

Review Article

Credit Scoring: A Review on Support Vector Machines and Metaheuristic Approaches

R. Y. Goh  ¹ and L. S. Lee  ^{1,2}

¹*Laboratory of Computational Statistics and Operations Research, Institute for Mathematical Research, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia*

²*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia*

Correspondence should be addressed to L. S. Lee; lls@upm.edu.my

Received 1 November 2018; Revised 28 January 2019; Accepted 18 February 2019; Published 13 March 2019

Academic Editor: Eduardo Fernandez

Copyright © 2019 R. Y. Goh and L. S. Lee. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Development of credit scoring models is important for financial institutions to identify defaulters and nondefaulters when making credit granting decisions. In recent years, artificial intelligence (AI) techniques have shown successful performance in credit scoring. Support Vector Machines and metaheuristic approaches have constantly received attention from researchers in establishing new credit models. In this paper, two AI techniques are reviewed with detailed discussions on credit scoring models built from both methods since 1997 to 2018. The main discussions are based on two main aspects which are model type with issues addressed and assessment procedures. Then, together with the compilation of past experiments results on common datasets, hybrid modelling is the state-of-the-art approach for both methods. Some possible research gaps for future research are identified.

1. Introduction

Credit scoring is the basis of financial institutions in making credit granting decisions. A good credit scoring model will be able to effectively group customers into either default or nondefault group. The more efficient it is, the more cost can be saved for a financial institution.

Credit scoring is used to model available data and evaluate every instance in the data with a credit score and probability of default (PD). Generally, score is a measurement of the credit-worthiness of customers, while the PD is the likelihood estimation of a customer fails to meet one's debt obligation in a given period of time. Hand and Henly [1] defined credit scoring as “the term used to describe formal statistical methods used for classifying applicants for credit into ‘good’ and ‘bad’ risk classes”. Since the final decision is binary, credit scoring is thus equivalent to a binary classification problem.

When credit cards started to be introduced in 1960s, necessity of credit scoring models is triggered. Financial institutions started to combine or replace purely judgemental-based credit granting decisions with statistical models as

customers database increased tremendously. In 2004, Basel II accord is released. Under credit risk, rather than the previous standardised method, Internal Rating Based (IRB) approach could be adopted by bank to compute the minimum capital requirement. This marked an evolution in the credit scoring field, where attempts to form sophisticated model have been actively researched. Together with the rapid growth of computer technology, formulation of sophisticated models is made possible.

Hand and Henley [1] first published review paper of credit scoring domain. They reviewed statistical methods and several data mining (DM) methods and concluded that the future trend of credit scoring models will be more complex methods. Thomas [2] also reviewed on past researches on credit scoring and pointed out the importance of profit scoring. The methods discussed ranged from statistical, operations research based, and DM approaches. Sadatrasoul et al. [3] reviewed DM techniques applied for credit scoring domain in year 2000-2012, showing the tendency of model building with DM methods in recent years.

There are also review papers that specifically focused on application scoring [4] and bankruptcy prediction [5, 6].

Martin [4] reviewed application scoring model in a different perspective, where procedure of scorecard performance assessment by past studies are categorized into consistency, application fit, and transparency. The author pointed out the weaknesses of past experiments that only pay attention to model development and neglected appropriate assessment procedures of the models. Sun et al. [6] reviewed on bankruptcy prediction credit models by providing clear definitions of bankruptcy prediction as resulted from literature throughout the years and then discussed the techniques based on three main approaches, i.e., modelling, sampling, and featuring. Alaka et al. [5] also focused on bankruptcy prediction models. They identified the popular statistical and artificial intelligence (AI) tools utilized and discovered 13 criteria of the tools usage. Based on the 13 criteria, they developed a framework as a guideline for tool selection. Besides, there is one review from Moro and Rita [7] that adopted a completely different approach than all the other review papers. An automated review procedure is conducted with text mining. Important issues in credit scoring domain and top ranked tools for model development are discovered through automated text mining.

Baesens et al. [8] are the first to build credit scoring model with Support Vector Machines (SVM) and compare its performance with other state-of-the-art methods. They experimented with standard SVM and Least Squares SVM (LS-SVM) and reported that LS-SVM yield good performance as compared to other methods. Thereafter, SVM has been actively researched in the credit scoring domain, being one of the mostly used DM methods to build credit scoring model. Recent review studies [5, 7] have also identified SVM as a significant tool to be selected among researchers for credit models development.

Metaheuristic Approaches (MA), especially Evolutionary Algorithm (EA), have also been introduced as the alternative to form credit scoring model. The surging applications in credit scoring can be seen in recent years, where most review papers included discussion of EA as one of the main DM methods (see Alaka et al. [5], Crook et al. [9], Lahsasna et al. [10], Lessmann et al. [11], and Louzada et al. [12]). In Louzada et al. [12] timeline discussion, it can be seen that, after year 2005, SVM and EA have increased research work in credit scoring domain. Besides, Marques et al. [13] have reviewed on credit scoring models specifically focusing in EA, indicating the popular usage of MA.

To date, reviews on credit scoring domain are mainly focusing on a wide category of methods to develop scorecard. The active research utilizing SVM prompted a need to particularly review on this method. There was only one review by Marques et al. [13] that focused on EA in credit modelling. However, EA are just a part of MA, where there are other MA that have received attention in the credit scoring domain in recent years due to the shift of modelling trend towards AI-based techniques. The surging application of both techniques has greatly increased contributed works, leading to a situation where general reviews are insufficient to peek into the development trend of these two methods. Hence, in contrast with past literature reviews which are general reviews, this study is a focused

review, particularly on the development of SVM and MA only.

All the research articles in this study are obtained from three online academic databases, Google Scholar, Science Direct, and IEEE Xplore/Electronic Library Online. Several main keywords applied to select the articles are credit scoring, credit risk prediction, metaheuristics, data mining, SVM, Genetic Algorithm, Genetic Programming, Evolutionary Computing, machine learning, and artificial intelligence. From the search results, there are 44 and 43 articles from SVM and MA models, respectively, with 12 articles utilizing both together, thus resulting in a total of 75 research articles being reviewed in this study ranging from year 1997 to 2018.

The objectives of this study are to review past literatures of using SVM and MA in developing credit scoring model, identify the contributions of both methods, and discuss the evolving trend that can lead to possible future works in credit scoring domain.

This paper is organized as follows. Section 2 discusses the evolving trend of the credit scoring models from traditional methods to DM methods. Section 3 briefly describes SVM and MA methods. Section 4 summarizes and discusses the results based on model type with issues addressed, assessment procedures, and results compilation from past experiments. Then, Section 5 suggests several possible future directions and draws conclusion for this study.

2. Trend

Before credit scoring model is developed, credit granting decision is purely judgemental-based. Statistical models started to be utilized since 1941 where Durand [14] is the first to pioneer the usage of discriminant analysis (DA) to classify good and bad loans. Altman [15] also used multiple discriminant analysis (MDA) to predict company bankruptcy. The financial ratios are treated as input variables and modelled with MDA. MDA model has shown good prediction ability and useful for analyst to provide investment recommendations. Ohlson [16] then introduced a probabilistic approach to predict credit-worthiness of companies in 1980. The method proposed is Logistic Regression (LOGIT). Few problems of using MDA are pointed out and LOGIT is believed to be more robust than MDA.

In 1985, Kolesar and Showers [17] introduced mathematical programming (MP) approach to solve credit granting problem. MP is compared with classical DA and reported results showed that MP is more robust and flexible for decision maker. Among the traditional methods, LOGIT turned out to be the standard credit scoring model because of its ability to fulfil all requirements from the Basel II accord.

Massive improvement in computer technology opens up DM approach in model building for credit scoring. There have been extensive researches done in the past that utilized DM methods in the credit scoring domain. Most of them compared the adopted methods with the standard LOGIT model and have shown the DM models are competitive. A

few comparative studies and review papers [1, 2, 8–12, 18] reported good performance of DM models, and among the various methods, SVM and MA (especially EA) have been widely researched to be the alternative to the credit scoring models.

3. AI Techniques

3.1. Support Vector Machines. Support Vector Machines (SVM) were first introduced by Vapnik [91] in 1998, in the context of statistical learning theory. There are many successful applications of SVM in pattern recognition, indicating SVM to be a competitive classifier.

SVM seeks for an optimal hyperplane with maximum margin that acts as the decision boundary, to separate the two different classes. Given a training set with labelled instance pairs (x_i, y_i) , where i is the number of instance $i = 1, 2, 3, \dots, m$, $x_i \in \mathbb{R}^n$ and $y_i \in \{-1, +1\}$, the decision boundary to separate two different classes in SVM is generally expressed as

$$w \cdot x + b = 0. \quad (1)$$

Optimal separating hyperplane is the one with maximum margin and all training instances are assumed to satisfy the constraint,

$$y_i (w \cdot x_i + b) \geq 1. \quad (2)$$

The convex optimization problem is then defined as

$$\begin{aligned} \min \quad & \phi(w, b) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \epsilon_i \\ \text{s.t.} \quad & y_i (w \cdot x_i + b) \geq 1. \end{aligned} \quad (3)$$

The optimal hyperplane is equivalent to the optimization problem of a quadratic function, where Lagrange function is utilized to find the global maximum. The ϵ_i is the slack variable introduced to take account for misclassification, with C as the accompanied penalty cost. For nonlinear classification, kernel trick is used to modify the SVM formulation. Popular kernel choices are linear, Radial Basis Function (RBF), and polynomial.

- (i) Linear: $x_i \cdot x_j$.
- (ii) Polynomial: $(x_i \cdot x_j + c)^d$, $c > 0$, $d \geq 2$.
- (iii) Radial Basis Function (RBF): $\exp\{-\gamma \|x_i - x_j\|^2\}$.

3.2. Metaheuristic Algorithms. Metaheuristic Algorithm (MA) is one of the AI-based data mining approaches which had gained attention in recent years. MA is an automated process to search for a near-optimal solution of an optimization problem. The search process is conducted with operators to ensure a balance between exploration and exploitation to efficiently seek for good solutions. Generally, MA consists of several categories; the description of the categories that have been used in credit scoring domain is given as follows:

(i) Evolutionary Algorithm (EA)

The EA approach has a mechanism to seek solution following the Darwinian principle, which is based on the “survival of the fittest” concept. There are four main procedures to search for a solution, i.e., selection, reproduction, mutation, and crossover. The solutions in the population are improved in an evolutionary manner, guided by the quality of the fitness function. The EA applied in credit scoring are Genetic Algorithm (GA) and Genetic Programming (GP).

(ii) Swarm Intelligence (SI)

The SI approach is a nature-inspired algorithm that conducts solution-seeking based on natural or artificial collective behaviour of decentralized and self-organized systems. The solutions in the population are improved through the interaction of agents with each other and also with the environment. Similarly, the generation of better solution is led by fitness function quality. The SI applied in credit scoring are Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Scatter Search (SS), Artificial Bee Colony Optimization (ACO), Honey Bees Mating Optimization (HBMO), Cuckoo Search (CS), and Harmony Search (HS).

(iii) Iterative Based (IB)

The IB approach focus to improve on one single solution in each iteration around its neighbourhood, where the solution improvement is based on the quality of fitness function. Thus, IB is different as compared to EA and SI which are population-based. The IB applied in credit scoring are Simulated Annealing (SA) and Tabu Search (TS).

4. Discussions

4.1. Model Types with Issues Addressed. SVM and MA credit models are categorized according to the type of models formed. There are four main categories of the SVM being utilized in credit scoring: *standard SVM and its variants*, *modified SVM*, *hybrid SVM*, and *ensemble models*. On the other hand, MA approaches applied in credit scoring can be divided into three categories: *standard MA*, *hybrid MA-MA*, and *hybrid MA-DM*. The models development and main issues addressed are discussed.

4.1.1. SVM

(1) Standard SVM and Its Variants. SVMs applied to build credit scoring models discussed here are the standard SVM and its variants, where these SVMs are being applied directly for model building without any modifications.

Investigation of Predictive Ability. The predictive ability of SVMs credit models is examined through two main approaches, i.e., comparative studies [8, 11, 12, 18, 26] and application on specific credit domain [19–25, 27].

Various state-of-the-art methods ranging from traditional statistical methods to nonparametric DM methods have been attempted to form credit scoring models in different researches. This prompted the necessity of a benchmark study conducted by Baesens et al. [8] in 2003. In their study, SVM and Least Squares SVM (LS-SVM) are first utilized to develop credit scoring model, and their performances are compared with the other state-of-the-art methods across eight different datasets. The methods are from the family of LOGIT, DA, Linear Programming (LP), Bayesian Network (BN), k-nearest neighbourhood (kNN), neural networks (NN), and decision trees (TREE). LS-SVM and NN reported statistically significant better results compared to the other models. Their results also showed that the methods are competitive to each other. Hence, SVM has been actively researched in the credit scoring domain.

Lessmann et al. [11] updated Baesens et al. [8] research by including comparison with more individual classifiers, ensembles, and hybrid methods. They commented the increasing research on credit scoring domain that urged for necessity to update the benchmark study where not only state-of-the-art classifiers but also advanced techniques like ensembles should be included as well. Thus, there are a total of 41 classifiers (from the family of BN, TREE, LOGIT, DA, NN, kNN, Extreme Learning Machine (ELM), and ensembles) being investigated across six datasets with wide range of size in the benchmark study. The experimental results showed that ensemble models took up top 10 ranking among all the 41 classifiers. For individual classifiers, NN showed highest ranking. Linear and RBF SVM were investigated; both SVM models showed similar performance as both scored similar ranking. They also pointed out the importance of developing business-valued scorecard, which should be taken into consideration in proposing new models.

Louzada et al. [12] conducted a systematic review on classification methods applied on credit scoring, covering research papers of year 1992-2015. Their review discussed several aspects in the credit scoring domain: objective of the study, comparison done in the research, dataset used for model building, data splitting techniques, performance measures used, application of missing data imputation, and application of feature selection. They carried out experiment on 11 main classification techniques (Linear Regression (LR), NN, TREE, DA, LOGIT, FUZZY, BN, SVM, Genetic Programming (GP), hybrid, and ensembles), where inclusion of ensembles started to receive attention as recommended in [11]. The problem of imbalance dataset is investigated and SVM showed stability in dealing with imbalance dataset. Generally, SVM has better performance and lower computational effort in the experiment.

The most recent comparative study is contributed by Boughaci and Alkhawaldeh [18]. They investigated performances of 11 machine learning techniques (from the family of kNN, BN, NN, TREE, SVM, LOGIT, and ensembles) across eight different datasets. A main difference of their study from the previous one is that the datasets utilized involved a mixture of application scoring and bankruptcy prediction dataset. The experimental results did not suggest a winner among the techniques investigated as there are no consistent

outperformances across all the datasets. They concluded BN and the Boosting ensemble as generally effective method as credit models.

These comparative studies have included a wide variety of classification techniques to formulate credit scoring models and their predictive ability is assessed across various credit datasets. However, among these credit models, there has not been a clear best technique for the credit scoring domain. Nonetheless, these comparative studies have provided a guideline on the update of latest available models in credit scoring, where SVM is initially being tagged as state of the art in [8] and then served as an important model in this domain. Besides, ensembles are another state of the art as demonstrated in [11].

Instead of comparative studies with various techniques being benchmarked, a SVM-focused study is conducted by Danenas et al. [26], particularly on bankruptcy prediction dataset. Various types of SVM originating from different libraries and software, i.e., LIBSVM, LIBLINEAR, WEKA, and LIBCVM, are included in their experiment. The SVM classifiers ranged from linear to nonlinear with a wide variety of kernels. Models are investigated on original dataset and reduced dataset already preprocessed with feature selection technique. Comparing the accuracy, different types of SVM classifiers showed comparable results, with application on reduced dataset having higher accuracy. Besides, in terms of computational effort, linear SVM is the fastest model.

Another approach to investigate the predictive ability is via the application on specific domains. The specific domains experimented are multiclass corporate rating [19, 20, 22, 23, 25], application scoring [21, 24], and behavioural scoring [27]. For multiclass corporate rating, credit scoring models are formed with LS-SVM in [19, 22] and SVM in [20, 23, 25] with the main aim to show the effectiveness of SVM in building corporate rating models. Van et al. [19] and Lai et al. [22] showed that LS-SVM is the best performing credit model compared to other traditional techniques when applied on Bankscope and England financial service data, respectively.

Huang et al. [20] tested SVM on Taiwan and US market bond rating data. They also conducted cross market analysis with variable contribution analysis from NN models. Lee [23] utilized SVM for Korea corporate credit rating analysis. Kim and Sohn [25] in 2010 have initiated the building of credit scoring model for small medium enterprises (SME). They focused on Korea SME and the explanatory variables included four main aspects: SME characteristics, technology evaluations, financial ratios, and economic indicators. They believed that SVM model would be suitable to be used for technology-based SMEs. For all the SVM models on these four researches on different market data of corporate rating, SVM have outperformed the other methods in every experiment.

For application scoring, Li et al. [21] compared SVM with NN on Taiwan bank consumer loan credit scoring. SVM reported higher accuracy than NN and the results is statistically significant. Besides, they also experimented the effect of different hyperparameter values of SVM on the type I error and type II error, i.e., misclassification error. They

demonstrated the effect on misclassification error across the hyperparameter values range can serve as a visualization tool for credit scoring model. Thus, they concluded SVM outperformed NN in terms of visualization. Bellotti and Crook [24] tested SVM of different kernels on large credit card datasets. Comparison with traditional statistical models reported only kNN and polynomial SVM have poorer results which may be due to overfitting. They suggested using support vectors' weight as an alternative to select significant features and compared the selected features with those from LOGIT's Wald statistics. The experiment indicated SVM models are suitable for feature selection for building credit scoring model.

For behavioural scoring, South African unsecured lending market is investigated by Mushava and Murray [27]. Despite having the aim to show the effectiveness of SVM, this study aimed to examine some extensions of statistical LOGIT and DA that have been less explored in credit scoring domain, with SVM being included as a benchmark comparison in their study. In their fixed window experiment, Generalized Additive Models have outperformed the others. Although SVM did not show superior performance, the inclusion of SVM in this study once again indicated SVM is perceived as a standard to overcome in credit scoring domain.

(2) *Modified SVM*. Modified SVM involved algorithmic change in the formulation of SVM. There are a few works that proposed modified SVM for solving different problems in the credit scoring domain, particularly in application scoring. The modifications required changes in the quadratic programming formulation of the original SVM.

Outlier Handling. Wang et al. [28] proposed a bilateral fuzzy SVM (B-FSVM). The method is inspired from the idea that no customer is absolutely good or bad customer as it is always possible for a classified good customer to default and vice versa. Thus, they utilized membership concept in fuzzy method where each instance in the training set takes positive and negative classes, but with different memberships. This resulted in bilateral weighting of the instances because each instance now has to take into account error from both classes. By including the memberships from fuzzy, SVM algorithm is reformulated to form B-FSVM. They used LR, LOGIT, and NN to generate membership function. B-FSVM are compared with unilateral fuzzy SVM (U-FSVM), and other standard methods. Linear, RBF, and polynomial kernels are used to form B-FSVM, U-FSVM, and SVM models.

Computational Efficiency. Harris [29] introduced clustered SVM (CSVM), a method proposed by Gu and Han [92], into credit scoring model. The research aimed to reduce computational time of SVM model. With k-means clustering to form clusters, these clusters will be included into the formulation of SVM optimization problem, changing the original SVM algorithm. Two CSVM models are developed with linear and RBF kernel and compared with LOGIT, SVM, and their hybrids with k-means. Excellent time improvement of CSVM is reported.

Dynamic Scoring. Yang [30] modified weighted SVM model to become a dynamic scoring model. The main idea is to enable an easy update of credit scoring model without the need to repeat variable selection process when new customers data became available. Original kernel in weighted SVM is modified to become an adaptive kernel. When there is increment in the data size, adaptive kernel can automatically update the optimal solution. Besides, with the trained model, Yang [30] suggested an attribute ranking measure to rank the kernel attributes. Thus, this became an alternative solution for the black box property of SVM.

Reject Inference. The aim of reject inference is to include rejected instances into model training, then improving classification performance. Li et al. [31] and Tian et al. [32] proposed new SVM to solve reject inference problem for online peer-to-peer lending. Li et al. [31] proposed a semisupervised L2-SVM (SSVM) to solve reject inference for a peer-to-peer lending data from Lending Club of different years. Thus, in the SSVM formulation, unlabelled rejected instances are added to the optimization constraints of SVM, converting the original quadratic programming problem to a mixed integer programming. SSVM reported better performance than other standard methods. Tian et al. [32] proposed a kernel-free fuzzy quadratic surface SVM (FQSVM). The main advantages of the proposed model are the ability to detect outliers, extract information from rejected customers, no kernel required for computation, and efficient convex optimization problem. The proposed model is benchmarked against other reject inference methods. FQSVM is reported to be superior than SSVM proposed by [31] in terms of several performance measures as well as computational efficiency.

Features Selection. There are two new formulation of SVM to carry out features selection in cost approach [33] and profit approach [34] on Chilean small and microcompanies. The dataset consisted of new and returning customers, indicating the credit scoring models formed involved both application and behavioural scoring. Maldonado et al. [33] included variable acquisition costs into formulation of SVM credit scoring model to do feature selection. Similar to Li et al. [21], they also added additional constraints into the optimization problem, converting it to a mixed integer programming problem, but the added constraints are the variable acquisition costs. They proposed two models where 1-norm and LP-norm SVM are both modified with the additional constraints, forming two new credit scoring models, namely, L1-mixed integer SVM (L1-MISVM) and LP-mixed integer SVM (LP-MISVM). Due to the ability of the proposed models to take into consideration of variable acquisition costs and good performance simultaneously, it is believed that the proposed methods are efficient tool for credit risk as well as business analytics.

On the other hand, Maldonado et al. [34] introduced a profit-based framework to do feature selection and classification with modified SVM as well as profit performance metrics. Instead of considering acquisition costs for variables, one by one, they treated the costs as a group to be penalized on whole set of variables. Therefore, the L_{∞} norm is

utilized to penalize the group cost. Two models are proposed with 1-norm SVM and standard SVM being modified by including L_{∞} into optimization objective function, forming LIL- ∞ SVM and L2L- ∞ SVM. Their proposed models are effective in selecting features with lower acquisition costs, yet maintaining good performance. The main difference as compared to the previous research in [33] is that, with this newly proposed model, the profit can be assessed, which posed as a crucial insight for business decision makers.

(3) *Hybrid SVM*. Hybrid SVM credit scoring models have been developed by collaborating SVM with other techniques for different purposes.

Reject Inference. Chen et al. [48] tackled this problem using the credit card data from a local bank of China. They hybridized k-means clustering with SVM, formulating a two-stage procedure. The first stage is the clustering stage where new and accepted customers are grouped homogeneously, isolated customers are deleted, and inconsistent customers are relabelled. The clustering procedure of dealing with inconsistent customer is a type of reject inference problem. These clustered customers from the first stage are input to SVM to do classification in the second stage. Instead of classifying customers into binary groups, they attempted to classify into three and four groups. Different cutoff points are also set for different groups. They believed the proposed method is able to provide more insight for risk management.

Rule Extraction. Black box property of SVM has always been the main weakness, which is also a main concern for practitioners not using SVM as credit scoring models. Martens et al. [36] proposed rule extraction technique to be used together with SVM. Three different rule extraction techniques, namely, C4.5, Trepan, and Genetic Rule Extraction (G-REX), are hybridized with SVM. Experiment is conducted on different fields that require comprehensibility of model where credit scoring is one of the field addressed in their research. The proposed models are advantageous in giving clear rules for interpretation. In 2008, Martens et al. [93] made an overview on the rule extraction issue, where the importance of comprehensibility in credit scoring domain is addressed again. Zhang et al. [38] proposed a hybrid credit scoring model (HCSM) which hybridized Genetic Programming (GP) with SVM. The main advantage of the proposed technique is the ability to extract rules with GP that solved the black box nature of SVM.

Computational Efficiency. Hens and Tiwari [46] integrated the use of stratified sampling method to form SVM model. Then, with the smaller sample, F-score approach is used to do feature selection to compute the rank of features based on importance. The proposed model achieved lowest runtime and comparable performance when compared with other methods considered in their experiment.

Features Extraction. Xu et al. [40] incorporated link analysis with SVM, where link analysis is first used to extract applicants' information and then input into SVM to do

classification. The proposed method is a two-phase procedure. In the first phase, they introduced to use three different algorithms of link relation to extract input features by linking the relations of applicants' data. So, there are three hybrid SVM models being built with respect to the three different algorithms. Recently, Han et al. [50] proposed a new orthogonal dimension reduction (ODR) method to do feature extraction. They used SVM as the main classifier as they believed ODR is an effective preprocessing for SVM and used LOGIT as benchmark classifier. There are three main parts in the experiment. First, they discovered that variables normalization posed large effect on classification performance of SVM but LOGIT is not strongly affected. Therefore, normalization is applied for all models. Second, comparison is done with existing feature reduction method, principal component analysis (PCA). Third, they suggested using LOGIT at the start to pick important variables with Wald statistics and then only extract features from the reduced variables, which they name as HLG. They concluded ODR is effective in solving dimension curse for SVM.

Features Selection. For feature selection hybrid SVM models, there are filter approach [37, 39, 43] and wrapper approach [54, 55] used. For filter approach, rough set theory is employed by Zhou and Bai [37] and Yao et al. [39], where rough set select input features in the first stage and carried out classification tasks in the second stage with respect to the hybridized techniques. There are three main differences in between the hybrid models collaborated with rough sets proposed in these two researches. First, Zhou and Bai [37] specified their study on Construction Bank in China whereas Yao et al. [39] study is generally on public datasets. Second, features selection is based on information function in [37] whereas computed variable importance is used to select features in [39]. Third, the hybrid models developed in [37] are hybridization with NN, SVM, and GA-SVM (GA to tune SVM hyperparameters) whereas the hybrid models developed in [39] are hybridization with SVM, TREE, and NN. Both experiments showed that SVM-based hybrid models obtained best performance. Chen and Li [43] also adapted a filter approach to do feature selection. They proposed four different filter approaches: LDA, TREE, rough set, and F-score. These four approaches are hybridized with SVM. In order for the models to be comparable, the same number of features is selected based on the four approaches based on variable importance.

Apart from the filter approach of conducting feature selection, Jadhav et al. [54] and Wang et al. [55] incorporated filter techniques in developing novel wrapper model for feature selection task. The main concept is to guide the wrapper model to do feature selection with obtained information from filter feature ranking techniques.

Jadhav et al. [54] proposed Information Gain Directed Feature Selection (IGDFS) with wrapper GA model, based on three main classifiers, i.e., SVM, kNN, and NB. Top rank features from information gain are passed on to the wrapper GA models. They compared their three different IFDFS (based on three different classifiers) with six other models: three standalone classifiers with all features included,

and three standard GA wrapper models of the classifiers that conducted feature selections without guided by any filter methods. Wang et al. [55] hybridized SVM with multiple population GA to form a wrapper model for feature selection. The method has a two-stage procedure. In the first stage, they utilized three filter feature selection approaches to find prior information of the features. Feature importance is sorted in descending order, then a wrapper approach is used to find optimal subset. With the three feature subsets from the three approaches and probability of a feature to be switched on, they formed the initial populations to be passed on to the second phase. In the second phase, HMPGA with SVM is run to find the final optimal feature subset; thus HMPGA-SVM is the model with prior information.

Hyperparameters Tuning. Other than input features, hyperparameters of SVM models pose great effect on the end model formed. Previous works that proposed models for feature selection have applied the conventional Grid Search (GS) method to find the appropriate hyperparameters. The researches that introduced hybrid SVM for finding hyperparameters are [47, 52] in bankruptcy prediction, [42, 44, 51] in application scoring, and [49, 53] in corporate rating.

With the success of linear SVM as experimented in [26], Danenas and Garsva [47] examined the use of different linear SVMs available in the LIBLINEAR package on bankruptcy prediction. All the techniques are hybridized with GA and PSO to do model selection and hyperparameters tuning. The hybrid models formed are, namely, GA-linSVM and PSO-linSVM. Sliding window approach is adapted for building models across different time periods to report model performances. GA-linSVM is concluded to be more stable than PSO-linSVM by consistently selecting same model across different time periods yet having good performance. In later years, Danenas and Garsva [52] conducted another research to improve on PSO-linSVM. They modified PSO by using integer values for velocity and position, instead of rounding up the values as in [47, 51].

PSO-linSVM continued to receive attention by Garsva and Danenas [51] in application scoring. Similarly, they carried out model selection and hyperparameters tuning with PSO-linSVM but with a mixed search space for PSO, which is a slight modification as compared to previous work [47]. Comparison is done with SVM and LS-SVM (of different kernels), of which the hyperparameters are tuned with PSO, Direct Search (DS), and SA, respectively. To address the data imbalance problem, they investigated the use of True Positive Rate (TPR) and accuracy as the fitness function. TPR is concluded as appropriate fitness function for imbalance dataset.

Zhou et al. [42] presented Direct Search (DS) to tune LS-SVM hyperparameters. They compared DS with GA, GS, and Design of Experiment (DOE). Among the four hybrid models, DS-LS-SVM is the recommended approach due to its best performance. Yu et al. [44] conducted a similar experimental setup with [42] where DS, GA, GS, and DOE are presented for hyperparameters tuning. The main difference is that they considered class imbalance problem, so the model tuned is weighted LS-SVM (WLS-SVM), where different

weights can be assigned to solve class imbalance during classification. Different from [42], they recommended the use of DOE for hyperparameters tuning due to competitive results reported as compared to DS, GA, and GS, but with lowest computational time. External benchmarked against results from [8, 45] are also included.

Chen et al. [49] and Hsu et al. [53] integrated ABC and SVM for hyperparameters tuning in corporate credit scoring. Chen et al. [49] applied the proposed ABC-SVM on Compustat database of America from year 2001-2008. PCA is the data preprocessing method used for extracting important features. Recently, Hsu et al. [53] also researched on ABC-SVM in corporate credit rating with dataset from the same database as in [49], but including more recent years 2001-2010. Similarly, they utilized PCA as the preprocessing step. They conducted a more detailed study on the data, where using information from PCA, they divided the dataset into three categories to study the ability of the credit models to account for changes in future credit rating trend.

Simultaneous Hyperparameters Tuning and Features Selection. Based on the previous discussed research works, feature selection and hyperparameter selection for SVM models in credit scoring are crucial procedures in model building. Therefore, two pieces of research [35, 41] aimed to solve both problems simultaneously with wrapper model approach.

Huang et al. [35] attempted three strategies for building SVM-based credit scoring models: first, GS to tune SVM hyperparameters with all features included; second, GS to find hyperparameters and F-score to find feature subsets for SVM model building; third, the initiation of hybrid GA-SVM to search hyperparameters and feature subsets simultaneously. This experimental result indicated GA-SVM as a suitable tool for alternative to solve both issues together but required high computational effort. Zhou et al. [41] also formulated a hybrid model using GA, but using different variants of weighted SVM, thus forming GA-WSVM. They mentioned in Huang et al. [35] research that the features found did not carry importance of the selected features. Thus, they proposed feature weighting as one of the addition procedure in the wrapper approach. The proposed model aimed to search for hyperparameters of WSVM as well as feature subsets with feature weighting. They compared the feature weighting method with t-test and entropy based method.

(4) *Ensemble Models.* The two main types of ensemble models are homogeneous (combining same classifiers) and heterogeneous (combining different classifiers). In credit scoring domain, [56-58] worked on homogeneous ensembles while Xia et al. [59] worked on heterogeneous ensembles.

Improve Predictive Ability. Zhou et al. [56] pointed out inductive bias problem of single classifier when using fixed training samples and parameter settings. Therefore, they introduced ensemble model based on LS-SVM to reduce bias for credit scoring model. The two main categories of ensemble strategies introduced are the reliability-based and weight-based. There are three techniques, respectively, for each category,

resulting in a total of six LS-SVM-based ensemble models being formed. Another research that proposed ensemble model is by Ghodselahi [57]. They recommended using fuzzy C-means clustering to preprocess the data before fed into SVM. Then, 10 of the hybrid SVM base models formed the ensemble models, using membership degree method to fuse all the base models as the final ensemble results. Xia et al. [59] introduced a new technique named as bstacking. The main idea is to pool models and fuse the end results in a single step. Four classifiers are used, i.e., SVM, Random Forest (RF), XGBoost, and Gaussian Process Classifier (GPC), as the base learners due to their accuracy and efficiency.

Data Imbalance. Yu et al. [58] developed Deep Belief Network SVM-based (DBN-SVM) ensemble model in a different approach where the main aim is to solve dataset imbalance problem. Their model has a three-stage procedure. In the first stage, data is partitioned into various training subsets with bagging algorithm, and each subset is resampled to rebalance the instances. In the second stage, SVM classifiers are trained on the rebalanced training subsets. In the third stage, DBN is employed to fuse the final results. Proposed method is compared with SVM and ensemble SVM with majority voting.

4.1.2. MA

(1) *Standard MA.* Standard MA being attempted to form credit scoring models are GA, GP, SA, ACO, and CS. The credit scoring problem is formulated with these MA as an optimization problem to be solved with respect to the objective functions.

Investigation of Predictive Ability. The predictive ability of MA credit models has been tested through application on specific credit domains, i.e., application scoring [60, 62, 63, 65] and bankruptcy prediction [68]. The experiments of Desai et al. [60], Finlay [62], and Abdou [65] are specific study on a particular country which are credit unions in Southeastern US, large data of UK credit application, and Egyptian public sector bank, respectively. In contrast, Cai et al. [63] and Yang et al. [68] conducted a general study based on the public German dataset and simulated database of 20 companies, respectively.

Back in 1997, Desai et al. [60] investigated the predictive ability of AI techniques by comparing with traditional credit models LOGIT and DA. The AI techniques studied are three variants of NN and GA. They classified the credit customers into three classes (good, poor, and bad) instead of the usual binary (good and bad). GA is used for discriminant analysis. With the aim to minimize the number of misclassification, an integer problem is formulated with GA to make it acting equivalently to branch-and-bound method, then, solving the dual problem gives the final separating hyperplane. Finlay [62] pointed out the advantage of developing credit scoring model with GA due to its ability to form model based on self-decide objective function. They proposed to build a linear scoring model with GA that maximizes the GINI coefficient. Large dataset of UK credit application is experimented. Four

different GA-derived models are developed which consider seeding (prior information from LOGIT as initial solution) and different encoding scheme (integer or binary).

Cai et al. [63] and Abdou [65] have included misclassification costs into their studies. Cai et al. [63] established credit scoring model with GA. The optimization problem is a linear scoring problem as in [62]. They computed the appropriate cutoff as the weighted average of good and bad customers critical values. A fitness function is formed that considered all components in the confusion matrix together with the associated misclassification costs. Abdou [65] compared performance of GP with profit analysis and weight-of-evidence (WOE) model. Two types of GP are examined here: single program GP and team model GP, which is a combination of single program GP for better results. They conducted sensitivity analysis based on different misclassification ratios and emphasized the importance to evaluate scorecard with misclassification costs. Only Yang et al. [68] addressed the bankruptcy prediction issue with CS. The authors developed a CS model which used Lévy's flight to generate new solution.

Rules Extraction. There are several researches focused on rules extraction with different MA which are GP in [61, 64], SA in [66], and ACO in [67, 69]. Ong et al. [61] recommended using GP as an alternative to form credit models. GP undergone the same procedures as in GA to search for solution, but the main difference is that GP generates rules to do classification. The authors concluded several advantages of GP to build credit scorecard: nonparametric that is suitable for any dataset, automatic discrimination function that do not require user-defined function as in NN, and better rules obtained compared to TREE and rough set that generated lower error. Huang et al. [64] proposed a 2-stage GP (2SGP). In the first stage, IF-THEN rules are derived. They formulated GP to ensure that only useful rules are extracted. Based on these rules, the dataset will be reduced by removing instances that do not satisfy any rules or satisfy more than one rule. Then, the reduced data is passed on to the second stage of GP where the discriminant function is used to classify the customers.

Dong et al. [66] established SA rule-based extraction algorithm (SAREA) for credit scoring. Similar to the previous rule extraction with GP in [64], the proposed SAREA is also in two-phase, and the extracted rules are the IF-THEN rules. In the first phase, SA is run on initial rules and their corresponding best accuracy rules are put into the final rule set of first phase. The best rule from the first phase is required for fitness function computation in second phase to penalize the accuracy. In the second phase, SA is run again on random initial rules to find their corresponding best accuracy rules to form the final rule set, but the fitness computation will be the penalize accuracy based on the best rule from first phase.

The importance of comprehensibility for credit scoring model had been pointed out [36, 93] as discussed in Section 4.1.1(2). Martens et al. [67] researched on establishing a novel model that has good performance for both accuracy and comprehensibility. They introduced ACO algorithm in AntMiner+ as the potential credit scoring model. The rules have high comprehensibility which is crucial for business

decisions and they also analysed the extracted rules to be integrated with Basel II accord. Recently, ACO-based classification rule induction (CRI) framework is introduced by Uthayakumar et al. [69]. They carried out their experiment on both qualitative and quantitative datasets, focusing on bankruptcy prediction. ACO algorithm is modified based on the concept of rule induction. Due to the ability of ACO to provide better results in CRI, reducing rules complexity and effective classification of abundant data, ACO is recommended in their study.

(2) *Hybrid MA-MA*. Hybrid MA-MA involves two different MA being integrated together to form a new method. There are only two research works that have been proposed thus far.

Parameters Tuning. Jiang et al. [70] proposed the idea of using SA to optimize GA, forming hybrid SA-GA. SA is integrated into GA to update population by selecting chromosomes using Metropolis sampling concept from SA. Two variants of NN are the main classifiers used in this experiment. SA-GA is utilized to search the input weight vector of the combined NN classifiers.

Rules Extraction. Aliehyaei and Khan [71] presented a hybrid model of ACO and GP with a two-step task. ACO is responsible for searching for rule sets from the training set. The rules extracted from ACO is then fed into GP for classification.

(3) *Hybrid MA-DM*. Hybrid MA-DM methods include the usage of a MA technique to assist the classification task of DM classifier, thus improving model performance.

Rules Extraction. Past researches that aimed to do rules extraction are presented in [38, 79] which utilized SA and GP, respectively. Zhang et al. [38] proposed a hybrid credit scoring model (HCSM) which is a 2-stage model incorporating GP and SVM. In the first stage, GP is used to extract rules with misclassification of type I error and type II error taken as the fitness function. In the second stage, SVM is used as the discriminant function to classify the customers. Jiang [79] incorporated SA and TREE as a new credit scoring model. Rules from TREE are input as initial candidate for SA, then SA produced new decision rules for classification. They formed three TREE-SA credit models with different discriminant function that account for type I error and type II error, which is similar to Zhang et al. [38].

Parameters and Hyperparameters Tuning. Generally, parameters and hyperparameters have significant effect on model performance. The difference between the two is that, parameters are involved in model training, where the value can be evaluated by the model, whereas hyperparameters are completely user-defined where its value cannot be evaluated by the model. Therefore, tuning appropriate values for parameters and hyperparameters are crucial in model building for credit scorecard. Metaheuristics is applied for parameters tuning of NN input weight and bias [77], hyperparameters

tuning of NN [75, 84], and hyperparameters tuning of SVM [37, 42, 44, 47, 49, 51–53].

For NN model tuning, Wang et al. [77] hybridized GA with NN to tune the input weight and bias parameters. They employed real-valued encoding for GA, using arithmetic crossover and nonuniform mutation and concluded that tuning parameters with GA improved learning ability of NN. On the other hand, Lacerda et al. [75] established GA to tune hyperparameters in NN. They proposed a modified GA based on consideration of redundancy, legality, completeness and causality. Training samples are clustered and they introduced cluster crossover algorithm for GA. Then, they utilized the proposed GA to form a multiobjective GA in seeking for NN hyperparameters. Correa and Gonzalez [84] presented two hybrid models, GA and binary PSO (BPSO) for hyperparameters tuning in NN. For both MA techniques, cost of the candidate solutions is computed before proceeding to the searching process. They presented a different approach in examining their proposed models, where they conducted cost study on three different scoring models, i.e., application scoring, behavioural scoring, and collection scoring.

Several models that utilized GA to tune SVM hyperparameters [37, 42, 44] have been discussed in Section 4.1.1(3). In Zhou and Bai [37] experiment, the proposed model worked best with GA tuned SVM. In Zhou et al. [42] experiment, investigation on SVM hyperparameters tuning is conducted on DS, GS, DOE, and GA. In Yu et al. [44] experiment, they investigated LS-SVM hyperparameters tuning also with DS, GS, DOE, and GA. From these three researches, it can be observed that they included GA tuned SVM-based classifiers into their experiments for comparison, implying that EA is a good alternative for hyperparameters tuning in SVM. Model selection and hyperparameters tuning on linear SVM from LIBLINEAR using GA and PSO have been investigated in a few studies [47, 51, 52], as discussed in Section 4.1.1(3) GA-linSVM showed better performance than PSO-linSVM on bankruptcy prediction [47]. Then, Garva and Danenas [51] further experimented on modified PSO to form PSO-linSVM on application scoring. PSO is further modified by Danenas and Garsva [52], forming a different version of PSO-linSVM to build bankruptcy prediction model. There are also two models that tuned SVM with ABC [49, 53] as discussed in Section 4.1.1(3). Both researches focused on corporate credit rating problem. Chen et al. [49] first formulated ABC-SVM then followed by a more detailed study in Hsu et al. [53]. Both studies reached the same conclusion with ABC-SVM being an effective method to tune hyperparameters and Hsu et al. [53] indicated that ABC-SVM is also able to capture changing trend of credit rating prediction.

Features Extraction. Fogarty and Ireson [72] hybridized the GA and BN. GA is utilized to optimize BN by selecting categories and attributes combinations from training data using cooccurrence matrix. The attributes combinations generated are analogous to extracted features. Liu et al. [76] designed GP to extract features by selecting derived characteristics, which are attribute combinations but determined with analysis method and human communication. To ensure the derived characteristics are practical, the characteristics are generated

with GP by maximizing information value together with application of human communication. Linear DA model is then built using these derived characteristics. Zhang et al. [78] formed hybrid model by incorporating GA, k-means clustering, and TREE together. The GA is introduced to do attribute reduction, which is a kind of feature extraction. Binary encoding is applied and the candidate solutions in GA consist of breakpoints. Then k-means clustering is assigned to remove noise and TREE to do classification.

Features Selection. Hybrid models of MA-DM developed for features selection are [54, 55, 74, 81, 83, 86, 89, 90]. GA from EA category is the most popular method to be hybridized with DM classifiers [54, 55, 81, 83, 86, 89] for solving feature selection in credit scoring domain. All the hybrid MA-DM for feature selection are based on wrapper approach except for [87–89] which presented filter approach.

Drezner et al. [74] constructed a new credit scoring model by incorporating TS with LR for feature selection, focusing on bankruptcy prediction. First, the selected feature subset from TS is compared with a known optimal subset, subset from stepwise regression, and subset from maximum R^2 improvement. The TS-feature subset is reported to be very competitive in selecting a near-optimal feature subset. Sustersic et al. [81] introduced GA to do feature selection with Kohonen subsets and random subsets. PCA is the benchmark feature selection method compared with the two GA-generated features. NN and LOGIT models are developed with the features generated. Their experiment is specified on a Slovenian bank loan data. The authors also discussed the effect of setting cutoff point on the change in type I error and type II error. Huang and Wu [83] examined the effectiveness of GA for feature selection. GA is wrapped with kNN, BN, LOGIT, SVM, NN, decision tables, and three different ensembles to carry out feature selection task. In the first part, standalone classifiers (without wrapped with GA) are compared. In the second part, GA-based feature selection is compared with four filter feature selection techniques, i.e., chi-squared statistics, information gain, ReliefF, and symmetrical uncertainty. In the third part, every standalone classifier is compared with their GA-wrapped counterpart.

Some proposed wrapper models included filter techniques to gain useful feature information for improving the standard wrapper approach. This type of wrapper model has been presented by [54, 55, 86]. Oreski and Oreski [86] tackled feature selection problem by proposing hybrid GA with NN. They incorporated four different filter techniques to develop the wrapper GA-NN. With the feature ranking from the filter methods, three main procedures are infused into the GA-NN. The three procedures are to reduce search space using the reduced features from the filter ranking, refine the search space with GA, and induce diversity in the initial population with incremental stage using GA. There are two other wrapper models for solving feature selection problem in credit modelling presented by Jadhav et al. [54] and Wang et al. [55] as discussed in Section 4.1.1(3). Similarly, both researches formed novel wrapper models with filter information. Jadhav et al. [54] formulated wrapper-based SVM, kNN, and NB models with GA, incorporating

with information gain. On the other hand, for Wang et al. [55], the idea is to formulate a wrapper-based SVM model with multiple population GA. They incorporated information from different filter techniques to be input as initial solutions for the multiple population GA.

Marinakis et al. [80] formed wrapper model with ACO and PSO. The kNN and its variants (1-NN and weighted kNN) are being wrapped to do feature selection and classification. The experiment is on multiclass problem using two datasets from a UK nonfinancial firm where the first dataset is to do credit risk modelling and the second dataset is for audit qualification. Later, Marinaki et al. [82] conducted a research with similar setup as in [80] and used the first dataset as in [80]. They proposed a different metaheuristics which is HBMO to wrap the kNN and its variants. A music-inspired SI technique, HS, is attempted by Krishnaveni et al. [90] recently to form a wrapper model with a kNN variant, i.e., 1-NN for feature selection. Besides, parallel computing is also attempted with the model and reported a significant time-saving with the paralleled version of the proposed method compared to other wrapper models.

Apart from wrapper approach, filter approach models have also been proven useful in the credit scoring domain. Wang et al. [87] hybridized TS with rough sets (TSRS) to search for minimal reducts that served as the reduced feature subsets to be input into classifiers. Japan dataset is experimented in the experiment. The feature subset from TSRS is fed into RBF network, SVM, and LOGIT. Later, Wang et al. [88] attempted hybridization of SS with rough sets (SSRS) for feature selection. Similar experiment setup as in [87] is conducted, but with two differences: one additional dataset is included and the classifiers included for comparison are different. Waad et al. [89] proposed another filter-based feature selection method with GA. The new method is a two-stage filter selection model. The first stage formulated an optimization problem to be solved with GA. The main idea in the first stage is to overcome selection trouble and rank aggregation problem and then sort the features according to their relevance. In the second stage, an algorithm is proposed to solve disjoint ranking problem for similar features and remove redundant features.

Features Discretization. There is only a single research contributed by Hand and Adams [73] for demonstrating a new model that does feature discretization in credit scoring. They formed collaborated SA with weight-of-evidence (WOE) and generalized WOE, forming two wrapper models. The main concept is to effectively discretize continuous attributes into appropriate intervals. The proposed SA discretization technique is compared with quantile partitioning and equal interval division.

Simultaneous Hyperparameters Tuning and Features Selection. The importance of hyperparameters tuning and feature selection has urged some researchers to resolve both problems simultaneously [35, 41, 85]. Huang et al. [35] and Zhou et al. [41] developed GA-based wrapper model together with SVM and LS-SVM respectively. Both studies have been discussed in Section 4.1.1(3) and their research results have indicated

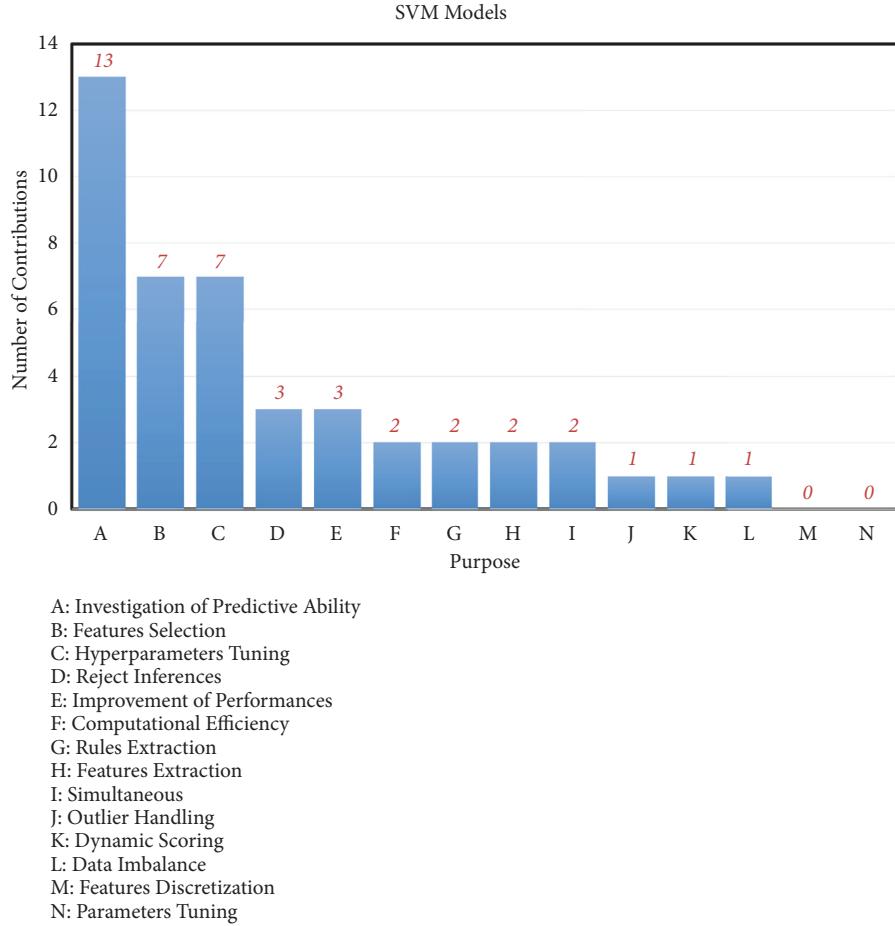


FIGURE 1: Summary of SVM models.

GA wrapper can solve both problems effectively. Oreski et al. [85] also solved hyperparameters tuning and feature selection simultaneously but with NN as the main classifier. The research is conducted on a Croatian bank credit applicant dataset. The proposed GA-based feature selection wrapped with NN (GA-NN) is benchmarked against other feature selection techniques, i.e., forward selection, information gain, gain ratio, GINI index, correlation, and voting. For hyperparameters tuning, the authors proposed Neural Network Generic Model (NNGM) which employed GA to tune hyperparameters in NN model. The features generated from the different methods are passed on to NNGM to do classification. They also examined the effect of different cutoff points on accuracy and study different misclassification ratios.

4.1.3. Summary

(1) **SVM Models.** Developments of SVM models are summarized in the upcoming tables and figures. Table 1 arranges all the reviewed studies in chronological order to show the development trend, summarizes the addressed issues, and provides additional information on the SVM type with respect to the kernel used as well as details of the other

models considered in each experiment for comparison. Table 2 reports the counts of papers categorized by the types of models. Figure 1 illustrates the categorization of all the SVM models based on the purposes.

From Table 1, early stage of credit models with SVM are basically standalone SVMs for investigation of the predictive ability. As a result from these investigative experiments that validated the effectiveness of SVM, it is soon labelled as one of the state-of-the-art credit scoring methods. Then, the development trend shifted to enhance the original SVM models, where hybrid models formulation is the most popular approach that remains active until recent years, with data preprocessing and hyperparameters tuning outnumbered the other research purposes. Ensemble models are the latest research trend in credit scoring due to its ability to improve classification performances. This leads to involvement of SVM in two situations; i.e., SVM is one of the benchmark models against ensembles and SVM is the base classifier used to form new ensembles. On the other hand, in view of the SVM type in credit modelling, standard SVM has been most frequently used while SVM variants have apparently lesser research works. In view of the kernel used, linear and RBF kernel have been widely utilized in this domain.

TABLE 1: Summary of literature for SVM models.

Authors	SVM type	Kernel	Other Methods	Remarks
<i>Standard SVM and Variants</i>				
Baesens et al. (2003) [8] Van Gestel et al. (2003) [19]	SVM, LS-SVM LS-SVM	linear, RBF RBF	-	(i) compare with LOGIT, DA, kNN, LP, BN, NN, TREE (i) multiclass corporate rating (ii) compare with LR, LOGIT, NN
Huang et al. (2004) [20]	SVM	linear, RBF, polynomial	-	(i) multiclass corporate rating (ii) compare with NN
Li et al. (2006) [21]	SVM	RBF	-	(i) compare with NN (ii) study misclassification error
Lai et al. (2006) [22]	LS-SVM, SVM	RBF	-	(i) multiclass corporate rating (ii) compare with NN, LR, LOGIT
Lee et al. (2007) [23]	SVM	RBF	-	(i) multiclass corporate rating (ii) compare with NN, DA, CBR
Bellotti and Crook (2009) [24]	SVM	linear, RBF, polynomial	-	(i) application on large dataset (ii) compare with NN, DA, CBR (iii) support vector weights to select significant features
Kim and Sohn (2010) [25]	SVM	RBF	-	(i) multiclass corporate rating on SME (ii) compare with NN, LOGIT
Danenas et al. (2011) [26]	SVM	linear, RBF, polynomial, Laplacian, Pearson, inverse distance, inverse square distance	-	(i) compare between SVMs of different libraries (LIBLINEAR, LIBSVM, WEKA, LIBCVM)
Lessmann et.al (2015) [11]	SVM	linear, RBF	-	(i) compare with LOGIT, TREE, ELM, kNN, DA, BN, ensembles (ii) recommendation to use different performance measures
Louzada et.al (2016) [12]	SVM	not mentioned	-	(i) compare with LR, NN, TREE, DA, LOGIT, FUZZY, BN, SVM, GP, hybrid, ensembles (ii) study class imbalance problem
Boughaci and Alkhawaldeh (2018) [18]	SVM	not mentioned	-	(i) compare with kNN, BN, NN, TREE, SVM, LOGIT and ensembles

TABLE 1: Continued.

Authors	SVM type	Kernel	Other Methods	Remarks
Mushayva and Murray (2018) [27]	SVM	RBF	-	(i) compare with LOGIT, DA, extensions of LOGIT and DA, ensembles
<i>Modified SVM</i>				
Wang et al (2005) [28]	SVM	linear, RBF, polynomial	fuzzy membership	(i) introduce bilateral weighting error into classification problem (ii) compare with U-FSVM, SVM, LR, LOGIT and NN
Harris (2015) [29]	SVM	linear, RBF	k-means cluster	(i) reduce computational time (ii) compare with LOGIT, k-means+LOGIT, SVM, k-means+SVM
Yang (2017) [30]	WSVM	RBF, KGPFF	-	(i) dynamic scoring with adaptive kernel (ii) ranking of kernel attributes to solve black box model (iii) compare with LOGIT
Li et al. (2017) [31]	L2-SVM	not mentioned	-	(i) reject inference (ii) compare with LOGIT, SVM
Tian et al. (2018) [32]	L2-SVM	no kernel	-	(i) reduce computational time (ii) reject inference and outlier detection (iii) compare with LOGIT, kNN, SVM, SSVM
Maldonado et al. (2017) [33]	SVM, 1-norm SVM	linear	-	(i) feature selection (ii) acquisition cost into formulation of SVM (iii) application and behavioural scoring (iv) study class imbalance problem (v) compare with SVM (filter and wrapper feature selection)
Maldonado et al. (2017) [34]	SVM, LP-norm SVM	linear	-	(i) profit-based feature selection (ii) group penalty function included in SVM formulation (iii) compare with LOGIT, SVM (filter, wrapper feature selection)
<i>Hybrid SVM</i>				
Huang et al. (2007) [35]	SVM	RBF	GA	(i) hyperparameters tuning, features selection (wrapper approach) (ii) compare with GP, NN, TREE
Martens et al. (2007) [36]	SVM	RBF	C4.5, Trepan, G-REX rough sets	(i) rules extraction (ii) compare with LOGIT, SVM, TREE
Zhou and Bai (2008) [37]	SVM	RBF	GP	(i) features selection (filter approach) (ii) compare with DA, NN, SVM, SVM wrapped by GA (i) rules extraction (ii) compare with SVM, GP, LOGIT, NN, TREE
Zhang et al. (2008) [38]	SVM	RBF	neighbourhood rough set link analysis	(i) features selection (filter approach) (ii) compare with DA, LOGIT, NN (i) features extraction with link relation of applicants (ii) compare with SVM
Yao (2009) [39]	SVM	RBF		
Xu et al. (2009) [40]	SVM	RBF		

TABLE I: Continued.

Authors	SVM type	Kernel	Other Methods	Remarks
Zhou et al. (2009) [41]	WSVM	linear, RBF	GA	(i) hyperparameters tuning, features selection (wrapper approach) (ii) features weighting (iii) compare with LR, LOGIT, NN, TREE, kNN, AdaBoost
Zhou et al. (2009) [42]	LSSVM	RBF	DS, GA, GS, DOE	(i) hyperparameters tuning (wrapper approach) (ii) compare with LOGIT, kNN, DA, TREE
Chen and Li (2010) [43]	SVM	RBF	DA, TREE, rough sets, F-score	(i) features selection (filter approach) (ii) compare with SVM
Yu et al. (2010) [44]	WLS-SVM	RBF	DS, GA, GS, DOE	(i) hyperparameters tuning (wrapper approach) (ii) study class imbalance problem (iii) compare with results from [8, 45]
Hens and Tiwari (2011) [46]	SVM	linear	stratified sampling	(i) reduce computational time
Danenas and Garsva (2012) [47]	SVM from LIBLINEAR	linear	PSO, GA	(i) model selection (ii) hyperparameters tuning (wrapper approach)
Chen et al. (2012) [48]	SVM	RBF	k-means cluster cluster ABC	(i) reject inference (ii) multiclass problem with different cutoff points (i) hyperparameters tuning (wrapper approach) (ii) compare with SVM tuned with GA and PSO
Chen et al. (2013) [49]	SVM	RBF	orthogonal dimension reduction PSO, DS, SA	(i) features extraction with dimension reduction (ii) compare with LOGIT (i) model selection (ii) hyperparameters tuning (wrapper approach)
Han et al. (2013) [50]	SVM	linear	PSO	(iii) study class imbalance problem (iv) compare among all SVM and LS-SVM tuned with PSO, DS, SA
Garsva and Danenas (2014) [51]	LS-SVM, SVM from LIBLINEAR	linear, RBF, polynomial, sigmoid	PSO	(i) model selection (ii) hyperparameters tuning (wrapper approach)
Danenas and Garsva (2015) [52]	SVM from LIBLINEAR	linear	PSO	(iii) compare with LOGIT, RBF network classifier, SVM tuned with DS

TABLE I: Continued.

Authors	SVM type	Kernel	Other Methods	Remarks
Hsu et al. (2018) [53]	SVM	RBF	ABC	(i) hyperparameters tuning (wrapper approach) (ii) compare with LOGIT, SVM tuned with GS, GA and PSO
Jadhav et al. (2018) [54]	SVM	RBF	GA	(i) features selection (wrapper approach) (ii) compare with standalone SVM, kNN, NB and their wrappers with GA with standard GA
Wang et al. (2018) [55]	SVM	RBF	multiple population GA	(i) features selection (wrapper approach) (ii) compare with MPGA-SVM, GA-SVM, SVM
<i>Ensemble Model</i>				
Zhou et al. (2010) [56]	LS-SVM	RBF	fuzzy C-means	(i) homogeneous ensemble (ii) compare with ensemble and single classifiers
Ghodssehahi (2011) [57]	SVM	linear, RBF, polynomial, sigmoid	-	(i) homogeneous ensemble (ii) compare with ensemble and single classifiers
Yu et al. (2018) [58]	SVM	RBF	DBN	(i) homogeneous ensemble (ii) study class imbalance problem (iii) compare with ensemble and single classifiers
Xia et al. (2018) [59]	SVM	RBF	RF, GPC, XGBoost	(i) heterogeneous ensemble (ii) compare with ensemble and single classifiers

TABLE 2: Type of SVM models.

Type	Count
Standard SVM and its Variants	13
Modified SVM	7
Hybrid SVM	20
Ensembles SVM	4
Total	44

As reported in Table 2, hybrid SVM is the most frequently adopted approach to construct new SVM credit models. This is followed by standalone SVM, modified SVM, and ensembles. Louzada et al. [12] review study also revealed the same trend where hybrid models are most popular among researches. In hybrid models, the method hybridized with SVM acts to assist the classification task without changing the SVM algorithm. Thus, the construction of hybrid models is perceived as a direct approach. Standalone SVM comes in second place due to its recognition as the state-of-the-art technique. Its involvement in recent studies as benchmark model further consolidated its recognition in the credit domain. Modified SVM requires a complicated process to alter the SVM algorithm, thus receiving relatively lesser attention. Ensemble models are new modelling concept which have just been researched very lately, leading to the least number of contributions.

Figure 1 provides a quick summary on the research purposes handled in past researches utilizing SVM models. According to the counts of papers for each purpose, the top research purposes that take up the majority in this review study are investigation of predictive ability, features selection, and hyperparameters tuning. Frequent usage of SVM models in various types of credit datasets and involvement in benchmark experiments to investigate models predictive ability is an indication of its significance in the domain. Data preprocessing with features selection and fine tuning of SVM hyperparameters in the second place verified the importance of these two to effectively ensure quality classification of SVM. Therefore, there are another two pieces of works which conduct both tasks simultaneously with the proposed new models. Besides, there are also few researches which used features extraction to preprocess the dataset instead of features selection. However, these do not solve the main drawbacks of SVM which are black box property and inefficient computation with large instances. Hence, research on rules extraction and computational efficiency is the remedy corresponding to the two problems. Other credit scoring issues confronted using SVM have a minority count of contributions. They are outlier handling, improvement of classification performances, reject inferences, dynamic scoring, and data imbalance. The attempts to solve various issues with SVM imply its worthiness to be considered as the alternative in credit scoring domain.

(2) *MA Models.* MA models development is summarized in the upcoming tables and figures. Table 3 is the chronological summary of all reviewed studies with MA to show the modelling trend, summarize the issues addressed, and

provide details of fitness function and models considered for comparison in each experiment. Then, Table 4 reports the count of MA models categorized by model type. Figure 2 illustrates the research paper counts corresponding to its research purposes.

Chronological order of the MA models in Table 4 shows the modelling trend. Early MA models are standalone MA with investigative purposes. Initiation of MA in credit modelling is due to the increasing popularity of AI techniques. The development trend in later years is formulation of new hybrid models that persists until recent years. Among the hybrid models, a majority of the studies are the hybrid MA-DM where MA techniques act as the assistant of DM to do the classification task. In view of the MA techniques utilized, GA is considered as the pioneer as well as the dominant MA in credit scoring field since its usage can be observed from the earliest study till recent while GP is the second popular MA. Promising performances with hybrid GA and GP opened up a new page for MA in credit modelling where other MA started to receive attentions.

Based on the types of models formed with MA as summarized in Table 4, hybrid MA-DM is the obvious dominant, followed by standard MA and hybrid MA-MA. The abundant studies to construct hybrid MA-DM indicates that MA can effectively enhance standalone DM credit models performances. Standard MA and hybrid MA-MA have much lesser research works. This may be due to the subjectivity of MA models in formulating the optimization problem to classify credit customers that pose a difficulty for a general usage.

A quick overview of the research purposes with MA models is illustrated in Figure 2. Features selection is the main issue dealt that has the most number of contributed works, followed by rules extraction and hyperparameters tuning. This outcome infers that MA is a useful tool to do data preprocessing. High comprehensibility is always the crucial criterion for credit models. Having a number of MA studies that solve this problem with rules extraction is recognition that MA can produce transparent credit scorecard. In addition, AI models are sensitive to hyperparameters; thus automated tuning with MA in place of manual tuning has been under continuous research. The success of features selection and hyperparameters tuning in ensuring good performance urged few studies to conduct both simultaneously. Other than preprocessing data with features selection, there are a few works that use MA to do features extraction and discretization. Other minority research purposes are investigation of predictive ability and parameter tuning.

(3) *Overall Summary.* Being two different AI techniques, both methods have been actively researched throughout the years that unleashed their great potential in the credit scoring domain. The roles of these two in credit modelling are illustrated in Figure 3 based on the research purposes.

Features selection is the issue most consistently addressed by both models. For features selection, both SVM and MA have almost the same number of works in addressing this issue. However, the main difference is that SVM is the tool to do classification directly whereas MA indirectly do classification as it acts as the assistant to the hybridized DM models

TABLE 3: Summary of literature for MA.

Authors	MA (category)	Other Methods	fitness function	Remarks
<i>Standard MA</i>				
Desai et al. (1997) [60]	GA (EA)	-	no. of misclassification	(i) multiclass problem (ii) compare with NN, LOGIT, DA
Ong et al. (2005) [61]	GP (EA)	-	mean absolute error	(i) rule extraction (ii) compare with LOGIT, NN, TREE, rough sets
Finlay (2006) [62]	GA (EA)	-	GINI coefficient	(i) large credit data application (ii) compare with LOGIT, LR, NN
Cai et al. (2009) [63]	GA (EA)	-	error rate	(i) include misclassification cost into fitness function (ii) rules extraction
Huang et al. (2006) [64]	2stage GP (EA)	-	mean absolute error	(i) compare with LOGIT, TREE, kNN, GP (ii) study effect of misclassification costs
Abdou et al. (2009) [65]	GP (EA)	-	α (sum of square error) + β (classification error)	(i) compare with profit analysis and weight-of-evidence measure
Dong et al. (2009) [66]	SA (IB)	-	accuracy × similarity function × e (e is a coefficient)	(i) rules extraction (ii) compare with DA, kNN, TREE
Martens et al. (2010) [67]	ACO (SI)	-	coverage + confidence	(i) rules extraction (ii) compare with TREE, SVM, majority vote
Yang et al. (2012) [68]	CS (SI)	-	error rates	(i) compare with Altman's Z-score and SVM (ii) rules extraction
Uthayakumar et al. (2017) [69]	ACO (SI)	-	$TPR + TNR$	(i) compare with LOGIT, NN, RF, RBF network (ii) compare with NN, RF, RBF network
<i>Hybrid MA-MA</i>				
Jiang et al. (2011) [70]	SA+GA (IB+EA)	NN	$\alpha(1 - FPR)$	(i) SA optimizes GA (ii) parameter tuning (iii) compare with standalone NN and GA-optimized NN
Aliehyaei and Khan (2014) [71]	ACO (SI), GP (EA), ACO+GP	-	mean absolute error	(i) rules extraction by ACO to input to GP (ii) compare with ACO and GP
<i>Hybrid MA-DM</i>				
Fogarty and Ireson (1993) [72]	GA (EA)	BN	accuracy	(i) features extraction (ii) large credit data application (iii) compare with default rule, BN, kNN, TREE

TABLE 3: Continued.

Authors	MA (category)	Other Methods	fitness function	Remarks
Hand and Adams (2000) [73]	SA (IB)	WOE, weighted WOE LR	R^2	(i) features discretization (ii) compare with LOGIT, DA, two other discretization methods
Drezner et al. (2001) [74]	TS (IB)			(i) features selection (wrapper approach) (ii) compare with Altman's Z-score
Lacerda et al. (2005) [75]	GA (EA)	NN	average of individuals with rank of individual from a population accuracy	(i) hyperparameters tuning (ii) compare with NN, consecutive learning algorithm, SVM
Huang et al. (2007) [35]	GA (EA)	SVM	accuracy+expected misclassification cost Information value	(i) hyperparameters tuning, feature selection (wrapper approach) (ii) compare with NN, GP, TREE
Zhang et al. (2008) [38]	GP (EA)	SVM		(i) rules extraction (ii) compare with SVM, GP, TREE, LOGIT, NN
Liu et al. (2008) [76]	GP (EA)	DA		(i) features extraction (select derived characteristics) (ii) large dataset from finance enterprise
Wang et al. (2008) [77]	GA (EA)	NN	1/MSE	(iii) compare with DA (i) compare with DA (ii) parameters tuning (iii) compare with NN
Zhang et al. (2008) [78]	GA (EA)	TREE	$\alpha(1\text{-info entropy}) + (1 - \alpha)(1\text{-info entropy})$	(i) features extraction (ii) compare with TREE, NN, GP, GA-optimized SVM, rough set
Jiang (2009) [79]	SA (IB)	TREE	(i) $\alpha \cdot FNR + \beta \cdot TNR$ (ii) $(FNR - \alpha)^2 + (FPR - \beta)^2$	(i) rules extraction (ii) compare with TREE
Marinakis et al. (2009) [80]	ACO (SI), PSO (SI)	kNN, 1-NN, weighted kNN	(iii) $\alpha \cdot FN/(TP + 1) + \beta \cdot TN/(FP + 1)$ not mentioned	(i) multiclass problem, features selection (wrapper approach) (ii) compare with wrapper models of GA and TS with kNN (and variants)
Sustersic et al. (2009) [81]	GA (EA)	NN	accuracy $> \theta_b$ and $RMSE < \theta_a$, θ_b, θ_a preset threshold error rate	(i) feature selection (wrapper approach) (ii) compare with NN, LOGIT (features from PCA)
Zhou et al. (2009) [42]	GA (EA)	SVM		(i) hyperparameters tuning (ii) compare with LOGIT, kNN, DA, TREE
Zhou et al. (2009) [41]	GA (EA)	WSVM	AUC	(i) hyperparameters tuning, feature selection (wrapper approach) (ii) compare with LR, LOGIT, NN, TREE, kNN, AdaBoost
Marinaki et al. (2010) [82]	HBO (SI)	kNN, 1-NN, weighted kNN	accuracy	(i) multiclass problem, features selection (wrapper approach) (ii) compare with wrapper models of GA, PSO, TS, ACO with kNN (and variants)
Huang and Wu (2011) [83]	GA (EA)	kNN, BN, TREE, LR, SVM, NN, AdaBoost, Logitboost, Multiboost	accuracy	(i) feature selection (wrapper approach) (ii) compare with kNN, BN, TREE, LOGIT, SVM (features from filter selection approach)

TABLE 3: Continued.

Authors	MA (category)	Other Methods	fitness function	Remarks
Yu et al. (2011) [44]	GA (EA)	LS-SVM	accuracy	(i) hyperparameters tuning (ii) study class imbalance problem (iii) compare with results from [8, 45]
Correa and Gonzalez (2011) [84]	BPSO (SI), GA (EA)	NN	AUC	(i) hyperparameters tuning (ii) large credit dataset (iii) cost study on application, behavioural and collection scoring (iv) compare with LOGIT, NN, Global Optimum
Oreski et al. (2012) [85]	GA (EA)	NN	accuracy	(i) hyperparameters tuning, feature selection (wrapper approach) (ii) study effect of misclassification costs (iii) compare their proposed model with features from filter selection approach (i) model selection (ii) hyperparameters tuning (i) multiclass problem (ii) hyperparameters tuning (iii) compare with SVM (tuned with GA and PSO)
Danenas and Garsva (2012) [47]	GA (EA),PSO (SI)		TPR	
Chen et al. (2013) [49]	ABC (SI)	SVM	not mentioned	
Garsva and Danenas (2014) [51]	PSO (SI)	SVM	(i) TPR, (ii) accuracy	(i) model selection (ii) hyperparameters tuning (iii) study class imbalance problem (iv) compare among all SVM, LS-SVM tuned with PSO, DS, SA
Oreski and Oreski (2014) [86]	GA (EA)	NN	accuracy	(i) feature selection (wrapper approach) (ii) compare with standard wrapper-based NN with GA
Danenas and Garsva (2015) [52]	PSO (SI)	SVM	TPR	(i) model selection (ii) hyperparameters tuning (iii) compare with LOGIT, RBF network
Wang et al. (2010) [87]	TS (IB)	NN, SVM, LOGIT	entropy	(i) features selection (filter approach)
Wang et al. (2012) [88]	SS (SI)	NN, TREE, LOGIT	entropy	(ii) compare with NN, SVM, LOGIT with full features
Waad et al. (2014) [89]	GA (EA)	LOGIT, SVM, TREE	$\sum_i weight_i \times Distance(\sigma, L_i)$	(i) features selection (filter approach) (ii) compare with NN, TREE, LOGIT with full features (i) feature selection (filter approach) (ii) compare with LOGIT, SVM, TREE (features from other filter selection and rank aggregation methods)
Hsu et al. (2018) [53]	ABC (SI)	SVM	$\frac{1}{1 + abs(f_i(x_{ij}))}, if f_i(x_{ij}) \geq 0$ $\frac{1}{1 + abs(f_i(x_{ij}))}, if f_i(x_{ij}) < 0$	(i) multiclass problem (ii) hyperparameters tuning (iii) compare with standalone SVM, kNN, NB and their wrappers with GA
Jadhav et al. (2018) [54]	GA (EA)	SVM, kNN, NB	accuracy	(i) features selection (wrapper approach) (ii) compare with LOGIT and SVM (tuned with GS, GA, PSO)
Wang et al. (2018) [55]	multiple population GA (EA)	SVM	accuracy	(i) features selection (wrapper approach)
Krishnaveni et al. (2018) [90]	HS (SI)	1-NN	accuracy	(ii) compare with MPGA-SVM, GA-SVM, SVM (i) features selection (wrapper approach) (ii) computational time reduction (iii) compare with standalone SVM, TREE, kNN, NB, NN and their wrappers with GA and PSO

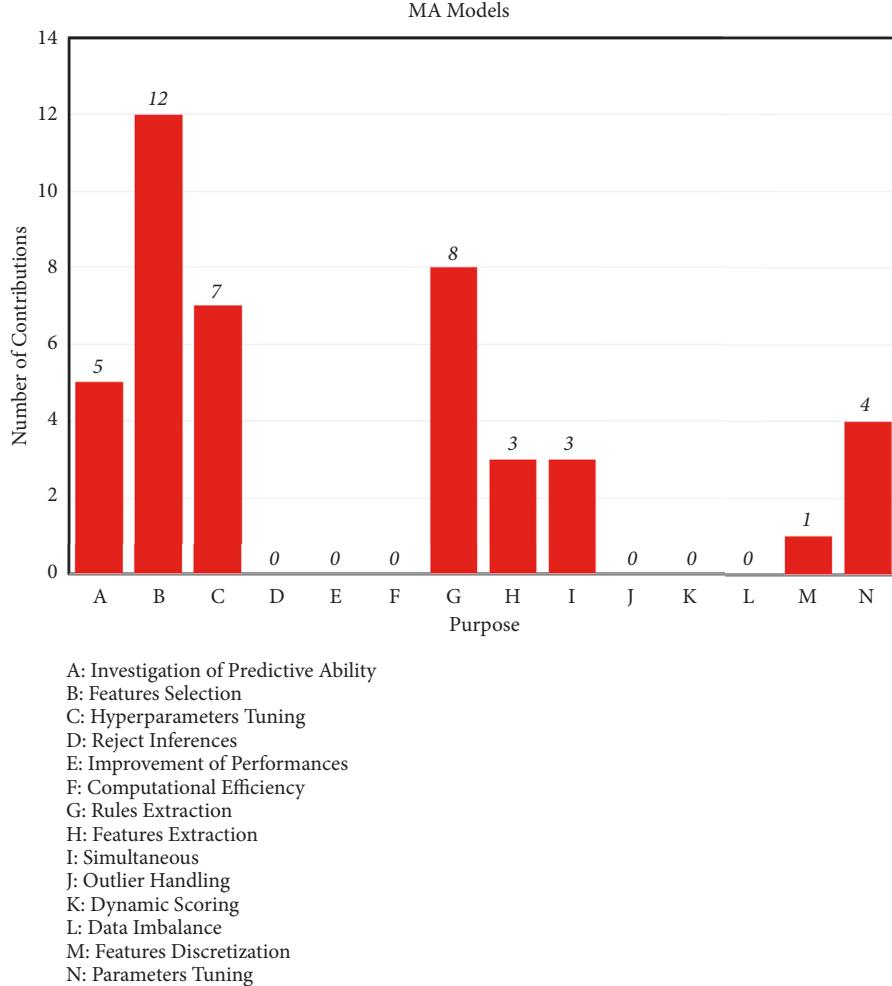


FIGURE 2: Summary of MA models.

TABLE 4: Type of MA models.

Type	Count
Standard MA	10
Hybrid MA-MA	2
Hybrid MA-DM	31
Total	43

which are responsible for the classification. Investigation of predictive ability comes as the second top research purposes. SVM models have much greater number of researches as compared to GA. This indicates that SVM is already a recognized credit scoring model as it is frequently included in comparative studies and attempted in different specific domains. MA has lesser works under this purpose as it is seldom involved in benchmark experiments that may be due to its subjectivity in model building. Then, rules extraction comes in the third largest number of researches, with MA models more than SVM. This shows the great ability of MA to develop transparent model. Hyperparameters tuning comes

as the next with seven contributions where all the seven are the collaboration of MA with SVM. Thus, MA can be viewed as a recommended tool to tune SVM hyperparameters. Simultaneous features selection and hyperparameters tuning result in a total of three studies, with two of them being hybrid MA and SVM.

The rest of the research purposes have shown dominance in either SVM or MA. Feature discretization has only been attempted by MA while reject inferences, improvement of performances, computational efficiency, dynamic scoring, imbalance datasets, and outlier handling have only been addressed using SVM. MA models have taken into account much lesser issues as compared to SVM since MA have been focused more to solve features selection and rules extraction.

4.2. Assessment Procedures. In order to assess credit models performance, they are usually compared with other standard credit models applied on the selected credit datasets and evaluated with appropriate performance measures. Thus, the assessment procedures are categorized into *benchmark experiments*, *performance measures*, and *credit datasets*.

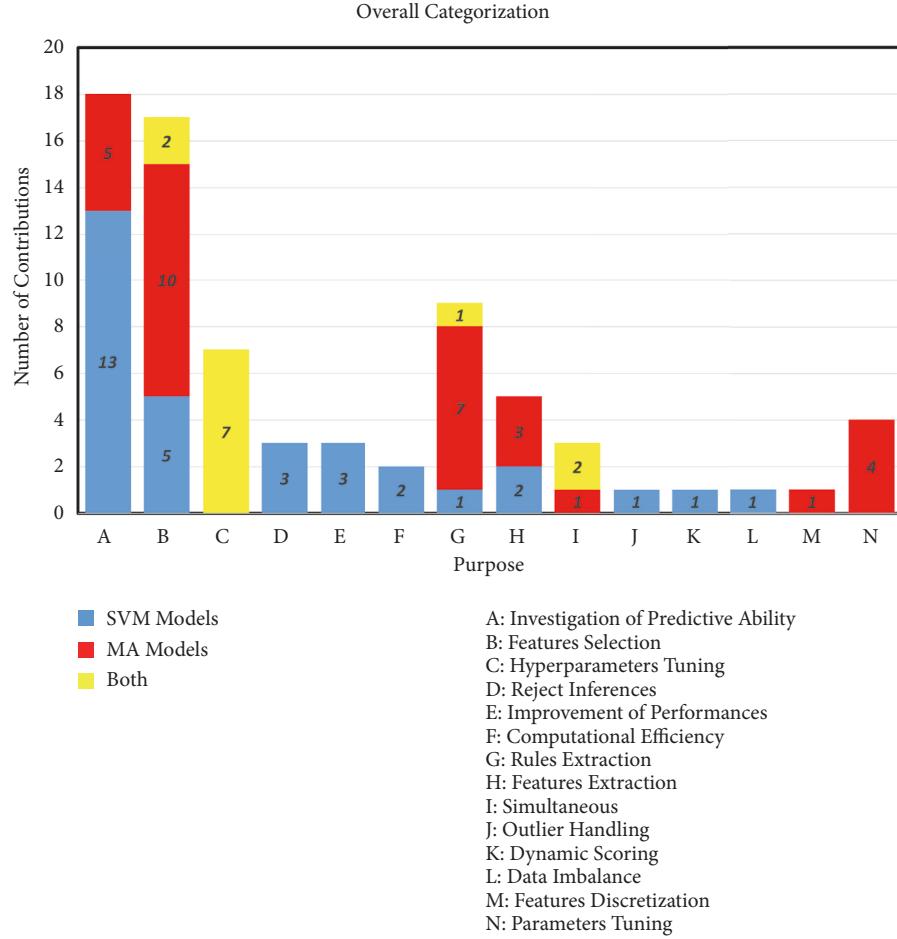


FIGURE 3: Overall summary for both models.

4.2.1. Benchmark Experiments. Benchmark experiments include comparisons of the proposed models with other standard credit models. Tables 1 and 2 provided brief summary on the models considered for comparison in every experiment. Detailed experiment setup shall be referred to the original paper. Table 5 presents the categorization of the type of benchmark experiment carried out with SVM and MA models.

As reported in Table 5, it can be seen that inclusion of model comparison has been a standard approach to make conclusion on the proposed models. Most of the studies adopted internal benchmark approach to make comparison with other models based on the same experimental setup for the credit data. Only Yu et al. [44] adopted external benchmark approach to compare their proposed models with other models. Chen et al. [48] and Cai et al. [63] are the rare cases in the literature which do not benchmark their proposed models with others.

Large scale benchmark is a rare approach with only several studies presented this in the past. It can be noticed that research work with large scale benchmark are those that conducted comparative studies and formulate new ensemble models for performance improvement. It is necessary for

comparative studies to include sufficient huge number of methods as benchmark to be able to provide sufficient information to serve as a guideline for future research. Since ensembles are formulated from assembly of a number of standalone classifiers, authors that proposed new ensembles usually have to compare new ensembles not only with standalone classifiers but also with standard available ensembles.

Small scale benchmark is the most common approach which can be further broken down into four main parts, i.e., comparison only with the counterpart techniques of the proposed model, comparison only with either statistical or AI techniques, and comparison with both statistical and AI techniques. For both SVM and MA models, the most preferred small scale benchmark is comparison with both statistical and AI techniques. Besides, LOGIT and NN are the most frequently involved statistical and AI techniques, respectively.

4.2.2. Performance Measures. There are four main types of performance measures being used to make inference on the models performances. Cutoff-dependent measures are those directly obtained or computed from the confusion matrix,

TABLE 5: Benchmark experiments.

Benchmark Type		SVM Models	MA Models
Large Scale		[8, 11, 12, 18, 27, 56–59]	-
Small Scale	Counterpart	[33, 40, 43]	[70, 71, 79, 80, 82, 83, 85, 86, 89] [77, 87, 88]
	Statistical	[30, 50]	[65, 73, 74, 76]
	AI	[20, 21, 46]	[67, 72, 75, 78]
	Statistical & AI	[22–25, 29, 31, 32, 34, 36] [19, 28, 37, 39]	[60–62, 64, 66, 69, 81, 84, 90] [68]
			[38, 41, 42, 52, 53]
External			[44]
None		[48]	[63]

TABLE 6: Performance measures.

Type	SVM Models	MA Models
Cutoff-dependent	[12, 19–21, 23, 25, 26, 28, 30, 36, 37, 39, 40, 46, 48] [35, 38, 42, 44, 47, 49, 51–53, 55]	[60, 61, 63–72, 74–79, 81–83, 85, 86, 89, 90]
Cutoff-independent	[24, 33, 34]	[84]
Mixture	[8, 11, 18, 22, 27, 29, 31, 32, 43, 50, 56, 57, 59] [54]	[1, 62, 87, 88]
Business-oriented	[33, 34]	[65, 72, 76, 86]
Others	[58]	-

where the cutoff point is often problem-dependent. Cutoff-independent measures are those computed to determine the discriminating ability of the model. Mixture indicates the usage of both cutoff-dependent and cutoff-independent measures while business-oriented measures are those computed with the misclassification costs. Table 6 summarizes the performance measures in the literature of SVM and MA models.

Cutoff-dependent measures are the most popular indicator utilized by researchers, especially accuracy or its counterpart error rate that measures the number of correct classifications in a straightforward manner. Among them, several studies of SVM models [21, 22, 28, 30–32, 40, 57] and MA models [70, 77, 81, 83] are interested in the misclassifications which is believed to pose higher risk for financial institutions; thus type I and type II errors are included. Although cutoff-dependent measures are direct in presenting performances of classifier, a main drawback has been denoted by [4, 11, 24], where researchers often do not address the cutoff point used. Thus, there are a few studies believed cutoff-independent performance measures are sufficient to serve as guideline for model performances as reported in Table 6, with Area Under Receiver Operating Characteristics Curve (AUC) being most popularly used.

Cutoff-dependent measures from confusion matrix and cutoff-independent measures of model discriminating ability are both informative for decision makers. Hence, a few studies have included both types of measures in their experiment

to provide different perspective of interpretation. In one of the latest and largest comparative studies by Lessmann et al. [11], they have recommended the use of more cutoff-independent measures with the aim to explain models in different perspective. Therefore, their recommendations have been adopted in recent research [59].

Profit and loss is often the final aim for financial institutions as credit scoring is treated as a risk aversion tool. Only a minority of studies explained their models with business-oriented measures. They are Expected Misclassification Costs (EMC) [65, 86], profitability [72], and self-defined profit or cost measures [33, 34, 76]. Note that all researches utilizing business-oriented measures also included cutoff-dependent or cutoff-independent measures to evaluate their models. Only Yu et al. [58] applied weighed accuracy that involved imbalance rate, revenue and cost to compute a new version of accuracy.

4.2.3. Model Evaluation. Researchers make inferences and conclusions based on the reported performance measures. The conclusions are often to show that the proposed models are competitive or outperform the other standard credit models based on the numerical results. In addition, some studies conducted statistical tests to show evident improvement of the proposed models.

There are several studies of SVM models [29, 36, 46, 59], MA models [60, 69, 71, 78, 84], and joint MA-SVM [41, 47, 51, 52] that do not have numerical outperformance of

TABLE 7: Credit datasets.

Type	SVM Models	MA Models
Specific	<p><i>CR:</i> BankScope [19], England [22], Taiwan/US [20], Korea [23, 25], Taiwan [21], credit card [24]</p> <p><i>BP:</i> US [26]</p> <p><i>AS:</i> German [30], LC P2P [31], LC/Huijin P2P [32], China [37, 48]</p> <p><i>BS:</i> Chilean [27, 33, 34]</p>	<p><i>CR:</i> UK [80, 82]</p> <p><i>BP:</i> Simulated data [68]</p> <p><i>AS:</i> Southeastern US [60], UK [62], Egyptian [65], Compustat [74], Shenzhen [70, 77], Croatian [85], UK/simulated [1]</p> <p><i>AS/BS/CS:</i> local bank [84]</p>
General	AS: [12, 39, 43, 57]	AS: [61, 63, 64, 66, 70, 71, 75, 78, 83, 87, 88, 90] AS: [35, 38, 41, 42, 44, 51, 55]
Specific & General	AS: UK CDROM [28], Barbados [29], Benelux [36]	AS: Croatian [86], Tunisian/home equity [89] <i>BP:</i> ANALCAT [69] <i>AS/BP/CR:</i> SME/ BankScope [67]
Large scale comparison	AS: [8, 11] AS/BP: [18]	AS: Taiwan [54] <i>CR:</i> Compustat [53]

their proposed models with the compared models. However, they have contributed to another aspect, i.e., outperformance in timing [29, 46], proper handling of imbalance [59], transparent rules [36, 69], and presentation of competitive new methods [41, 47, 51, 52, 60, 71, 78, 84].

Some comparative studies [8, 11, 12] identified certain methods to have best performance and recommended for future research but they did not penalize the use of other techniques since the reported performance is still very competitive, whereas comparative studies by [18, 26] did not result in outperformance of any methods among the compared models. The rest of the research articles reviewed in this study have reported their proposed models having better results than the others.

Instead of solely depending on numerical improvement to detect outperformance, a minority of the studies [8, 11, 12, 19, 21, 23, 29, 34–36, 43, 50, 53, 55, 56, 59, 60, 67, 89] have utilized statistical tests for a more concrete support of their results. Most commonly used statistical tests are paired t-test, Mc Nemar test, Wilcoxon signed rank test, and Friedman test.

4.2.4. Credit Datasets. SVM and MA credit models have been applied on different types of credit datasets, i.e., application scoring (AS), behavioural scoring (BS), collection scoring (CS), corporate rating (CR), and bankruptcy prediction (BP). There have been two main types of studies which are specific studies particularly on a country's financial institutions and general studies using publicly available datasets from the UCI repository [94]. Table 7 summarizes the credit datasets usage.

For specific studies, they have focused on particular financial institutions, with application scoring being the mostly utilized data type, followed by corporate rating and

bankruptcy prediction. Behavioural scoring has received very less attention. For general studies, there are three datasets: German, Australian, and Japanese datasets available in the UCI repository, and all of them are application scoring data. Among them, German and Australian have been widely involved in research. There are also several specific studies that included UCI datasets as evaluation purpose, and similarly, German and Australian are still the dominant usage among researchers. It can be noticed that almost all of the studies utilized small numbers of datasets for investigation, with only comparative studies involving large amount of datasets.

4.3. Results from Literature. German (G) and Australian (A) datasets have been frequently included in credit scoring domain using SVM and MA models. This section compiles those contributed works based on the research purposes to provide information on which model type with the handled issues have shown good performance on both datasets. Note that the compilation is only based on accuracy as it is the common reported measures. For researches that reported only error rate, it is converted to accuracy. Researches that have utilized these datasets but do not report accuracy are not included here. Then, for usage of more than one standard SVM and its variants, as well as studies that formulated more than one new model, only the best performing results are recorded. The results are compiled in Table 8.

The computed mean in Table 8 is the overall general performance of the models on both datasets. The high value of mean accuracy is an indication of good performance from SVM and MA models in the literature. The computed standard deviation in Table 8 is viewed as an indicator to generalize the range of the accuracy that is believed to take

TABLE 8: Results from literature.

Purpose	Model Type	Authors	G	A
Investigation of predictive ability	standard SVM and its variants	Baesens et al. [8] Lessmann et al. [11]	74.30 75.30	89.10 86.00
	standalone MA	Boughaci and Alkhawaldeh [18] Cai et al. [63]	69.90 80.00	80.70 -
	modified SVM	Harris [29]	77.10	-
	hybrid SVM	Hens and Tiwari [46]	75.08	85.98
Improvement of classification performance	ensembles	Zhou et al. [56]	78.13	-
		Ghodsehahi [57]	81.42	-
		Xia et al. [59]	78.32	86.29
Rules extraction	hybrid SVM	Martens et al. [36]	-	85.10
	standalone MA	Ong et al. [61] Huang et al. [64] Dong et al. [66] Martens et al. [67]	77.34 79.49 72.90 80.80	88.27 89.17 -
	hybrid MA-MA	Uthayakumar et al. [69] Aliehyaei and Khan [71]	- 70.70	86.37 84.30
	hybrid MA-DM	Zhang et al. [38] Jiang et al. [79]	79.88 73.10	89.45 -
	hybrid SVM	Xu et al. [40] Han et al. [50]	- 75.00	89.28 -
	hybrid MA-DM	Zhang et al. [78]	77.76	89.33
	hybrid SVM	Yao [39] Chen and Li [43]	76.60 76.70	87.52 86.52
	hybrid MA-DM	Jadhav et al. [54] Wang et al. [55] Huang and Wu [83] Oreski and Oreski [86] Krishnaveni et al. [90] Wang et al. [88]	82.80 78.53 -	90.75 86.96 87.54 -
	hybrid MA-DM	Zhou et al. [42] Yu et al. [44] Garsva and Danenas [51] Hsu et al. [53] Lacerda et al. [75]	77.10 78.46 81.30 84.00 -	86.96 90.63 87.40 92.75 86.05
Hyperparameters tuning	hybrid MA-DM	Huang et al. [35]	77.92	86.90
Simultaneous features selection & hyperparameters tuning		mean standard deviation	77.56 3.35	87.75 2.64

into account the variation in different experimental setup for every model development.

The general mean and standard deviation in Table 8 provide an overview on the models performances through compilation from all studies regardless of the data splitting strategies adopted. Data splitting methods are influential on the experiments end results. The general compilation in Table 8 is further detailed in Table 9 to take into consideration the effect of different data splitting strategies. The detailed analysis is conducted by categorizing the studies (except Jiang [79] as no data splitting mentioned) into four main types

of data partitioning, i.e., k-fold cross validation (k-fold), holdout validation (holdout), repeated k-fold cross validation (rep k-fold), and repeated holdout validation (rep holdout).

Table 9 shows that k-fold and holdout are the most adapted data splitting strategies while rep k-fold and rep holdout are less popular, which may be due to the high computational effort required for both rep k-fold and rep holdout. Instead of a general mean and standard deviation as in Table 8, the specific mean and standard deviation for each category are reported in Table 9, which is believed to be less biased and more reliable metrics as being grouped

TABLE 9: Results categorized with data splitting methods.

Data Split	Authors	G	A
k-fold cv	Dong et al. [66]	72.90	-
	Uthayakumar et al. [69]	-	86.37
	Zhang et al. [38]	79.88	89.45
	Xu et al. [40]	-	89.28
	Han et al. [50]	75.00	-
	Zhang et al. [78]	77.76	89.33
	Yao [39]	76.60	87.52
	Chen and Li [43]	76.70	86.52
	Jadhav et al. [54]	82.80	90.75
	Wang et al. [55]	78.53	86.96
	Huang and Wu [83]	-	87.54
	Oreski and Oreski [86]	78.90	-
	Wang et al. [88]	-	88.90
	Zhou et al. [42]	77.10	86.96
	Yu et al. [44]	78.46	90.63
	Hsu et al. [53]	84.00	92.75
	Huang et al. [35]	77.92	86.90
	mean	78.20	88.56
	standard deviation	2.93	1.92
holdout	Baesens et al. [8]	74.30	89.10
	Cai et al. [63]	80.00	-
	Boughaci and Alkhawaldeh [18]	69.90	80.70
	Harris [29]	77.10	-
	Zhou et al. [56]	78.13	-
	Ghodselahei [57]	81.42	-
	Martens et al. [36]	-	85.10
	Martens et al. [67]	80.80	-
	Aliehyaei and Khan [71]	70.70	84.30
	Garsva and Danenas [51]	81.30	87.40
	mean	77.07	85.32
	standard deviation	4.47	3.20
rep k-fold	Lessmann et al. [11]	75.30	86.00
	Xia et al. [59]	78.32	86.29
	Krishnnaveni et al. [90]	80.40	93.50
	mean	78.01	88.60
	standard deviation	2.56	4.25
rep holdout	Ong et al. [61]	77.34	88.27
	Huang et al. [64]	79.49	89.17
	Lacerda et al. [75]	-	86.05
	mean	78.42	87.83
	standard deviation	1.52	1.61

homogeneously according to the data splitting methods. For all the categories, across both datasets, the mean accuracies are relatively high, showing the effectiveness of building new credit models with both SVM and MA.

However, for the holdout category, it can be noticed that the standard deviation is much higher than the other categories. Observing the holdout category, the high deviation is deduced to be contributed from an unusual lower accuracy from Boughaci and Alkhawaldeh [18]. This lower accuracy

may be due to the usage of SVM without the hyperparameters tuning procedure. Besides, another high standard deviation comes from the rep k-fold category only in Australian dataset. This high deviation is from a higher accuracy results from Krishnnaveni et al. [90]. This may be due to the nature of this dataset that suits well to the proposed method.

To provide information on which models are more effective than the others, models with accuracy higher than the mean is viewed as having greater potential to deal with

the credit scoring problem and would be recommendable for future research. Higher potential models are reported in italic (see Table 9), and they are compared with the respective mean accuracy in every category. All accuracies written in italic in Table 8 correspond to the higher potential credit models obtained from Table 9. Generally, rules extraction, hyperparameters tuning and features selection are observed as effective measures to be undertaken for well-performed credit models. This aligned with our previous discussion in Section 4.1.3.

For features selection and hyperparameters tuning, both SVM and MA appears to be the crucial tools for credit modelling, where SVM is the main classifier and MA is the assistant to be fused with SVM to carry out the targeted tasks. For rules extraction, MA is the best choice to build transparent credit scoring models, either in standalone MA or hybridized with other black box DM. Among the different variants of SVM, standard SVM is most commonly adopted by researchers for new model development. On the other hand, among the different MA, GA and GP appear to be the dominant tool, while recent trend has shifted to other types of MA for new model building.

Inclusion of German and Australian datasets into large scale benchmark comparative studies of credit models is an indication of the status of these two becoming the standard credit datasets in this domain. This leads to frequent usage of both datasets throughout the years until recent.

5. Conclusions and Future Directions

This study presented a literature review of credit scoring models formulated with SVM and MA. From the two aspects, model type with issues addressed and assessment procedures, together with past results of models applied on UCI credit datasets, hybrid approach is identified as the state-of-the-art modelling approach for both techniques in the credit scoring domain. SVM and MA have been the current trend for credit modelling with SVM being the main classifier and MA being the assistant tool for model enhancement with hybrid approach. Aligned with the views from [1, 11, 12] that concluded sophisticated models are the future trend, both SVM and MA will also have the similar future modelling trend. Various issues and assessment procedures in the literature are concluded to point out some future directions as follows.

5.1. Model Type with Issues Addressed

5.1.1. SVM Models. SVM is an ongoing active research in credit scoring, where the future development trend is perceived as building new SVM models based on hybrid and ensembles method. Several directions are pointed out for possible future works. There are various SVM variants available that are able to account for more flexible classification. A benchmark experiment comparing standard SVM and its variants can be conducted to give insight on which SVM type is more adaptable in this domain. When building new models, SVM variants that include different regularization

terms shall be considered. Danenas et al. [26] provided information on the various choice of available kernels for SVM; thus instead of the common linear and RBF, other kernels shall be investigated. Modified SVM in the literature involved modifications on the hyperplane optimization problem. Introduction of new kernels can be perceived as a possible future work for modified SVM category, due to the flexibility of SVM itself, where any kernels that follow the Karush-Kuhn Tucker conditions can be used in SVM. Hybrid approach is the majority for features selection, only two works [33, 34] adopted cost and profit view to handle this issue. Future features selection can consider incorporation of cost and profit in model building process as profit scoring is suggested as the main future trend in [2, 11]. Besides, other research purposes that have few publications are valuable future directions to be considered.

5.1.2. MA Models. Hybrid MA-DM has been the leading model type throughout the years, and it is believed that this trend will persist in the future since hybrid models formulation is considered as a direct way to propose new models yet able to improve the standalone DM. EA, i.e., GA and GP, are the most popular MA to be applied in credit scoring domain. It can be observed that MA (other than EA family) have increasing publications in recent years. Since there are various choices for MA in each family, those that have not yet been investigated in credit scoring domain, such as, Social Cognitive Optimization, Bat Algorithm, Local Search, Variable Neighbourhood Search, etc., should be considered for new model development. Although standard MA is seldom used for model development, the transparent property, and flexibility of MA to be tailored for to solve for specific credit data is a plus point to use MA in credit modelling. In addition, business-oriented model is an important prospect for decision makers. Hence, the flexibility of MA shall be employed to incorporate costs and benefits into model formulation. In view of computational efficiency, MA is very adaptable where the operators can be carefully modified or parallelized to achieve high efficiency.

5.1.3. Overall. Features selection, hyperparameters tuning and rules extraction are concluded as the few popular issues addressed with SVM and MA models, with higher tendency to deal with hybrid methods. Results from past experiments with German and Australian datasets further validate these three as the trend for both models in credit scoring. As compared to SVM, MA is mostly limited to deal with these three issues, while SVM dealt with wider varieties of issues. Thus, instead of limiting MA to build models based on the few popular purposes, other issues that have been attempted as in SVM models shall be considered since flexibility of MA to make it tailored to specific problem is always possible. MA is mostly incorporated with SVM to form hybrid. Other AI models that have not yet been attempted for the same research purpose are worth being investigated. Besides, MA is advantageous in rules extraction while SVM is a black box model. These two can be collaborated to formulate a transparent yet competitive model. Lastly, formation of

business-oriented SVM models with hybrid approach is a more direct task than to modify the algorithm in SVM. The adaptable property of MA is a good prospect to join it with SVM, where MA is responsible to take account of cost and profit in the modelling procedure. Ensembles modelling is also a future research direction as it has just received attention in credit scoring domain lately. Business-oriented ensemble with MA-tuned SVM model is a recommendable future research.

5.2. Assessment Procedures. The whole assessment procedures in each experiment are a subjective decision based on the experimental setup of the researchers. Internal small scale benchmark experiments including both statistical and AI techniques, reporting the numerical improvement of cutoff-dependent measures, is the leading trend in credit scoring domain. Internal small scale benchmark is the recommended experimental setup for future experiments due to two main considerations. First, newly developed model has to be compared with standard credit scoring models from both statistical and AI families as a concrete support of the usefulness of the proposed models but large scale comparison is not necessary to avoid loss of focus. Second, external benchmark comparing with results from other researches can be considered but have to be used very cautiously because different experiments have different design that may be difficult to have an apple-to-apple comparison. It is also recommended to report both cutoff-dependent and independent measures together and disclose the assumed cutoff point when using cutoff-dependent measures. There have been varieties of performance measures being recommended in Lessmann et al. [11], but the choice shall be carefully made instead of using as many as possible of these measures. This is because every measure evaluates models in different approach; researchers shall choose based on what is the most important information needed to be conveyed as it is not always “the more the better” [4]. Business-oriented measures can be viewed as a future trend to evaluate models, provided that the credit dataset should have information for the profit and costs. Although numerical improvement is considered a valuable contribution, it is recommended to include statistical tests for further validation on the statistical significance of model performances. The trend of focusing in application scoring especially with public UCI datasets may be due to the confidentiality issue of credit scoring domain. Recently, there are some datasets available from data mining competitions and also Lending Club (LC) online P2P lending platform. Thus, other types of dataset shall be involved to check models robustness across different types of data. Besides, LC datasets provided more information that can be considered to build models for behavioural scoring. For specific research using private datasets, profit scoring may be a future research trend and it is recommended to include UCI datasets as a benchmark due to their general usage in the literature.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research is funded by Geran Putra-Inisiatif Putra Siswazah (GP-IPS/2018/9646000) supported by Universiti Putra Malaysia (UPM).

References

- [1] D. J. Hand and W. E. Henley, “Statistical classification methods in consumer credit scoring: a review,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 160, no. 3, pp. 523–541, 1997.
- [2] L. C. Thomas, “A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers,” *International Journal of Forecasting*, vol. 16, no. 2, pp. 149–172, 2000.
- [3] S. M. Sadatrasoul, M. R. Gholamian, M. Siami, and Z. Hajimohammadi, “Credit scoring in banks and financial institutions via data mining techniques: a literature review,” *Journal of AI and Data Mining I*, vol. 1, no. 2, pp. 119–129, 2013.
- [4] N. Martin, “Assessing scorecard performance: a literature review and classification,” *Expert Systems with Applications*, vol. 40, no. 16, pp. 6340–6350, 2013.
- [5] H. A. Alaka, L. O. Oyedele, H. A. Owolabi et al., “Systematic review of bankruptcy prediction models: Towards a framework for tool selection,” *Expert Systems with Applications*, vol. 94, pp. 164–184, 2018.
- [6] J. Sun, H. Li, Q.-H. Huang, and K.-Y. He, “Predicting financial distress and corporate failure: a review from the state-of-the-art definitions, modeling, sampling, and featuring approaches,” *Knowledge-Based Systems*, vol. 57, pp. 41–56, 2014.
- [7] S. Moro, P. Cortez, and P. Rita, “An automated literature analysis on data mining applications to credit risk assessment,” in *Artificial Intelligence in Financial Markets*, pp. 161–177, Palgrave Macmillan, London, 2016.
- [8] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, “Benchmarking state-of-the-art classification algorithms for credit scoring,” *Journal of the Operational Research Society*, vol. 54, no. 6, pp. 627–635, 2003.
- [9] J. N. Crook, D. B. Edelman, and L. C. Thomas, “Recent developments in consumer credit risk assessment,” *European Journal of Operational Research*, vol. 183, no. 3, pp. 1447–1465, 2007.
- [10] A. Lahsasna, R. N. Ainan, and T. Y. Wah, “Credit scoring models using soft computing methods: a survey,” *International Arab Journal of Information Technolog*, vol. 7, no. 2, pp. 115–123, 2010.
- [11] S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas, “Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research,” *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, 2015.
- [12] F. Louzada, A. Ara, and G. B. Fernandes, “Classification methods applied to credit scoring: systematic review and overall comparison,” *Surveys in Operations Research and Management Science*, vol. 21, no. 2, pp. 117–134, 2016.
- [13] A. I. Marques, V. García, and J. S. Sánchez, “A literature review on the application of evolutionary computing to credit scoring,” *Journal of the Operational Research Society*, vol. 64, no. 9, pp. 1384–1399, 2013.
- [14] D. Durand, *Risk Elements in Consumer Instalment Financing*, NBER Books, 1941.
- [15] E. I. Altman, “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy,” *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.

- [16] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *Journal of Accounting Research*, vol. 18, no. 1, pp. 109–131, 1980.
- [17] P. Kolesar and J. L. Showers, "A robust credit screening model using categorical data," *Management Science*, vol. 31, no. 2, pp. 123–133, 1985.
- [18] D. Bougaci and A. A. Alkhawaldeh, "Appropriate machine learning techniques for credit scoring and bankruptcy prediction in banking and finance: a comparative study," *Risk and Decision Analysis*, pp. 1–10, 2018.
- [19] I. T. Van Gestel, B. Baesens, I. J. Garcia, and P. Van Dijcke, "A support vector machine approach to credit scoring," *Bank- en Financiewezen*, vol. 2, pp. 73–82, 2003.
- [20] Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu, "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decision Support Systems*, vol. 37, no. 4, pp. 543–558, 2004.
- [21] S.-T. Li, W. Shiue, and M.-H. Huang, "The evaluation of consumer loans using support vector machines," *Expert Systems with Applications*, vol. 30, no. 4, pp. 772–782, 2006.
- [22] K. Lai, L. Yu, L. Zhou, and S. Wang, "Credit risk evaluation with least square support vector machine," *Rough Sets and Knowledge Technology*, pp. 490–495, 2006.
- [23] Y.-C. Lee, "Application of support vector machines to corporate credit rating prediction," *Expert Systems with Applications*, vol. 33, no. 1, pp. 67–74, 2007.
- [24] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3302–3308, 2009.
- [25] H. S. Kim and S. Y. Sohn, "Support vector machines for default prediction of SMEs based on technology credit," *European Journal of Operational Research*, vol. 201, no. 3, pp. 838–846, 2010.
- [26] P. Danenas, G. Garsva, and S. Gudas, "Credit risk evaluation model development using support vector based classifiers," *Procedia Computer Science*, vol. 4, pp. 1699–1707, 2011.
- [27] J. Mushava and M. Murray, "An experimental comparison of classification techniques in debt recoveries scoring: evidence from South Africa's unsecured lending market," *Expert Systems with Applications*, vol. 111, pp. 35–50, 2018.
- [28] Y. Wang, S. Wang, and K. K. Lai, "A new fuzzy support vector machine to evaluate credit risk," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 6, pp. 820–831, 2005.
- [29] T. Harris, "Credit scoring using the clustered support vector machine," *Expert Systems with Applications*, vol. 42, no. 2, pp. 741–750, 2015.
- [30] Y. Yang, "Adaptive credit scoring with kernel learning methods," *European Journal of Operational Research*, vol. 183, no. 3, pp. 1521–1536, 2007.
- [31] Z. Li, Y. Tian, K. Li, F. Zhou, and W. Yang, "Reject inference in credit scoring using semi-supervised support vector machines," *Expert Systems with Applications*, vol. 74, pp. 105–114, 2017.
- [32] Y. Tian, Z. Yong, and J. Luo, "A new approach for reject inference in credit scoring using kernel-free fuzzy quadratic surface support vector machines," *Applied Soft Computing*, vol. 73, pp. 96–105, 2018.
- [33] S. Maldonado, J. Pérez, and C. Bravo, "Cost-based feature selection for support vector machines: an application in credit scoring," *European Journal of Operational Research*, vol. 261, no. 2, pp. 656–665, 2017.
- [34] S. Maldonado, C. Bravo, J. López, and J. Pérez, "Integrated framework for profit-based feature selection and SVM classification in credit scoring," *Decision Support Systems*, vol. 104, pp. 113–121, 2017.
- [35] C. L. Huang, M. C. Chen, and C. J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Systems with Applications*, vol. 33, no. 4, pp. 847–856, 2007.
- [36] D. Martens, B. Baesens, T. Van Gestel, and J. Vanthienen, "Comprehensible credit scoring models using rule extraction from support vector machines," *European Journal of Operational Research*, vol. 183, no. 3, pp. 1466–1476, 2007.
- [37] J. Zhou and T. Bai, "Credit risk assessment using rough set theory and GA-based SVM," in *Proceedings of the 3rd International Conference on Grid and Pervasive Computing Symposia/Workshops (GPC)*, pp. 320–325, IEEE, 2008.
- [38] D. Zhang, M. Hifi, Q. Chen, and W. Ye, "A hybrid credit scoring model based on genetic programming and support vector machines," in *Proceedings of the 4th International Conference on Natural Computation (ICNC)*, vol. 7, pp. 8–12, 2008.
- [39] P. Yao, "Hybrid classifier using neighborhood rough set and SVM for credit scoring," in *Proceedings of the International Conference on Business Intelligence and Financial Engineering (BIFE)*, pp. 138–142, IEEE, 2009.
- [40] X. Xu, C. Zhou, and Z. Wang, "Credit scoring algorithm based on link analysis ranking with support vector machine," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2625–2632, 2009.
- [41] L. Zhou, K. K. Lai, and J. Yen, "Credit scoring models with AUC maximization based on weighted SVM," *International Journal of Information Technology & Decision Making*, vol. 8, no. 4, pp. 677–696, 2009.
- [42] L. Zhou, K. K. Lai, and L. Yu, "Credit scoring using support vector machines with direct search for parameters selection," *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, vol. 13, no. 2, pp. 149–155, 2009.
- [43] F.-L. Chen and F.-C. Li, "Combination of feature selection approaches with SVM in credit scoring," *Expert Systems with Applications*, vol. 37, no. 7, pp. 4902–4909, 2010.
- [44] L. Yu, X. Yao, S. Wang, and K. K. Lai, "Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15392–15399, 2011.
- [45] D. West, "Neural network credit scoring models," *Computers & Operations Research*, vol. 27, no. 11–12, pp. 1131–1152, 2000.
- [46] A. B. Hens and M. K. Tiwari, "Computational time reduction for credit scoring: an integrated approach based on support vector machine and stratified sampling method," *Expert Systems with Applications*, vol. 39, no. 8, pp. 6774–6781, 2012.
- [47] P. Danenas and G. Garsva, "Credit risk evaluation modeling using evolutionary linear SVM classifiers and sliding window approach," *Procedia Computer Science*, vol. 9, pp. 1324–1333, 2012.
- [48] W. Chen, G. Xiang, Y. Liu, and K. Wang, "Credit risk evaluation by hybrid data mining technique," *Systems Engineering Procedia*, vol. 3, pp. 194–200, 2012.
- [49] M. Y. Chen, C. C. Chen, and J. Y. Liu, "Credit rating analysis with support vector machines and artificial bee colony algorithm," in *Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 528–534, Springer, Berlin, Heidelberg, Germany, 2013.

- [50] L. Han, L. Han, and H. Zhao, "Orthogonal support vector machine for credit scoring," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 2, pp. 848–862, 2013.
- [51] G. Garsva and P. Danenas, "Particle swarm optimization for linear support vector machines based classifier selection," *Non-linear Analysis: Modelling and Control*, vol. 19, no. 1, pp. 26–42, 2014.
- [52] P. Danenas and G. Garsva, "Selection of support vector machines based classifiers for credit risk domain," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3194–3204, 2015.
- [53] F. J. Hsu, M. Y. Chen, and Y. C. Chen, "The human-like intelligence with bio-inspired computing approach for credit ratings prediction," *Neurocomputing*, vol. 279, pp. 11–18, 2018.
- [54] S. Jadhav, H. He, and K. Jenkins, "Information gain directed genetic algorithm wrapper feature selection for credit rating," *Applied Soft Computing*, vol. 69, pp. 541–553, 2018.
- [55] D. Wang, Z. Zhang, R. Bai, and Y. Mao, "A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring," *Journal of Computational and Applied Mathematics*, vol. 329, pp. 307–321, 2018.
- [56] L. Zhou, K. K. Lai, and L. Yu, "Least squares support vector machines ensemble models for credit scoring," *Expert Systems with Applications*, vol. 37, no. 1, pp. 127–133, 2010.
- [57] A. Ghodselahei, "A hybrid support vector machine ensemble model for credit scoring," *International Journal of Computer Applications*, vol. 17, no. 5, pp. 1–5, 2011.
- [58] L. Yu, R. Zhou, L. Tang, and R. Chen, "A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data," *Applied Soft Computing*, vol. 69, pp. 192–202, 2018.
- [59] Y. Xia, C. Liu, B. Da, and F. Xie, "A novel heterogeneous ensemble credit scoring model based on bstacking approach," *Expert Systems with Applications*, vol. 93, pp. 182–199, 2018.
- [60] V. S. Desai, D. G. Conway, J. N. Crook, and G. A. Overstreet, "Credit-scoring models in the credit-union environment using neural networks and genetic algorithms," *IMA Journal of Management Mathematics*, vol. 8, no. 4, pp. 323–346, 1997.
- [61] C.-S. Ong, J.-J. Huang, and G.-H. Tzeng, "Building credit scoring models using genetic programming," *Expert Systems with Applications*, vol. 29, no. 1, pp. 41–47, 2005.
- [62] S. M. Finlay, *Using Genetic Algorithms to Develop Scoring Models for Alternative Measures of Performance*, Springer, Lancaster, England, UK, 2006.
- [63] W. Cai, X. Lin, X. Huang, and H. Zhong, "A genetic algorithm model for personal credit scoring," in *Proceedings of the International Conference on Computational Intelligence and Software Engineering (CiSE)*, pp. 1–4, IEEE, 2009.
- [64] J. J. Huang, G. H. Tzeng, and C. S. Ong, "Two-stage genetic programming (2SGP) for the credit scoring model," *Applied Mathematics and Computation*, vol. 174, no. 2, pp. 1039–1053, 2006.
- [65] H. A. Abdou, "Genetic programming for credit scoring: the case of Egyptian public sector banks," *Expert Systems with Applications*, vol. 36, no. 9, pp. 11402–11417, 2009.
- [66] G. Dong, K. K. Lai, and L. Zhou, "Simulated annealing based rule extraction algorithm for credit scoring problem," in *Proceedings of the International Joint Conference on Computational Sciences and Optimization (CSO)*, vol. 2, pp. 22–26, IEEE, 2009.
- [67] D. Martens, T. Van Gestel, M. De Backer, R. Haesen, J. Vanthienen, and B. Baesens, "Credit rating prediction using ant colony optimization," *Journal of the Operational Research Society*, vol. 61, no. 4, pp. 561–573, 2010.
- [68] X.-S. Yang, S. Deb, M. Karamanoglu, and X. He, "Cuckoo search for business optimization applications," in *Proceedings of the National Conference on Computing and Communication Systems (NCCCS)*, pp. 1–5, IEEE, 2012.
- [69] J. Uthayakumar and T. Vengattaraman, "Swarm intelligence based classification rule induction (CRI) framework for qualitative and quantitative approach: an application of bankruptcy prediction and credit risk analysis," *Journal of King Saud University - Computer and Information Sciences*, 2017.
- [70] M. Jiang, F. Ji, and R. Li, "The applied research of credit scoring combination model based on SA-GA algorithm," in *Proceedings of the Fourth International Conference on Business Intelligence and Financial Engineering (BIFE)*, pp. 491–494, 2011.
- [71] R. Aliehyaei and S. Khan, "Ant colony optimization, genetic programming and a hybrid approach for credit scoring: a comparative study," in *Proceedings of the 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, pp. 1–5, IEEE, 2014.
- [72] T. C. Fogarty and N. S. Ireson, "Evolving bayesian classifiers for credit control—a comparison with other machine-learning methods," *IMA Journal of Management Mathematics*, vol. 5, no. 1, pp. 63–75, 1993.
- [73] D. J. Hand and N. M. Adams, "Defining attributes for scorecard construction in credit scoring," *Journal of Applied Statistics*, vol. 27, no. 5, pp. 527–540, 2000.
- [74] Z. Drezner, G. A. Marcoulides, and M. H. Stohs, "Financial applications of a Tabu search variable selection model," *Journal of Applied Mathematics and Decision Sciences*, vol. 5, no. 4, pp. 215–234, 2002.
- [75] E. Lacerda, A. C. Carvalho, A. P. Braga, and T. B. Ludermir, "Evolutionary radial basis functions for credit assessment," *Applied Intelligence*, vol. 22, no. 3, pp. 167–181, 2005.
- [76] S.-A. Liu, Q. Wang, and S. Lv, "Application of genetic programming in credit scoring," in *Proceedings of the Chinese Control and Decision Conference (CCDC)*, pp. 1106–1110, IEEE, China, 2008.
- [77] S. Wang, S. Yin, and M. Jiang, "Hybrid neural network based on GA-BP for personal credit scoring," in *Proceedings of the 4th International Conference on Natural Computation (ICNC)*, vol. 3, pp. 209–214, IEEE, 2008.
- [78] D. Zhang, S. C. Leung, and Z. Ye, "A decision tree scoring model based on genetic algorithm and k-means algorithm," in *Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology (ICCIT)*, vol. 1, pp. 1043–1047, IEEE, 2008.
- [79] J. Yi, "Credit scoring model based on the decision tree and the simulated annealing algorithm," in *Proceedings of the World Congress on Computer Science and Information Engineering (WCI)*, vol. 4, pp. 18–22, IEEE, 2009.
- [80] Y. Marinakis, M. Marinaki, M. Doumpos, and C. Zopounidis, "Ant colony and particle swarm optimization for financial classification problems," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10604–10611, 2009.
- [81] M. Šušteršič, D. Mramor, and J. Zupan, "Consumer credit scoring models with limited data," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4736–4744, 2009.
- [82] M. Marinaki, Y. Marinakis, and C. Zopounidis, "Honey bees mating optimization algorithm for financial classification problems," *Applied Soft Computing*, vol. 10, no. 3, pp. 806–812, 2010.
- [83] S.-C. Huang and C.-F. Wu, "Customer credit quality assessments using data mining methods for banking industries,"

- African Journal of Business Management*, vol. 5, no. 11, pp. 4438–4445, 2011.
- [84] A. B. Correa and A. M. Gonzalez, “Evolutionary algorithms for selecting the architecture of a MLP neural network: a credit scoring case,” in *Proceedings of the 11th IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 725–732, IEEE, 2011.
 - [85] S. Oreski, D. Oreski, and G. Oreski, “Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment,” *Expert Systems with Applications*, vol. 39, no. 16, pp. 12605–12617, 2012.
 - [86] S. Oreski and G. Oreski, “Genetic algorithm-based heuristic for feature selection in credit risk assessment,” *Expert Systems with Applications*, vol. 41, no. 4, pp. 2052–2064, 2014.
 - [87] J. Wang, K. Guo, and S. Wang, “Rough set and tabu search based feature selection for credit scoring,” *Procedia Computer Science*, vol. 1, no. 1, pp. 2425–2432, 2010.
 - [88] J. Wang, A.-R. Hedar, S. Wang, and J. Ma, “Rough set and scatter search metaheuristic based feature selection for credit scoring,” *Expert Systems with Applications*, vol. 39, no. 6, pp. 6123–6128, 2012.
 - [89] B. Waad, B. M. Ghazi, and L. Mohamed, “A new feature selection technique applied to credit scoring data using a rank aggregation approach based on: optimization, genetic algorithm and similarity,” in *Knowledge Discovery Process and Methods to Enhance Organisational Performance*, K. Osei-Bryson and C. Barclay, Eds., pp. 347–376, CRC Press, Boca Raton, Fla, USA, 2014.
 - [90] V. Krishnaveni, K. G. Santhiya, M. P. Ramesh, and M. S. Jaganathan, “A credit scoring prediction model based on harmony search based 1-NN classifier feature selection approach,” *International Journal of Advance Research in Science and Engineering*, vol. 7, no. 2, pp. 468–476, 2018.
 - [91] V. N. Vapnik, *Statistical Learning Theory*, Wiley- Interscience, New York, NY, USA, 1998.
 - [92] Q. Gu and J. Han, “Clustered support vector machines,” *Artificial Intelligence and Statistics*, pp. 307–315, 2013.
 - [93] D. Martens, J. Huysmans, R. Setiono, J. Vanthienen, and B. Baesens, “Rule extraction from support vector machines: an overview of issues and application in credit scoring,” in *Rule Extraction from Support Vector Machines*, vol. 80, pp. 33–63, Springer, Berlin, Heidelberg, Germany, 2008.
 - [94] M. Lichman, *UCI Machine Learning Repository*, School of Information and Computer Science, University of California, Irvine, CA, USA, <http://archive.ics.uci.edu/ml/>.

