)}\left[\left(\frac{w}{\|w\|}\right)^T x^{(i)} + \frac{b}{\|w\|}\right]. \quad (15)$$

Here, if $\|w\| = 1$, then the geometric margin is equal to the functional margin. In addition, the geometric margin is invariant to rescaling of the parameters $(w,b)$. As a result, given a training set $S = \{(x^{(i)},y^{(i)}), i = 1,\ldots,m\}$, the geometric margin is the smallest of the geometric margins on the individual training examples (16).

$$\gamma = \min_{i=1,\dots,m} \gamma^{(i)} \tag{16}$$

Accordingly, when given a training dataset of past clients, it seems natural that the financial institution would want to find a decision boundary that maximises the geometric margin, since this would reflect a very confident set of predictions on the training data. Specifically, this will result in a SVM classifier that separates "good" and "bad" past clients effectively, thus giving the institution reliable information with which to make judgments about future credit applicants. As a result, to find the hyperplane that achieves the maximum geometric margin the following optimisation problem is posed:

$$\begin{aligned} &\max_{\gamma,w,b} \gamma, \\ &\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geqslant \gamma, \ \ i = 1, \dots, m, \\ &\|w\| = 1. \end{aligned} \tag{17}$$

However, because the $\|w\| = 1$ constraint is non-convex, the problem is transformed into one more suited for optimisation, as in (18). Here, if, $\hat{\gamma} = 1$, then $\hat{\gamma}/\|w\| = 1/\|w\|$, and maximising this is the same thing as minimising $\|w\|^2$.

$$\begin{aligned} &\min_{\gamma,w,b} \frac{1}{2}\|w\|^2, \\ &\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geqslant \hat{\gamma}, \ \ i = 1, \dots, m. \end{aligned} \tag{18}$$

At this point, a regularisation term $\xi$, is added to the optimisation problem posed in (18) to modify the algorithm so that it works for non-linearly separable datasets, as is often the case with credit scoring data. The term $C$ is a turning parameter which weights the significance of a classification error to the overall model.

$$\begin{aligned} &\min_{\gamma,w,b} \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i, \\ &\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geqslant 1 - \xi_i, \ i = 1, \dots, m, \\ &\xi_i \geqslant 0, \ \ i = 1, \dots, m. \end{aligned} \tag{19}$$

Eq. (19) represents the primal form of the optimisation problem for finding the optimal margin classifier to separate "good" and "bad" clients. Given that this equation satisfies the Karush–Kuhn–Tucker (KKT) conditions, the condition $g_i(w) \leqslant 0$ is an active constraint. As a result, the constraint to the primal problem can be rewritten as follows:

$$g_i(w) = -y^{(i)}(w^T x^{(i)} + b) + 1 - \xi_i \leqslant 0. \tag{20}$$

To develop the dual form of the problem, the Lagrangian for the optimisation problem is constructed, as in (21). Where the $\alpha_i$'s and the $r_i$'s are Lagrangian multipliers.

$$L(w,b,\xi,\alpha,r)\frac{1}{2}\|w\|^2 - c\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i[y^{(i)}(w^T x^{(i)} + b) - 1 + \xi_i] - \sum_{i=1}^{m}\mathrm{r}_i\xi_i \tag{21}$$

Eq. (21) is minimised with respect to $w$ and $b$ by taking partial derivatives with respect to $w$ and $b$ and setting them to zero. The equations derived are as follows:

$$\frac{\partial}{\partial w}L(w,b,\xi,\alpha,r) = w - \sum_{i=1}^{m}\alpha_i y^{(i)} x^{(i)} = 0, \tag{22}$$

$$\frac{\partial}{\partial b}L(w,b,\xi,\alpha,\mathrm{r}) = \sum_{i=1}^{m}\alpha_i y^{(i)} = 0. \tag{23}$$

Solving (22) for $w$ produces;

$$w = \sum_{i=1}^{m}\alpha_i y^{(i)} x^{(i)}. \tag{24}$$

Therefore, substituting the definitions of $w$(24) and $b$ (23) in (21) and including the constraints $0 \leqslant \alpha_i \leqslant C$ and $\sum_{i=1}^{m}\alpha_i y^{(i)} = 0$ the dual optimisation problem is derived as;

$$\begin{aligned} &W(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} y^{(i)}y^{(j)}\alpha_i\alpha_j < x^{(i)}, x^{(j)} >, \\ &\text{s.t. } 0 \leqslant \alpha_i \leqslant C, \ \ i = 1, \dots, m, \\ &\sum_{i=1}^{m}\alpha_i y^{(i)} = 0. \end{aligned} \tag{25}$$

This dual form (25) can be solved in lieu of the primal problem, in order to derive the parameters $\alpha_i$'s that maximise $W(\alpha)$ subject to the constraints. These parameters can then be used in (24) to find the optimal $w$'s. Having found $w^*$, the primal problem can be used to find the optimal value for the intercept term $b$.

Accordingly, after the classification model has been trained, when presented with a new credit applicant the equation $w^T x + b$, would calculate and predict $y = 1$ if and only if this quantity is bigger than zero.

$$(w^T x + b) = \left(\sum_{i=1}^{m}\alpha_i y^{(i)} x^{(i)}\right)^T x + b \tag{26}$$

Eq. (26) can be rewritten as;

$$\sum_{i=1}^{m}\alpha_i y^{(i)} < x^{(i)}, x > +b. \tag{27}$$

This representation allows for the inclusions of kernels to deal more effectively with datasets which have multiple dimensions. Kernels map attributes to higher order feature spaces, and this is represented by replacing the $x$'s in the equation with the feature vector $\phi(x)$, as shown in (28).

$$\sum_{i=1}^{m}\alpha_i y^{(i)} K(x^{(i)}, x) + b, \tag{28}$$

where

$$K(x^{(i)}, x^{(j)}) = < \phi(x^{(i)}), \phi(x^{(j)}) > . \tag{29}$$

To solve the dual problem, the Sequential Minimal Optimation (SMO) algorithm as proposed by Platt (1999) can be used. Recall the dual optimisation problem, as in (25), given a set of $\alpha_i$'s which satisfy the constraints, the values of any $\alpha_i$ and $\alpha_j$ are updated, where $i \neq j$, simultaneously in-order to continue satisfying the constraints. A description of the SMO algorithm is presented in Fig. 2 given below.

## 5. Data

In order to build the proposed credit-scoring models, a loan dataset was provided by the credit union. This dataset measured 20 client attributes and contained records of over 250,000 instances dating from 1997 to 2012. The client features measured by the dataset included: the number of months at current address, applicant's marital status, the number of dependents, the age of first dependent, the age of second dependent, the age of third dependent, the age of fourth dependent, the age of sixth dependent, the age of seventh dependent, the age of eight dependent,

Repeat till convergence {

Select some pair $\alpha_i$ and $\alpha_j$ to update using some heuristic.

Re-optimise $W(\alpha)$ with respect to $\alpha_i$ and $\alpha_j$, while holding all other $\alpha_k$'s constant.

}

**Fig. 2.** Illustration of sequential minimal optimation (SMO) algorithm.

the age of ninth dependent, the age of tenth dependent, applicant's employment status, the number of years employed with current employer, loan amount, loan purpose, loan type, applicant's monthly income, and applicants monthly expenditure.

Given the number of instances in the loan dataset, it was computationally prohibitive to build a model using the full dataset. This meant that sampling techniques had to be used to reduce the size of the dataset in order to facilitate its use. In addition, given the 15 year time period of historical data the issue of population drift emerges. As pointed out by Hand (2006) the class distribution of data can shift over time for many classification problems. This is certainly true for credit risk assessment data where changing economic, political, social, and legal conditions impact customer behaviour.

Dealing with these two issues (data-file size and population drift) simultaneously, the decision was taken to focus the study on the years 2011 and 2012. Thus, a stratified sampling technique was used, where all data instances for the years 2011 and 2012 were chosen. This sample dataset consisted of about 21,117 examples of creditworthy applicants and 503 examples of customers who should not have been granted credit based on the fact that they were past due on their month payment 30 days or greater. This resulted in an imbalanced ratio of approx. 42:1 (broad default definition) in favour of creditworthy applicants. This first sample is referred to as sample$_1$.

The sample$_1$ dataset was pre-processed so as to transform all categorical data into numerical data for analysis. Missing values were substituted using the variable median. In addition, the data was normalised so as to improve the performance of the SVM.

## 6. Methodology

Testing began by randomly sorting the processed sample$_1$ data-file before splitting it into two data-files—test (20%) and cross-validation (80%). Due to the imbalanced nature of sample$_1$ two other sample datasets were produced using sample$_1$'s cross-validation data-file. These samples were designed to reduce the class imbalance problem. As a result, the second sample dataset, referred to as sample$_2$, oversampled the un-creditworthy class from sample$_1$'s cross-validation data-file to produce an approx. 1:1 class ratio. The third sample, called sample$_3$, under-sampled the creditworthy class to produce an approx. 1:1 class ratio. Note that from this point on when sample$_1$ is mentioned the Author of this paper is referring to the cross-validation data-file of sample$_1$ as opposed to the full sample (test and cross-validation).

The withheld test dataset was strictly used to test model performance. This approach gives an idea as to the generalisablity of the credit scoring models. Arguably, a model's performance on the test dataset reasonably approximates real world performance. The test dataset was made up of 4,220 creditworthy applicants and 102 un-creditworthy applicants (30 days past due).

The cross-validation data-files (sample$_1$, sample$_2$, and sample$_3$) were fed into the parameter selection algorithm. Here, a five (5) fold cross-validation technique was used to select the parameters *Gamma* and *C* for the Support Vector Machine (RBF Kernel) algorithm.

The Support Vector Machine algorithm was implemented in the OCTAVE 3.2.4 programming language (Chang & Lin, 2011). A grid search technique was used to find the parameters which maximised the AUC using the cross-validation datasets (sample$_1$, sample$_2$, and sample$_3$).

Having found the parameters that maximised the AUC on the cross-validation datasets, these parameters were used to build the credit-scoring models using sample$_1$, sample$_2$, and sample$_3$. The performances of the models were evaluated using the withheld test dataset. These performances are presenting in the following section.

## 7. Results and analysis

Normally, credit data is not easily separable. This is because the data collected from applicants cannot capture the ever facet in each customer's life that may lead to future delinquency. Furthermore, the use of some information (e.g., race, ethnicity, and religion) that could result in better decision boundaries is prohibited by law when building credit scorecards. This results in higher misclassification rates than would normally be acceptable for other classification problems (Baesens et al., 2003). Nevertheless, the literature on credit risk assessment evidences the fact that credit-scoring is a useful means of credit risk evaluation. In this section the performances of the credit-scoring models developed are reported.

### *7.1. Credit scoring models for the credit union*

#### *7.1.1. Model$_1$*

The first model, model$_1$, was built using the imbalanced sample dataset, sample$_1$. The performance of this model is displayed in the table (Table 2) below. Here, the default definition used was 30 days past due.

Table 2 shows the performance of model$_1$ on the withheld test dataset. This model achieves good performance in terms of test accuracy, which measures the proportion of clients from the test dataset who were correctly classified as "good" or "bad" for credit. At first glance this performance may seem excellent; however, if the skewness of data is taken into consideration this performance is not as impressive. A better measure of the discriminatory properties of the model is the AUC metric. Here it is observed that this first model achieves average performance on the withheld test dataset. Furthermore, the level of Type I accuracy is extremely poor (Type I accuracy is a measure of how accurately the model classifies applicants who are un-creditworthy as un-creditworthy). As a result, it can be concluded that this model has no meaningful discriminatory properties and is not suitable for the credit union.

#### *7.1.2. Model$_2$*

To compensate for the imbalanced nature of sample$_1$ a second model, model$_2$ (Table 3) was built using sample$_1$. This model addressed the 42:1 class imbalance favouring the creditworthy applicants by weighting the importance of the un-creditworthy applicants 42 times more important to the model than creditworthy applicants. Like model$_1$, the default definition used when developing this model was 30 days past due.

The second model achieved better performance (66.03%) in terms of AUC when compared to model$_1$ while remaining reasonably accurate at 67.51% test accuracy. The overall performance of this model can be attributed to the fact that a more severe misclassification penalty was placed on making Type I errors. Accordingly, the SVM algorithm produced the separating hyperplane that placed more emphasis on rejecting "bad" applicants. This results in the model achieving reasonable Type I and Type II accuracy rates. Therefore, because there is good reason to believe that performance on the

**Table 2**
Showing performance of Model$_1$ built using Sample$_1$.

| Performance metric | Percentage |
|---|---|
| Accuracy | 97.96 |
| Precision | 98.03 |
| Recall | 99.93 |
| F-score | 98.97 |
| AUC | 53.74 |
| Type I Accuracy | 8.21 |
| Type II Accuracy | 99.92 |

**Table 3**
Showing performance of $Model_2$ built using $Sample_1$.

| Performance metric | Percentage |
|---|---|
| Accuracy | 67.51 |
| Precision | 98.89 |
| Recall | 67.61 |
| F-score | 80.31 |
| AUC | 66.03 |
| Type I Accuracy | 62.35 |
| Type II Accuracy | 67.61 |

**Table 5**
Showing performance of $Model_4$ built using $Sample_3$.

| Performance metric | Percentage |
|---|---|
| Accuracy | 62.59 |
| Precision | 99.07 |
| Recall | 62.42 |
| F-score | 76.59 |
| AUC | 67.37 |
| Type I Accuracy | 71.08 |
| Type II Accuracy | 62.24 |

withheld test dataset is generalisable to the population it can be argued that were this model implemented, the credit union would be able to correctly identify approximately 62.35% of applicants who would reveal to be un-credit worthy (Type I accuracy) and approximately 67.61% of applicants who would turn-out to be creditworthy (Type II accuracy).

### 7.1.3. $Model_3$

A third model was built, $model_3$, using $sample_2$. Recall that this sample was produced by oversampling the un-creditworthy class so as to balance the data for analysis. The performance of this model on the withheld test dataset is shown in Table 4. Default was defined for this model as 30 days past due.

Like $model_1$, this model achieves excellent performance in terms of the test accuracy, precision, recall, and f-score performance metrics. However, the AUC result from this model is poor when compared to $model_2$. Similarly, Type I accuracy is insignificant; as a result, this model is not suitable for practical use.

### 7.1.4. $Model_4$

Another model, $model_4$, was built using $sample_3$. Recall that this sample dataset was built by under-sampling the creditworthy class of $sample_1$. The performance of $model_4$ on the withheld test dataset is shown in the table below (Table 5). Here again, the default definition used was 30 days past due.

This model compares well with the weighted model, $model_2$. As a result, its performance makes it a suitable candidate for application in the credit union environment. If used, one could reasonably expect that this model would correctly identify 71.08% of "bad" applicants as "bad, and 62.24% of "good" applicants as "good". Additionally, not only does this model result in the highest AUC score noted thus far, but the under-sampling technique has the added benefit of shorter computation times (due to smaller sample size). As a result, this method (under-sampling the larger creditworthy class) is seen as superior when building credit-scoring models from the credit union dataset. Accordingly, all subsequent models presented in this paper utilise this technique.

### 7.1.5. $Model_5$

A fifth model, $model_5$, was built using $sample_3$. Unlike the 2models presented thus far, this model was designed to predict the likelihood that an applicant will experience 60 days delinquency. Accordingly, 60 days past due was used as the default definition. The results of this model are shown in Table 6.

The results presented in Table 6 reveal that this model was reasonably predictive. Here, it was able to correctly predict the class label of 32.39% of "bad" applicants and 89.27% of "good" applicants. While on its own this model may seem to be a relatively average classifier, particularly with regard to the ability to correctly classify un-creditworthy applicants as un-creditworthy (Type I Accuracy). However, this model can be used in conjunction with $model_4$ or $model_2$ (the two acceptable 30 day models). The combined system could achieve significant classification gains for the credit union with $model_4$ or $model_2$ predicting the likelihood of 30 day default and $model_5$ predicting the likelihood of 60 day default. Accordingly, a simple decision rule as follows could be adopted (Fig. 3).

### 7.1.6. $Model_6$

Another model, $model_6$, was trained using $sample_3$. This model was built to predict the likelihood that an applicant will experience 90 days past due delinquency. The results of this model are shown below in Table 7.

The results presented in Table 7 indicate that this model achieves relatively poor performance in terms of the AUC metric. This suggests that this model is a relatively weak classifier. Nevertheless, like $model_6$ this model could be use with an ensemble of other classifiers in order to improve overall bad credit prediction.

### 7.1.7. $Model_7$

$Model_7$ was built using a more severe (Narrow) definition of default than that defined by the Basle Committee on Banking and Supervision. This model was built using $sample_3$ with a default definition of 120 days past due. Test results for this model are presented in Table 8.

Despite this model performing reasonably well in terms of the predictive accuracy performance metric, the discriminatory properties of this model are relatively weak. This conclusion is suggested by an AUC measure of 55.5%. Accordingly, this model is unsuitable to be used for credit-scoring in isolation. A better approach would be to use it in combination with other classifiers.

**Table 4**
Showing performance of $Model_3$ built using $Sample_2$.

| Performance metric | Percentage |
|---|---|
| Accuracy | 95.51 |
| Precision | 98.15 |
| Recall | 97.71 |
| F-score | 97.93 |
| AUC | 55.43 |
| Type I Accuracy | 9.41 |
| Type II Accuracy | 97.23 |

**Table 6**
Showing performance of $Model_5$ built using $Sample_3$.

| Performance metric | Percentage |
|---|---|
| Accuracy | 88.52 |
| Precision | 99.00 |
| Recall | 89.27 |
| F-score | 93.88 |
| AUC | 63.40 |
| Type I Accuracy | 32.39 |
| Type II Accuracy | 89.27 |

Begin
1. if $model_x$ == bad AND $model_y$ == bad
   reject
2. if $model_x$ == good AND $model_y$ == bad
   reject OR investigate further
3. if $model_x$ == bad AND $model_y$ == good
   reject OR investigate further
4. if $model_x$ == good AND $model_y$ == good
   accept

End

**Fig. 3.** Showing simple decision rule for two model system.

**Table 7**
Showing performance of $Model_6$ built using $Sample_3$.

| Performance metric | Percentage |
|---|---|
| Accuracy | 86.12 |
| Precision | 99.04 |
| Recall | 88.71 |
| F-score | 89.82 |
| AUC | 56.10 |
| Type I Accuracy | 13.31 |
| Type II Accuracy | 88.61 |

**Table 8**
Showing performance of $Model_7$ built using $Sample_3$.

| Performance metric | Percentage |
|---|---|
| Accuracy | 84.24 |
| Precision | 83.34 |
| Recall | 79.23 |
| F-score | 84.21 |
| AUC | 55.40 |
| Type I Accuracy | 13.29 |
| Type II Accuracy | 87.70 |

**Table 9**
Showing performance of $Model_8$ built using $Sample_3$.

| Performance metric | Percentage |
|---|---|
| Accuracy | 83.24 |
| Precision | 98.35 |
| Recall | 87.53 |
| F-score | 89.43 |
| AUC | 53.75 |
| Type I Accuracy | 13.22 |
| Type II Accuracy | 87.64 |

#### *7.1.8. $Model_8$*

The final model built, $model_8$, was built using $sample_3$. Like $model_7$ this model was built using a Narrow default definition. Specifically, 150 days past due was used as the default criterion when training and testing the model. Table 9 reports the performance of this model.

Results indicate that this model is a poor discriminator with AUC achieving a meagre 53.75%. As a result, on its own this classifier is unsuitable for use in the credit union.

### *7.2. Comparison of models built using different definitions of default*

A comparative analysis was conducted to investigate the questions as to whether model performance (in terms of 90 days past due delinquency detection), could be improved by training the credit-scoring models using (i) Broader definitions of default, or (ii) Narrower definitions of default. Here, $model_4$ a 30 days past model, $model_5$ a 60 days past due model, $model_6$ a 90 days past due model, $model_7$ a 120 days past due model, and $model_8$ a 150 days past due model are compared. The performances of the models on a withheld test dataset where 30 days past due, 60 days past due, 90 days past due, 120 days past due, and 150 days past due are the default criteria, are presented in Table 10.

The results displayed in Table 10 reveal an interesting trend. That is, models built using Broad definitions of default performed noticeably better, in terms of the AUC, when predicting the likelihood of default where the default definition were more severe (Narrow). Conversely, it can be seen that models that were built using Narrow definition of default, performed progressively worst when used to predict default on test sets that have more relaxed (Broad) definitions of applicant default. Fig. 4 below presents this information more concisely.

The results presented in Table 10 and Fig. 4 suggest that models built using a broader definition of default can be more accurate when predicting more severe cases of default than models built specifically to predict those more severe cases. At first glance this result may appear to go against intuition, as one may expect that a model built to identify a severe case of delinquency would do a better job at this than a model designed to identify a less critical level of default. However, upon deeper inspection some good reasons for this occurrence are revealed. For instance, models built using a broader definition of default will inherently have more cases of default in the training dataset, and as a result the classifier will "see" more instances of what a potentially "bad" applicant looks like. Hence the classifier would have a better opportunity to learn a general pattern of applicant delinquency. Conversely, when a more severe case of default is used, the classifier may not be fed as many instances of "bad" applicants; therefore, the classifier may not be able to learn a broad pattern describing uncreditworthiness.

## 8. Conclusion

In this paper, the SVM algorithm is used to develop credit-scoring models for a credit union in Barbados. This quantitative approach to credit risk assessment at financial institutions in Barbados and other SIDS is currently an under-utilised (if not non-existent) practice. Interviews with key personal at a large local credit union reveal that presently a majority of Barbadian credit unions use the traditional judgmental approach when making credit approval decisions. This was the case at the credit union chosen to conduct this study. Brief analysis of the institutions annual reports indicated that there is a serious situation with regard to the non-performing loans. Here the percentage increase in non-performing loans when compared to the closing balances for 2011 and 2012 present an alarming circumstance. To address this situation a number of credit-scoring models were developed. Results presented in this paper suggest that the use of the suitable credit-scoring models developed would lead to improve the decision-making at the institution.

When investigating whether models built using (i) more relaxed default definitions (Broad) and/or (ii) more severe default definitions (Narrow) can have an impact on classifier performance. Findings from this study suggest that models built using a broader definition of default were more accurate when used to predict more severe cases of default. One possible explanation for this is that when a broader definition for default is used to develop the model it is usually fed more cases of default applicants. As a result, the classifier learns a better pattern of un-creditworthiness. On the other hand, when a more severe definition of default is used, the classifier may not be fed as many instances of the un-creditworthy class. Therefore, the classifier may not be able to learn an accurate description of un-creditworthiness. This finding suggests that

**Table 10**
Showing comparative model performances.

| Models | Default | Accuracy (%) | Precision (%) | Recall (%) | F-Score (%) | AUC (%) | Type I accuracy (%) | Type II accuracy (%) |
|---|---|---|---|---|---|---|---|---|
| $Model_4$ | **30 days** | **62.59** | **99.07** | **62.42** | **76.59** | **67.37** | **71.08** | **62.24** |
| | 60 days | 61.84 | 99.43 | 61.70 | 76.14 | 68.28 | 73.24 | 61.70 |
| | 90 days | 61.56 | 99.67 | 61.46 | 76.03 | 90.14 | 74.41 | 61.46 |
| | 120 days | 61.13 | 99.62 | 61.33 | 75.85 | 92.12 | 77.14 | 61.12 |
| | 150 days | 60.67 | 99.71 | 61.21 | 75.86 | 93.54 | 79.12 | 61.08 |
| $Model_5$ | 30 days | 88.85 | 98.45 | 90.03 | 94.06 | 62.73 | 29.41 | 90.04 |
| | **60 days** | **88.52** | **99.00** | **89.27** | **93.88** | **63.40** | **32.39** | **89.27** |
| | 90 days | 88.94 | 99.48 | 89.31 | 89.31 | 93.79 | 41.86 | 89.31 |
| | 120 days | 88.43 | 99.52 | 88.67 | 89.53 | 93.98 | 55.52 | 89.33 |
| | 150 days | 89.02 | 99.35 | 89.23 | 89.42 | 94.22 | 62.23 | 89.45 |
| $Model_6$ | 30 days | 88.21 | 98.56 | 90.53 | 90.68 | 53.46 | 10.98 | 90.51 |
| | 60 days | 87.65 | 98.61 | 89.89 | 90.53 | 54.65 | 11.54 | 89.56 |
| | **90 days** | **86.12** | **99.04** | **88.71** | **89.82** | **56.10** | **13.31** | **88.61** |
| | 120 days | 85.65 | 99.56 | 88.78 | 89.64 | 60.35 | 22.56 | 87.13 |
| | 150 days | 86.21 | 99.63 | 89.83 | 89.94 | 66.34 | 27.78 | 86.98 |
| $Model_7$ | 30 days | 82.32 | 83.66 | 72.21 | 82.21 | 51.13 | 12.62 | 89.87 |
| | 60 days | 82.54 | 83.08 | 80.56 | 80.56 | 51.32 | 12.97 | 89.60 |
| | 90 days | 83.45 | 83.15 | 80.59 | 80.54 | 54.33 | 13.03 | 88.62 |
| | **120 days** | **84.24** | **83.34** | **79.23** | **84.21** | **55.40** | **13.29** | **87.70** |
| | 150 days | 84.54 | 84.15 | 80.48 | 84.91 | 65.59 | 35.16 | 88.18 |
| $Model_8$ | 30 days | 84.51 | 98.89 | 87.61 | 89.31 | 50.03 | 11.35 | 89.61 |
| | 60 days | 84.40 | 98.44 | 86.33 | 89.58 | 50.63 | 11.54 | 89.32 |
| | 90 days | 83.27 | 98.61 | 86.26 | 89.68 | 52.01 | 12.44 | 89.26 |
| | 120 days | 84.13 | 98.25 | 86.32 | 89.32 | 53.21 | 12.57 | 89.21 |
| | **150 days** | **83.24** | **98.35** | **87.53** | **89.43** | **53.75** | **13.22** | **87.64** |

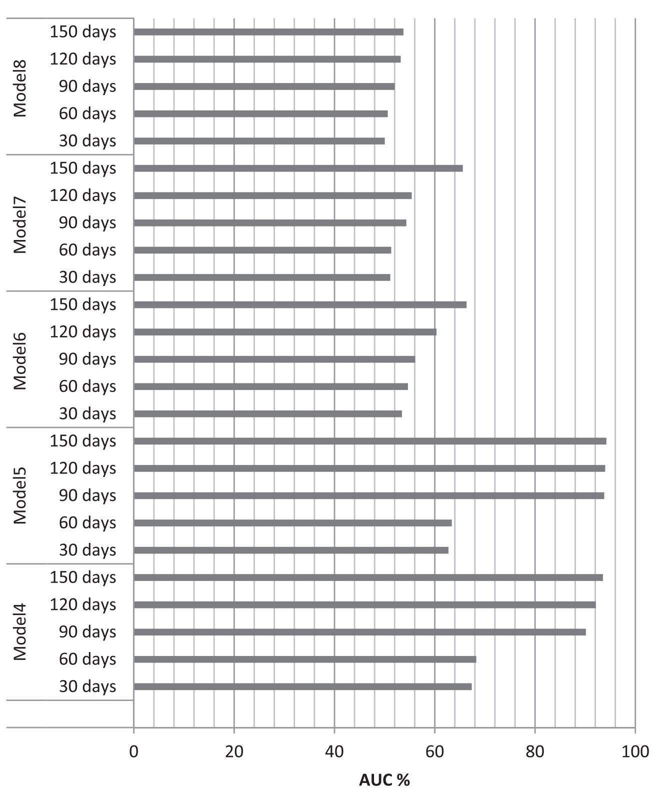


**Fig. 4.** Showing comparative performances of the models on the test datasets.

model developers need to assess the impact of the predetermined default definition on model performance.

Future work will consider the generalisablity of the sampling approaches and findings to other classifiers and classification problems. In addition, other studies will investigate the advantages and disadvantages of using other evaluation metrics as the primary model evaluation metric. Furthermore, future studies will consider ensemble techniques—bagging and boosting—along with other performance enhancing techniques.

## References


Altman, E. I. (1986). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance, 23*(4), 589–609.

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*, 1082–1088.

Basel, Committee. (2006). International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Comprehensive Version. Switzerland: Bank for International Settlements.

Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications, 36*(2), 3302–3308.

CBB. (2011). Financial Stability Report. Bridgetown, Barbados: The Central Bank of Barbados.

CBB. (2012). Financial Stability Update August 2012 Bridgetown, Barbados: The Central Bank of Barbados.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST), 2*(3), 27.

Cortes, C., & Vapnik, V. (1995). Support vector networks. *Machine Learning, 20*(3), 273–297.

Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research, 183*(3), 1447–1465.

Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research, 95*(1), 24–37.

Durand, D. (1941). *Risk elements in consumer instalment financing. Studies in Consumer Instalment Financing*. NY: National Bureau of Economic Research..

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Human Genetics, 7*(2), 179–188.

Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science, 21*(1), 1–14.

Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 160*(3), 523–541.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: John Wiley & Sons.

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.

Huang, C. L., Chen, M. C., & Wang, C. J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications, 33*(4), 847–856.

Lee, Y. C. (2007). Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications, 33*(1), 67–74.

Lee, T. S., & Chen, I. F. (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.

Orgler, Y. E. (1970). A credit scoring model for commercial loans. *Journal of Money, Credit and Banking, 2*(4), 435–445.

Orgler, Y. E. (1971). *Evaluation of bank consumer loans with credit scoring models*. Tel-Aviv University, Department of Environmental Sciences.

Platt, J. (1999). *Fast training of support vector machines using sequential minimal optimization. Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press.

Thomas, L. C., Edelman, D. B., & Crook, J. N. (1987). Credit Scoring and its Applications: Society for Industrial Mathematics.

Thomas, L. C., Oliver, R. W., & Hand, D. J. (2005). A survey of the issues in consumer credit modelling research. *Journal of the Operational Research Society, 56*(9), 1006–1015.

Wang, Y., Wang, S., & Lai, K. K. (2005). A new fuzzy support vector machine to evaluate credit risk. *Fuzzy Systems, IEEE Transactions on, 13*(6), 820–831.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research, 27*(11), 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.

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.

Zhou, L., Lai, K. K., & Yen, J. (2009). Credit scoring models with AUC maximization based on weighted SVM. *International Journal of Information Technology & Decision Making, 8*(04), 677–696.