



# Credit risk evaluation: a comprehensive study

Arijit Bhattacharya<sup>1</sup>  · Saroj Kr. Biswas<sup>2</sup> · Ardhendu Mandal<sup>3</sup>

Received: 17 February 2022 / Revised: 15 August 2022 / Accepted: 12 September 2022 /

Published online: 4 October 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

To date, there has been relatively little research in the field of credit risk analysis that compares all of the well known statistical, optimization technique (heuristic methods) and machine learning based approaches in a single article. Review on credit risk assessment using sixteen well-known approaches has been conducted in this work. The accuracy of the machine learning approaches in dealing with financial difficulties is superior to that of traditional statistical methods, especially when dealing with nonlinear patterns, according to the findings. Hybrid or Ensemble algorithms, on the other hand have been found to outperform their traditional counterparts – standalone classifiers in the vast majority of situations. Finally, the paper compares the models with nine machine learning classifiers utilizing two benchmark datasets. In this study, we have encountered with 46 datasets, among them 35 datasets have been utilized for once; whereas among the other 11 datasets, Australian, German and Japanese are the three most frequently utilized datasets by the researchers. The study showed that the performance of ensemble classifiers were very much significant. As per the experimental result, for both datasets ensemble classifiers outperformed other standalone classifiers which validate with the prior research also. Although some of these approaches have a high level of accuracy, additional study is required to discover the right parameters and procedures for better outcomes in a transparent manner. Additionally this study is a valuable reference source for analyzing

---

✉ Arijit Bhattacharya  
barijit@hotmail.com

Saroj Kr. Biswas  
bissarojkum@yahoo.com

Ardhendu Mandal  
am.csa.nbu@gmail.com

<sup>1</sup> Department of Computer Science, Gour Mahavidyalaya, Malda, West Bengal, India

<sup>2</sup> Department of Computer Science & Engineering, National Institute of Technology, Silchar, Assam, India

<sup>3</sup> Department of Computer Science & Technology, University of North Bengal, Darjeeling, West Bengal, India

credit risk for both academic and practical domains, since it contains relevant information on the most major machine learning approaches employed so far.

**Keywords** Credit risk assessment · Statistical methods · Machine learning models · Optimization techniques

## 1 Introduction

Lending credit may be the oldest but still considered the major source of income in economic sector. Financial institutes make a significant amount of income through it; however it is considered as one of the most risk-associated businesses. The impact of credit risk can be visualized by the UN Conference on Trade and Development (UNCTAD) 2020 report. As per the report, financial corporation's owe to banks and other creditors indebtedness comes to around \$75 trillion globally, which is considerably up from \$45 trillion in 2008 [32].

A performance-based strategy that would reduce the risk factors associated with credit, at the same time would also balance profitability, and the security of the financial institution would surely be the choice of any progressive institution [52]. For banking supervision to all financial institutions, the Basel Committee recommended that it must implement powerful credit scoring systems to estimate their credit risk levels and different risk factors and improve capital allocation and credit pricing [30]. The most conventional and traditional approach for financial institutions like banks for credit risk assessment is to generate an internal rating of the lender. Usually, it considers various types of quantitative and subjective factors, like lender profile, earnings, etc., through a manual credit scoring system. Credit scoring was originally initiated with personal experiences, but later on, to reduce human biasness, it changed based on 5C's: the character of the assessed, the collateral, the capital, the capacity, and the financial conditions [3]. With the increase of applicants and its heterogeneous type; it's almost unthinkable to do it manually in present time. Financial institutions providing credit, counts on credit scorings to make critical investment decisions. It means the accuracy of the credit scoring model is critical for any financial institution's profitability. Even a mere change in the accuracy of credit scoring of applicants can decrease a certain loss for the financial institutions. The first credit scoring model was designed by Altman [2]. Since then organizations in the credit industry have been developing new models that support credit decisions. The objective of these new credit scoring models is to improve the accuracy, that means more credit-worthy applicants are assisted with credit, and consequently increasing profits. Based on statistical and intelligent system, the automated credit evaluation system can be used to predict with higher accuracy level.

Machine learning, a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Different strategic methods have been implemented over time. Supervised and unsupervised learning methods are the most common. Supervised machine learning techniques have been used predominantly to find a relation between the characteristic of a lender and potential failures. Widely applied credit risk measurement methods are based on optimization techniques, statistical methods, and machine learning models with their hybrid counterparts. This study aims to conduct an overall revision of the credit risk evaluation techniques and their implementations from time to time. However, the research area is specified only to all scientific research in the academic field that includes applications of optimization techniques, statistical methods, and machine learning techniques to assess credit risk. The goal of this

research is to analyze and examine the most recent machine learning algorithms and other techniques for credit risk analysis, as well as classify them based on their performance. The study makes an attempt to report on various credit risk evaluation models suggested by scholars from time to time. The paper also delves into the most recent machine learning techniques to see how well it performs. Finally, the article conducts an analytical study on the cited publications in order to draw some conclusions and determines the main tendencies of future research.

The following is how the paper is arranged. In Section 2, introduces credit risk with its types followed by related works in credit risk assessment. Section 3 provides a brief description of statistical methods employed in credit risk assessment. Section 4 provides description of machine learning based methods employed in credit risk assessment. Optimization technique based models have been explored in Section 5. Section 6 discusses the hybrid and ensemble methods utilized in this domain. Section 7 and 8 covers the experimental works, and result & discussion respectively. Section 9 then follows with the conclusion.

## 2 Credit risk and different assessment methods

Credit risk is the danger of a bank borrower failing to satisfy its commitments according to agreed upon conditions. The most significant risk that financial institutions face is credit risk [68]. The Basel Accord permits banks to employ a credit risk management approach based on internal ratings. For assessing projected loss, banks can design their own credit risk models internally. Probability of Default (PD), Loss Given Default (LGD), and Exposure at Default (EAD) are the essential risk metrics to assess. Expected loss for any financial institution is calculated by multiplying PD, LGD, and EAD. In the risk management arena, there are several forms of credit risk. In general, credit risks can be categorized as Credit Default Risk, Concentration Risk, Country Risk, and Counterparty Risk [51]. Following are some of them.

**Credit Default Risk** - A type of credit risk in which a bank suffers losses when a borrower fails to meet his or her obligations in terms of interest or principal for a period of time. All credit-sensitive transactions, such as loans, securities, and derivatives, may be affected by default risk.

**Concentration Risk** - The risk associated with any one or set of exposures that have the potential to cause big enough losses to jeopardize a bank's main activities. It might manifest as a single name concentration or an industry focus.

**Country Risk** - The risk of loss resulting from a sovereign state by freezing foreign currency payment (transfer/conversion risk) or defaulting on its commitments (sovereign risk) known as country risk.

**Counterparty Credit Risk (CCR)** - CCR is the risk that a transaction's counterparty may default before the ultimate settlement of the transaction's cash flows. If the transaction or portfolio of transactions with the counterparty has a positive economic value at the moment of default, an economic loss will result. Unlike a firm's exposure to credit risk through a loan, which is unilateral and only the lending bank bears the chance of loss, CCR generates a bidirectional risk of loss: the market value of the transaction might be positive or negative to either counterparty to the transaction. The market value is speculative and might fluctuate over time according to the movement of underlying market conditions.

**Table 1** Comparison of the proposed survey with existing surveys

| Paper        | Year | Evaluation Model   | Dataset details | predictor | Performance | Software Platform | Validation methods | Remarks |
|--------------|------|--|-----------------|-----------|-------------|-------------------|--------------------|---------|
| [6]          | 2001 | ANN  | x               | x         | ✓           | x                 | x                  | x       |
| [31]         | 2006 | Different existing models by different firms   | x               | x         | x           | x                 | x                  | x       |
| [10]         | 2010 | ANN, Expert System and Hybrid Intelligent System   | x               | x         | ✓           | x                 | x                  | ✓       |
| [22]         | 2016 | Statistical Models, Semi Parametric method, ANN, SVM, DT, Case Based Reasoning, Hybrid System, and Multiple Classifier System. | x               | x         | ✓           | x                 | ✓                  | ✓       |
| Present Work | 2022 | LDA, LA, HMM, DT, ANN, k-NN, SVM, BN, k-means, GA, ACO, SA and Hybrid & Ensemble Learning Algorithms.                          | ✓               | ✓         | ✓           | ✓                 | ✓                  | ✓       |

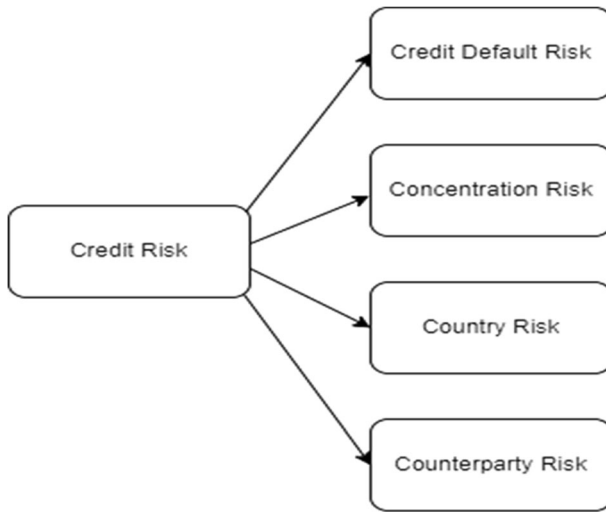


Fig. 1 Types of credit risk

Among the four different types of risk, credit default risk has been given major focus in this research work. Over the time, several survey papers on credit risk assessment had been carried out with their pros and cons. To establish the effectiveness of the present research work, comparison has been made with the previous survey works and it's been observed that numerous issues arose that need to be carefully evaluated in order to make better decisions. Comparison of previous survey works has been shown in Table 1, which clearly describes the effectiveness of the work (Fig. 1).

**2.1 Different assessment methods**

Several approaches based on statistical, optimization technique, and machine learning techniques have been employed to assess credit risk for a long time. Some of the most well-known statistical techniques include Logit Analysis (LA), Hidden Markov Model (HMM) etc. Invariably most utilized optimization techniques are Genetic Algorithm (GA), Ant Colony Optimization (ACO), and Simulated Annealing (SA). Artificial Neural Network (ANN), Support Vector Machine (SVM), Bayesian Network (BN), k-Nearest Neighbor (k-NN),

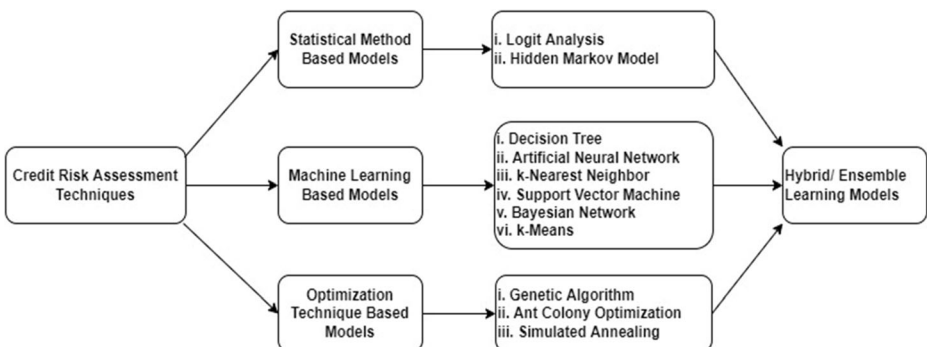


Fig. 2 Different Model based credit risk assessment techniques

Decision Tree (DT), k-means etc. are undoubtedly the most entrusted machine learning algorithms. For Hybrid algorithms, different combination of statistical, optimization technique and machine learning techniques have been utilized. Ensemble learning includes Random Forest (RF), Adaboost, Xgboost etc. The following section is the synopsis of previous research done by various researchers in the field of credit risk assessment. Figure 2 represents all of these aforementioned methods categorically.

### 3 Statistical methods

The traditional statistical models comprise Logit Analysis (LA), and Hidden Markov Model (HMM). The rationale behind statistical models is to find an optimal linear combination of explanatory input variables able to model, analyze and predict default risk. But unfortunately, they tend to overlook the complex nature, boundaries and interrelationships of the financial variables, due to some strict assumptions such as linear separability, multivariate normality, independence of the predictive variables and pre-existing functional form. However, statistical models still belong to the most popular tools for some famous institutions. LA and HMM have been analyzed here in view of credit risk and to make better decision.

#### 3.1 Credit risk assessment using Logit analysis

Logit Analysis (LA) is a commonly used approach, in which the likelihood of a dichotomous result is linked in the following way with a series of possible predictor variables. It transforms the input data nonlinearly, reducing the effect of outliers.

$$\log \left[ \frac{p}{1-p} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

where  $p$  is the probability of the trade credit risk occurrence,  $\beta_0$  is the intercept term, and  $\beta_i$  ( $i = 1, \dots, n$ ) represents the  $\beta$  coefficient associated with the corresponding explanatory variable  $x$  ( $i = 1, \dots, n$ ) [9].

Miller and LaDue [59] employed a logit model to link loan default to dairy farm debtors' financial measures. The accuracy rate in this classification case was 85.7%, which was significantly higher than the prior researches. Liquidity, Profitability and Operating efficiency were found the most suitable predictors for their study. Chi and Tang [23] also utilized LA to build a credit risk model based on data from seven Asia Pacific capital markets from 2001 to 2003. The classification accuracy of the model has been affected by country- and industry-specific effects which results about 85% train and test accuracy. The study examines data three years before a financial crisis begins, and the sample only includes companies that have been in operation for at least three years prior to the study

data gathering. As a result, the sample selection procedure may introduce a survivorship bias, because most businesses may fail during their initial periods.

For credit risk evaluation, utilizing China's 130 Shanghai and Shenzhen stock exchange listed firms from the year 2009, Konglai and Jingjing [49] entrusted on Discriminant Analysis (DA) and Logistic Regression (LR) models. They employed Principal Component Analysis (PCA) to avoid the influence of the listed firms' credit data's high correlation and dimensionality. They utilized 6 key components from the available 14 financial ratios. As per the

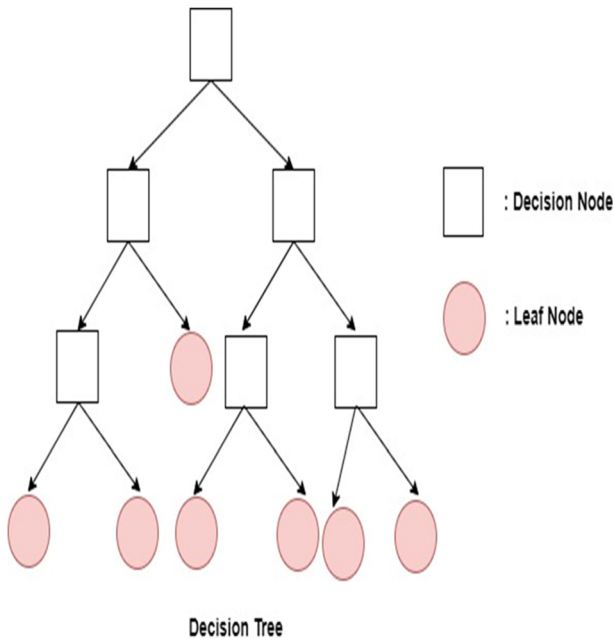
**Table 2** Comparison of existing Logit Analysis based models

| Paper | Evaluation Model         | Dataset   | Input / Output         | Dataset Year | Size of dataset | predictor | Accuracy/ Error rate  | Software |
|-------|--------------------------|---|------------------------|--------------|-----------------|-----------|---|----------|
| [59]  | Logit Model              | Agricultural loan portfolio of a case study bank in upstate New York  | Input: 23<br>Output: 1 | 1984         | 203             | 23        | Accuracy: 85.7%   | SAS      |
| [23]  | Logit Analysis           | Seven Asia Pacific capital markets (Hong Kong, Japan, Korea, Malaysia, Singapore, Thailand, and the Philippines) PACAP database | 24                     | 2001 to 2003 | 60              | 24        | Accuracy: 91.7% (training sample) and 86.1% (testing sample).   | NA       |
| [49]  | Logistic Regression & DA | Shanghai and Shenzhen stock exchange.   | 14                     | 2009         | 130             | 6         | For LDA, Average Accuracy: 74.4% (training sample) and 70.0% (test samples).<br>For LR, Average Accuracy: 87.8% (training sample) and 75.0% (test samples). | SPSS     |

**Table 3** Comparison of existing Hidden Markov based models

| Paper | Evaluation Model | Dataset                       | Input / Output | Dataset Year                 | Size of dataset                 | predictor   | Accuracy/ Error rate                        | Software |
|-------|------------------|-------------------------------|----------------|------------------------------|---------------------------------|---|---|----------|
| [63]  | HMM, LR, k-NN    | Australian and German dataset | 11             | German: 1994, Australian: NA | German: 1000<br>Australian: 690 | German: 21 attributes;<br>Australian: 14 attributes | German: 71.66%<br>Australian:<br>HMM-84.83% | NA       |





**Fig. 3** An example of a DT

experimental results, LR model outperformed the DA by a certain range based on both the average accuracy and misclassification cost. Since the borrower characteristics on the likelihood of returning or not repaying the loan is not consistent and tends to alter over time, Yurynets et al. [82] developed a credit risk assessment model using LR to reduce loan risks in the Ukrainian banking sector. They emphasized about adjusting on a regular basis to ensure the scoring model's continued operation. They used STATISTICA to create the model, which included only three independent variables and the Chi-square test was carried out to establish the validity of the three parameters. However, the author did not demonstrate the model's efficacy with other benchmark datasets and existing models to verify its productivity.

Despite its better performance on most of the occasions, flows also came up during several researchers' study. Accuracy level of the LR algorithm is directly proportional with the set of attributes which are selected as independent variables. It means accuracy level has dependency on the choice of these attributes. With better choice it can give good results but opposite outcomes are sometimes inevitable because of the poor choice of these variables. More research is needed to better understand these statistical approaches and how they may be applied to get the best outcomes in credit risk assessment situations. Table 2 shows the findings in details.

### 3.2 Credit risk assessment using Hidden Markov model

The Hidden Markov Model (HMM) is a statistical model in which a series of unobserved states generates a series of observations. It is assumed that the concealed state transitions follow a first-order Markov chain. Baum et al. [12, 13] developed the theory of HMMs in the late 1960s to demonstrate it effectively.

**Table 4** Comparison of existing Decision Tree based models

| Paper | Evaluation Model                               | Dataset                          | Attribute | Dataset Year                   | Size of dataset       | Predictor | Accuracy/ Error rate                                 | Software |
|-------|--|----------------------------------|-----------|--------------------------------|-----------------------|-----------|--|----------|
| [19]  | Logistic Regression, SMOTE based DT, Cox model | Real loan data of SMEs in Taiwan | 30        | November 2001 to December 2010 | 12,514/17174          | 13        | Recall: 88.8%  | NA       |
| [77]  | DT   | Pre-loan Survey                  | 5         | NA                             | 3422 consumer (75:25) | 5         | 81.2% (Cross-Validation) & 83.6% (Random-Validation) | SPSS     |

The performance of the HMM for credit risk analysis in terms of categorization and Probability of Default (PD) modelling was explored by Oguz and Gurgun [63]. Instead of categorizing credit customers as good or bad, the model offered them a probabilistic value. HMM's probability of default models was studied and compared with LR and k-NN methods. Experiments were conducted using six fold cross validation on two datasets, the Australian dataset and the German dataset. According to the experimental study, HMM (71.66%) came out on top with the German dataset, whereas k-NN (85%) came out superior for the Australian dataset, followed by HMM (84.83%). The findings proved that HMM's decent performance which financial institutions may employ to make smarter decisions about their consumers for credit risk analysis. Anagnostou and Kandhai [4] proposed a model that combined currency rates with a HMM to create counterparty credit risk scenarios. Apart from that, other studies on various aspects of credit risk assessment have been conducted, and it has been shown that HMM performed pretty well when compared to other statistical approaches.

Financial institutions heavily used credit risk assessment for assessing risk associated with a credit application. An automated system reads all details of the applicant and processes it to generate the risk percentage eventually to classify it as good or bad borrower. Although the quantity of research utilising HMM in this domain is quite limited but the findings are very much promising. Scope for more future work to explore still exists. Table 3 showed the details categorically.

## 4 Machine learning based methods

Machine learning models in credit risk assessment are of two types – a. Supervised and b. Unsupervised. Supervised algorithms learnt through the available data by means of some supervision; whereas, unsupervised learning carried out from the available data without any supervision. In the following section discusses the various supervised and unsupervised models employed in credit risk assessment. Application of machine learning methods such as Decision Tree (DT), Artificial Neural Network (ANN), k-Nearest Neighbour (k-NN), Support Vector Machine (SVM), Bayesian Network (BN), and K-means have been addressed in the next section.

### 4.1 Credit risk assessment using Decision Tree

A rule-based classifier such as DT [67], is a mapping of observations about an object to inferences about the item's target value. Each internal node represents a variable, and each link

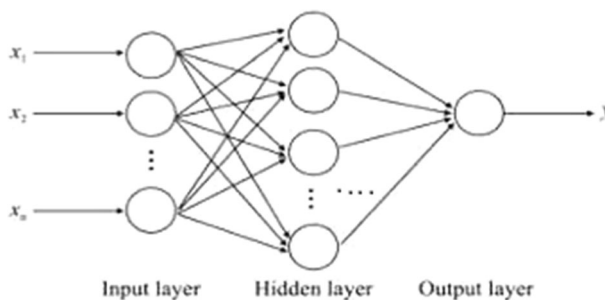


Fig. 4 An example of MLPNN

to a child represents one of the variable's potential values. A leaf indicates the destination variable's expected value given the values of the variables represented by the path from the root. In the case of learning for a DT, it depicts a tree structure in which leaves indicate classifications and branches reflect feature conjunctions that lead to those classifications. By dividing the source set into subgroups based on an attribute value test, a decision tree may be trained. This technique is done recursively on each derived subset. When splitting is not possible or when a solitary classification can be applied to each element of the resulting subset, the recursion is finished. Figure 3 depicts an example of a DT. The benefit of a DT-based learning algorithm is that it does not require users to have a lot of prior knowledge. The technique may be used as long as the training case can be presented in an attribute conclusion way [45, 71].

To measure credit risk, Chang et al. [19] developed a decision tree-based short-term default credit risk assessment model. The objective was to utilize a DT to filter short-term defaults in order to create a highly accurate model that can detect default loans. To improve the decision tree stability and performance on imbalanced data, Chawla et al. [20] incorporated Bootstrap Aggregating (Bagging) and a Synthetic Minority Over-sampling Technique (SMOTE) into the credit risk model. The model was designed to assist financial institutions assess their prospective financial losses and alter their credit policies to filter short-term defaults. The model employed 10-cross-validation and the performance was quite promising. Although the suggested model proved adequate for supporting financial institutions in making loan decisions, more work is needed for extremely low default datasets. To make the prediction more accurate, further DT was utilized by Wang and Duan [77] considering the personal information of 3422 consumers from a pre-loan survey. Verification of the model was done using the cross-validation method and the random validation method; it showed that the risk estimation accuracy was 81.2% and 83.6% respectively.

When compared to other machine learning approaches such as LR, SVM, etc. the decision tree classifier did exceptionally well. It's been observed that when the predictor variables are chosen precisely, DT performed better and all the rules generated through it are also quite simple in nature. Table 4 compared the detail of the research works categorical.

## 4.2 Credit risk assessment using Artificial Neural Network

ANN is one of the most promising machine learning approaches that has been widely employed by credit risk prediction experts. Figure 4 illustrates a basic multilayer neural network; where,  $x_1, x_2, \dots, x_n$  are the inputs for the input layer. Intermediate layers are known as hidden layers. Final layer is known as output layer. Multilayer perceptrons with back-propagation algorithms and multi-classifier-based hybrid neural networks were the most utilized algorithms. Credit risk is also evaluated using an emotional neural network [47]. The extraction of rules from ANN for credit risk assessment has also been discussed.

### 4.2.1 Credit risk assessment using conventional ANN methods

Multilayer perceptron with a back-propagation algorithm or a back-propagation neural network is the basic ANN model which is employed mostly for credit risk analysis. These procedures are referred to as typical or conventional methods in this study.

The effect of several neural network models such as Back-Propagation Neural Network (BPNN), Radial Basis Function (RBF), General Regression Neural Network (GRNN), and

**Table 5** Comparison of existing conventional Neural Network based models

| Paper | Evaluation Model                        | Dataset                                  | Attribute | Dataset Year     | Size of Dataset                            | Predictor                                | Accuracy/ Error rate  | Software                 |
|-------|---|--|-----------|------------------|--|--|---|--------------------------|
| [43]  | BPNN,RBF, GRNN,PNN                      | 46 Chinese SME's                         | NA        | NA               | NA   | 3 indicators                             | Error rate: 0.17% for PNN   | MATLAB,SPSS software     |
| [60]  | MLPNN,BP,LM, GD,CGD,R,BFGS and OS       | German, Australian and Japanese datasets | NA        | German: 1994, NA | German: 1000 Japanese: 690 Australian: 690 | German: 21; Japanese: 15; Australian: 14 | For (MLPNN):German 79% Japanese:89.28% Australian:89.71%  | MATLAB version R2012a    |
| [62]  | ANN                                     | Iranian Commercial Bank                  | 18/1      | 2006–2011        | Number of samples was 497.                 | 18                                       | Accuracy for Training dataset: 87.1% for good customer and 89.5% for bad customer. Accuracy for Testing dataset: 70% for good customer and 100% for bad customer. | SPSS and MATLAB software |
| [65]  | Feed Forward Multi Layer Neural network | CentraledeiBilanci in Turin (Italy)      | 24/1      | 1982–1992        | 1000                                       | 24                                       | Accuracy for safe cases: 84.2%; Accuracy for NA vulnerable cases: 73.9%; Accuracy for risk cases: 34.8%   |                          |

Probabilistic Neural Network (PNN) on enterprise credit risk evaluation was investigated in this article by Huang et al. [43]. The research carried out utilizing 46 Chinese small and medium-sized enterprises in the Yangtze River Delta Region, of which 21 businesses defaulted. The Area Under Curve (AUC) value in the RoC curve was utilized in this article to test the credit risk assessment model's predictive capacity. The number of neurons in the hidden layer affected the performance of the BP neural network, according to the study. The overall error rate of the test set and the Type-II errors were the least at 0.248 and 0.128, respectively, considering the number of neurons in the hidden layer 7. The credit risk assessment model based on PNN had the lowest rate of misclassification, followed by GRNN and RBF neural network. PNN produced the greatest classification results among other neural network models, as well as the highest AUC value and mean AUC in the 5 set of tests. The experimental findings showed that the model based on PNN was successful in assessing credit risk. Further, Mohammadi et al. [60], utilized the Multilayer Perceptron Neural Network (MLPNN) model, trained using various back-propagation algorithms such as Levenberg-Marquardt (LM), Gradient Descent (GD), Conjugate Gradient Descent (CGD), Resilient (R), BFGS Quasi-newton (BFGS) and One-step Secant (OS) for the purposes of better classification, and generalization in the field of credited risk assessment. The German, Australian and Japanese datasets were utilized to compare the model's performance. At first, 0.5 was viewed as a cut-off point between the two groups- good (0) & bad (1) applicant. The entire score of the applicant was compared with the cut-off point to determine the applicant's status. They utilize sigmoid and hyperbolic tangent as activation functions in their experimental setup. Mean Squared Error (MSE) was employed to determine the optimal number of neurons in the intermediate hidden layer. The highest accuracy of learning using Levenberg Marquardt (LM) Back-propagation model was recorded as 76.80%, 88.49%, and 88.31% respectively with MSE 0.1821, 0.1069, and 0.1118 for German, Japanese, and Australian datasets respectively. The accuracy of the model when assessed with LR and DA; it achieved 78.40% accuracy for the German dataset with MSE 0.1593; 89.12% accuracy for the Japanese dataset with MSE 0.1117; and 88.40% accuracy for the Australian dataset with MSE 0.0946. Although MLPNN had the best accuracy in terms of prediction, but Type II error rate was too high. Therefore RoC curve was used to determine an acceptable cut-off point for the model. After setting the cut-off point to 0.4417 for the German dataset, 0.4992 for the Japanese dataset, and 0.3621 for the Australian dataset; MLPNN achieved 79% accuracy with MSE 0.1593; 89.28% accuracy with MSE 0.1117; and 89.71% accuracy with MSE 0.0946 for the respective datasets.

As per the performance for almost all the models were quite acceptable, but identification of the categorization criteria between good and bad clients is very much essential requirement. To emphasis on this, Nazari and Alidadi [62] utilized records of applicants of Iranian commercial bank from 2006 to 2011 to find categorization criteria for good and bad clients in Iranian banks. They utilized ANN to achieve this task. The dataset was split into three sections- 70% for training, 20% for testing, and 10% for holdout. Only one hidden layer with hyper-tangent activation function was used to create the model. As per the result, in training phase, 87.1% of consumers categorized properly into good customers and 89.5% of customers classified into bad customers correctly. According to testing, the accurate anticipated percentages were 75% and 100%. The accurate projected percentages for Holdout were 80% and 75% respectively. The study also emphasized that individual loan frequency and the amount of borrowing had the major impact on identifying classification criteria of good and bad customers.

Pacelli et al. [65] compared a feed-forward multilayer neural network with another feed-forward neural network model developed previously in their study. The two neural networks are quite similar; however the activation function they used differs. Both network models were trained using the supervised back propagation technique. The approach optimized the network weights in order to reduce the difference between the desired and actual output. The weights were changed after each training pattern via an incremental method called the conventional back propagation algorithm. The logistic function was used as the activation function for the first network, whereas the sigmoid symmetric stepwise function was utilized for the network created for this study. The dataset on which the two networks were made up consists of a group of Italian manufacturing businesses companies. The dataset was split into three sections: training, validation, and test part. The firms were classified into three groups: safe, vulnerable, and risk. The training was conducted with 70% of the firms in each of the above-mentioned rating classes, with the remaining 30% of each class used for validation testing. The goal of this selection was to have uniform data in terms of classes for submission to the training stage. As per the result, the model can accurately identify 84.2% as safe, 73.9% as vulnerable and 34.8% as risk cases. Finally, the author concluded that the flexibility and objectivity of neural network models, when combined with linear techniques, can give great support for the efficiency of a bank's credit risk management procedures.

Throughout this part, traditional models have been scrutinized, and it has been observed that neural networks did quite well almost in all, excepting a few cases where further thought is required to enhance the situation. Certain situations observed where datasets weren't adequately pre-processed, or the model's performance had been evaluated with a minimal number of datasets. Discovering associated features might also assist the model to enhance its performance. Table 5 showed the comparative review in the following part. Performance of conventional model is very much promising but there exists scope of further improvement.

#### 4.2.2 Credit risk assessment using Hybrid models based on ANN

The hybrid prediction model, a combination of statistical techniques with machine learning based approaches like ANN, DT etc., has lately been noted its large applicability since it enhanced assessment power when compared to single statistical or machine learning methods. Hybridization also been used in selecting and providing the algorithm with essential input characteristics. They determine the relative relevance of each characteristic in relation to the data mining job and then rank the features accordingly. It helps the learning algorithms construct simpler models, which improves the classifier's speed and accuracy. Several researches have carried out on developing machine learning models to judge credit risks automatically. A Hybrid model was suggested by Chi et al. [24] that combine LR and MLPNN. It had higher prediction capability than any other approaches according to the research. They compared 16 hybrid models that combined LR, DA, and DT with four different types of neural networks: Adaptive Neuro-Fuzzy Inference Systems (ANFISs), Deep Neural Networks (DNNs), Radial Basis Function networks (RBFs), and Multilayer Perceptrons (MLPs). As indicated by ten different performance measures such as Accuracy, AUC, Type I error, Type II error, EMCC, G-Mean, DP, F-Score, Kappa, and Youden's Index; they showed hybrid model's capacity to develop a credit risk prediction technique different from all other approaches, was expressed in the experimental outcome. The classifier was validated using five real-world credit score data sets from a prominent public commercial bank in China, Australian, German, and Japanese, as well as two project datasets based on historical loan

**Table 6** Comparison of existing Hybrid models based on AN

| Paper | Evaluation Model       | Dataset   | Input / Output     | Dataset Year | Size of dataset   | predictor   | Accuracy/ Error rate   | Software                          |
|-------|------------------------|---|--------------------|--------------|---|---|--|-----------------------------------|
| [24]  | LR,MLP,DA,DT,ANFIS,DNN | 5 dataset; 2 private dataset are Chinese SME and Chinese Farmer. 3 public dataset are Australian, Japanese and German | No. of feature / 1 | NA           | Australian: 690<br>Japanese: 690<br>German: 1000,<br>Chinese SME: 3111<br>Chinese Farmer: 2036<br>Total: 7527 cases | Total Attributes / Selected Attributes by LR<br>Australian: 14/6,<br>Japanese: 15/7<br>German: 20/12,<br>Chinese SME: 81/14<br>Chinese Farmer: 44/6 | Chinese SME credit dataset: Accuracy of MLP4: 99.73% and MLP1: 99.76%.<br>Chinese Farmer dataset: Accuracy of MLP4: 99.73% and DT based model: 99.42%.<br>Australian credit dataset: Accuracy of MLP4: 96.51 % and MLP1: 94.76%<br>German credit dataset: Accuracy of MLP4: 96.34% and MLP1: 93.98%<br>Japanese credit dataset: Accuracy of MLP4: 98.07% and MLP1: 98.17%. | Matlab (version 1.2.0.0 and 1.19) |
| [69]  | CRJ,SVM                | Australian, German and Kaggle dataset   | NA                 | 2007–09      | Australian: 690<br>German: 1000,<br>Kaggle: 150,000   | Australian: 15 mixed attributes;<br>German dataset: 20 input variable;<br>Kaggle dataset: 59 attributes.  | Accuracy of Australian dataset: 93.6%<br>Accuracy of German dataset: 82.32%<br>Accuracy of Kaggle dataset:94.2%  | NA                                |



**Table 7** Comparison of existing models based on Emotional Neural Network

| Paper | Evaluation Model         | Dataset                             | Input / Output | Dataset Year | Size of dataset | predictor | Accuracy/ Error rate  | Software |
|-------|--------------------------|-------------------------------------|----------------|--------------|-----------------|-----------|---|----------|
| [28]  | EmNNs, & Conventional NN | Australian credit approval datasets | 14/1           | NA           | 690             | 14        | Accuracy of EmNN-1: 99% for training dataset, 81.03% for validation dataset and 88.84% for overall. | NA       |

information. They utilized a three-layer feed-forward neural network in this experiment. The Back-propagation method was used to train the network. They used sigmoid activation function in their network. Finally, the author suggested that MLP4 (LR plus MLP) can be utilized in banks and financial institutions since it performed the best in their tests. Uddin [76] also validated the findings in another research.

For categorizing credit approval requests, a unique hybrid technique based on a neural network model called Cycle Reservoir with Regular Jumps (CRJ) and Support Vector Machines (SVM) was suggested in this study by Rodan and Faris [69]. Rather than utilizing LR, which considered as the traditional method of training in the reservoir computing community, CRJ's readout learning was taught using SVM. It has the benefit of addressing a quadratic optimization problem with linear circumstances. The experiments were run on three prominent datasets: the German, Australian and the 'Give me some credit' dataset from the Kaggle. All datasets were normalized to eliminate the effects of various feature sizes. The following common machine learning methods were compared to CRJ with SVM-readout model: Classical SVM, MLPNN, k-NN, Naive Bayes, Decision Trees C4.5, Bagging, AdaBoost, and Random Forest. The number of neurons in the hidden layer of MLPNN was set to six. The number of neighbours was fixed to 10 in k-NN. 10-fold cross validation method was employed for the German and Australian datasets. On the other hand, Kaggle dataset was split into 80% training and 20% testing data ratio. The Australian, German, and Kaggle datasets showed that CRJ with SVM readout surpassed all other models, with accuracy of 93.6%, 82.32%, and 94.2%, respectively. In addition, the suggested model had the highest specificity. In terms of sensitivity, it achieved fifth place for the Australian and German datasets, and second place in the kaggle dataset.

In contrast to traditional neural network models, hybrid neural network models performed reasonably well. On diverse datasets, these models were compared to several current machine learning-based models, and the results revealed that hybrid neural networks outperformed other existing models. The hybrid model worked very well for non-linear models and, in particular, datasets with a large variety of features. Using the hybridization approach to discover relevant parameters may also help the model improve its performance. As per the comparative review shown in Table 6, the performances of hybrid methods are quite promising, but there is still room for improvement.

#### 4.2.3 Credit risk assessment using Emotional Neural Network

Credit applications differ slightly from one applicant to the next, which led to this method. The decision to approve or deny a credit application is normally determined by an expert. Using traditional neural networks to make a choice can be effective, but it lacks the human emotional aspects. With applications such as credit assessment, research has been done to combine emotional variables, as modelled in Emotional Neural Networks. Khashman [47] examined the efficacy of Emotional Neural Networks (EmNNs) and compared their performance with the traditional neural networks to assess credit risk. To classify whether a credit application will be granted or denied, 12(six emotional and six conventional) neural networks were evaluated. The Australian credit approval dataset was used to train and test the emotional and conventional neural models. Three learning techniques to train the neural network models were utilized in order to determine the optimal training to validation data ratio. Learning schemes- LS1, LS2, and LS3 were represented with 43.5:56.5, 50:50, and 56.5:43.5

**Table 8** Comparison of existing models exploring rule extraction from ANN

| Paper | Evaluation Model     | Dataset   | Input / Output   | Dataset Year | Size of dataset  | predictor  | Accuracy/ Error rate   | Software                             |
|-------|----------------------|---|--|--------------|--|--|--|--------------------------------------|
| [7]   | RxREN                | German credit   | crediting: 21/2; Australian  | NA           | crediting: 1000 German: 1000, Australian credit approval:690,      | crediting: 20 Australian Credit Approval: 14, German: 20 | crediting (72.2%). 78.26% for Australian credit approval, 70% for German             | JDK1.5<br>MATLAB 7.14.0.739 (R2012a) |
| [16]  | RxNCM, RxREN         | 9 datasets from the UCI repository: German credit data, Australian Credit Approval.                         | Credit Approval: 14/1, German: 20/1, Credit Approval: 47/1, Australian Credit Approval: 39/1 | NA           | Credit Approval: 690, Australian Credit Approval: 690              | Credit Approval: 14, Australian Credit Approval: 14      | 75.57%, Australian credit approval: 74.64%.  | MATLAB                               |
| [17]  | RE-Re-RX             | Six datasets from UCI repository. -Credit Approval, Australian Credit Approval                              | NA   | NA           | Credit approval: 690, Australian credit approval:690, German: 1000 | Credit approval: 16, Australian credit approval: 14      | 86.62% for credit approval, 85.51% for Australian credit approval, 73.1% for German. | MATLAB                               |
| [18]  | ERENNR, Re-RX, RxREN | Dataset from UCI and Keel machine learning repository- Credit approval, Australian credit approval, German. | NA   | NA           | Credit approval: 690, Australian credit approval:690, German: 1000 | Credit approval: 16, Australian credit approval: 14      | 86.62% for credit approval, 85.51% for Australian credit approval, 73.1% for German. | MATLAB                               |

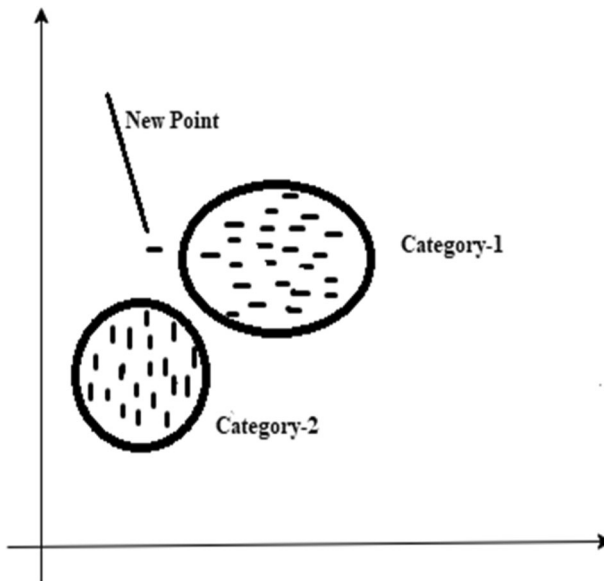
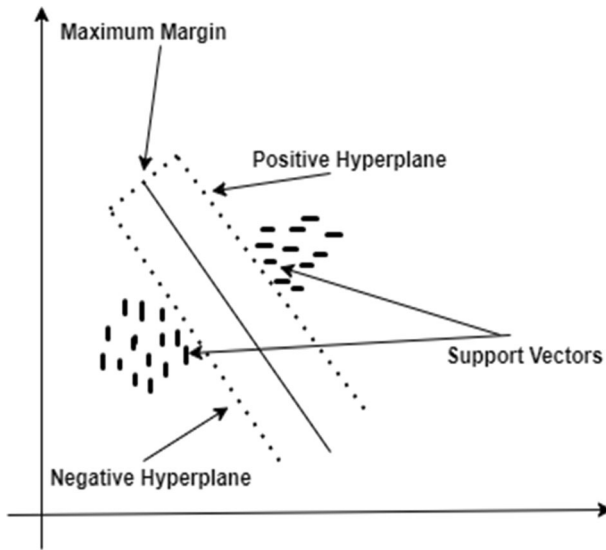


Fig. 5 An example of k-NN

respectively. Among them LS1 worked best, with a training dataset accuracy rate of 99%, validation dataset accuracy rate of 81.03%, and overall accuracy rate of 88.84%. According to the findings, EmNN-1, out of the six emotional neural models, had the lowest level of anxiety and also had the least amount of inaccuracy. The author found that emotional models outperformed their traditional equivalents in terms of decision-making speed and accuracy, making them appropriate for use in automated credit application processing. Table 7 pointed the performance of the EmNN model.

#### 4.2.4 Credit risk assessment using rule extraction from ANN

ANN generally attain higher classification accuracy rates; however, because of their lack of explanation capabilities, they are regarded as black boxes. To address this issue, Augusta and Kathirvalavakumar [7] presented RxREN, a novel rule extraction technique. Rule extraction can be classified into three categories - Decompositional, Pedagogical, Eclectical. Decompositional approaches refer to rule extraction algorithms that function on the neuron level rather than the entire neural network architectural level. If an artificial neural network is seen as a black box, regardless of design, then these algorithms are classified as pedagogical. The third method is a hybrid of decompositional and pedagogical approaches. It's referred to as eclectic method. The proposed technique derives rules from trained neural networks for datasets with mixed-mode characteristics in the pedagogical approach. To prune the unimportant input neurons and identify the technical principles of each significant input neuron of a neural network in classification, the method used reverse engineering approach. RxREN acquired accuracy by doing a 10-fold cross-validation test on German credit (72.2%). In terms of accuracy and comprehensibility, the suggested method RxREN compared to five current rule extraction techniques, and it showed that the proposed approach extracts the smallest



**Fig. 6** An example of SVM

amount of rules with higher level of accuracy. Another variation of the RxREN, Rule Extraction from Neural Network Using Classified and Misclassified Data (RxNCM) technique was proposed by Biswas et al. [16]. Using a pedagogical approach, the proposed technique extracts rules from the trained neural network for datasets with mixed-mode characteristics. Nine datasets from the UCI repository were used to compare the proposed method to RxREN. Accuracy, Precision, FP-Rate, Recall, f-Measure, and MCC were six performance metrics used in their study. According to the results of the experiment, accuracy on the nine datasets was 78.26% for Australian credit approval, 66% for German. RxNCM definitely surpassed RxREN by a significant margin, making it superior to existing algorithms, according to the result.

For symbolic rule extraction from neural networks with mixed characteristics, Chakraborty et al. [17] presented a method called Reverse Engineering Recursive Rule Extraction (RE-Re-RX), an expansion of Recursive Rule Extraction (Re-RX) [72]. To deal with nonlinearity, Re-RX produced a linear hyper-plane for continuous characteristics. A simple rule for continuous attributes was generated using the RE-ReRX algorithm to circumvent this restriction. RE-Re-RX generates rules by reverse engineering the NN (RxREN) algorithm. To compute input data ranges for rules, RxREN exclusively utilized misclassified patterns, while RE-Re-RX used both classified and misclassified patterns for each continuous characteristic. There were six benchmark datasets used to validate the proposed algorithm. It also used six performance metrics like the previous article. The accuracy for 80/20 fold Credit approval dataset was 75.57%, Australian credit approval was 74.64%. The study showed RE-Re-RX beats Re-RX by a wide margin.

With the goal of generating simple and accurate rules, Chakraborty et al. [18] introduced the Eclectic, Rule Extraction from Neural Network Recursively (ERENNR), rule extraction technique. The ERENNR method used a single-layer feed-forward neural network to derive symbolic classification rules. It analysed a hidden node using logical combinations of hidden node outputs with regard to output class, and it analysed an output node using data ranges of

input attributes with respect to its output. Finally, beginning from the output layer, it produced a rule set in a backward manner. Eleven benchmark datasets from the UCI and Keel machine learning repositories were used to verify the method. For the eleven datasets, the proposed algorithm had an accuracy of 86.62% for credit approval, 85.51% for Australian credit approval, 73.1% for German. Additionally, the article assessed performance utilizing measures such as Recall, FP rate, Specificity, Precision, f-measure as well as the MCC. For neural networks with a single hidden layer, the produced rules were straightforward and accurate, according to the results. However, the technique yielded no results for deep neural networks with more hidden layers.

As per the study, neural networks have attained the best classification accuracy for nearly all classification tasks; nevertheless, the generated results may not be interpretable because they are frequently regarded as a black box. To address this shortcoming, researchers have created a variety of rule extraction methods. This section compares four types of rule extraction models that take into account all sorts of rule extraction approaches, namely decompositional, pedagogical, and eclectic. According to the comparison in Table 8, ERENNR constructs basic rules with high accuracy and fidelity, although the model fails for deep neural networks with more hidden layers. It allows researchers to concentrate on rule extraction approaches for deep neural networks.

The subject of credit risk assessment using NN was discussed. According to the study, overall the performance of NN typically better than alternative approaches, although the model of choice is dependent on the dataset provided. Various performance methods, including Accuracy, Precision, FP-Rate, Recall, f-Measure, MCC etc. were employed to assess the model's performance. According to the findings, the accuracy of the NN in dealing with financial issues is very much promising in contrast to traditional statistical methods, particularly when dealing with nonlinear patterns. Hybrid algorithms, on the other hand, have been shown to beat conventional algorithms in the great majority of cases. Although some of these methods are highly accurate, further research is needed to determine the best parameters and techniques for achieving better results in a transparent manner.

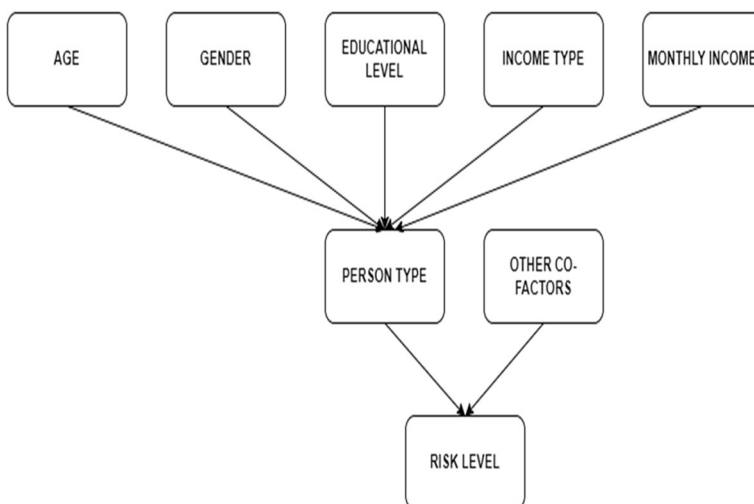


Fig. 7 An example of BN

### 4.3 Credit risk assessment using k-Nearest Neighbor

k-Nearest Neighbor (k-NN) is an example of lazy learning algorithm. It is a non parametric technique; it means that it does not make any assumptions on the underlying data distribution. This is very useful, as in the real world most of the practical data does not obey the typical theoretical assumptions made. The k-nearest neighbor classifier is commonly based on the Euclidean distance between a test sample and the specified training samples. Figure 5 demonstrates two different category points. The main idea of k-NN algorithm is that whenever there is a new point to predict, its k nearest neighbors are chosen from the training data. Then, the prediction of the new point can be the average of the values of its k nearest neighbors [1, 21, 37, 44].

Abdelmoula [1] attempted to deal with the issue of default prediction for a Tunisian commercial bank of short-term loans. For this, a database of 924 credit applications of Tunisian companies issued from 2003 to 2006 by a Tunisian trade bank was used. The procedure for the K-Nearest Neighbor Classifier was done and the findings suggest that the best information was in the order of 88.63% (for  $k = 3$ ) for accrual and cash flow. In another study by Henley [38], the credit scoring model was built using the k-nearest neighbor approach. It begins with an overview of the k-NN technique and distance matrices before moving on to a realistic k-NN classification model. To accomplish so, they used sample data from the Littlewoods Organization, which included 19,186 applications for mail-order credit, 54.5% of whom were rated as poor risks. The entire data set was split five times at random into a training set 80% and a test set remaining 20%. The k-NN classification rules were developed using the modified Euclidean metric and their performance was tested. The k-NN classification rate of the four other approaches - linear and logistical regression, reverse projection and decision trees - is somewhat below that of the other three.

Credit lenders may be categorized into 'good' or 'bad' clusters using this lazy learning approach, but the choice of k is critical for achieving a high degree of accuracy. The k-NN algorithm can be beneficial in this sort of binary classification problem, however the rationale for the classification is yet unknown. At the same time, it's a lengthy procedure that will take a long time to complete. Several studies have been conducted, with the conclusion that k-NN is not as accurate as the other techniques in all cases. However, due to its simplicity, it has been utilized mostly in conventional or fuzzy based systems. Despite the fact that the researchers included datasets from many areas, the models' classification accuracy may have been skewed due to the impacts of imbalanced dataset effects. More study is needed to fully comprehend these difficulties in order to improve credit risk assessment accuracy.

### 4.4 Credit risk assessment using Support Vector Machine

Vladimir Vapnik [26] proposed the Support Vector Machine (SVM) [8, 35, 36, 70] as a supervised learning approach for categorizing high-dimensional data. The kernel function is used in order to convert an original training dataset into a larger dimension. This allows the SVM to find an ideal separation hyper-plane that serves as an effective decision boundary, separating classes with the greatest possible margin. Figure 6 illustrates the support vectors and the maximum margin. Several kernel functions exists in the literature, few are the following.

In this study, Harris [36] used the German credit scoring dataset from the UCI Machine Learning Repository. They created seven (7) classifiers- LR, K- means + LR, CSVM-RBF, K means + SVM-RBF, SVM-RBF, CSVM-linear, K means + SVM-linear, and SVM-linear to

compare their results. The dataset was randomly divided into two parts: test (20%) and training and cross validation (80%). According to the study, CSVM-RBF outperforms the competition with a mean training accuracy of 84.525%. In terms of AUC, it competes favourably with nonlinear SVM-based approaches, while surpassing them in terms of training time and training effort. CSVM's cutting-edge performance paired with its comparably low computing cost makes it an attractive choice for credit scoring.

In order to improve credit evaluation accuracy on the basis of SVM parameters, optimization is necessary. The work by Wang and Li [78] employed IFOA(Improved Fruit Fly Optimization Algorithm)to enhance SVM model parameters and hence increase performance in P2P (peer-to-peer) lending credit scoring. The dataset utilized in this study comes from RenRenDai website, a 2010 lending platform. As shown by the results, they compared among four models- Linear Regression, Classical SVM, FOA-SVM, and IFOA-SVM; the accuracy rate of IFOA-based SVM models outperformed others by reaching up to 93%.

Khemakhem et al. [48] discussed the issues with imbalanced credit datasets for credit risk assessment. The paper investigate the relevance and performance of sampling models combined with statistical prediction and artificial intelligence techniques to predict and quantify the default probability based on real-world credit data. They suggested that due to the categorization of this imbalanced data was skewed toward the majority cluster, which effectively implied that it tends to mistakenly label “extremely hazardous borrowers” as “good borrowers”. With Random over-Sampling(RoS) technique (92.22%) and Synthetic Minority Oversampling Technique(SMOTE) (92.105%), the SVM approach achieved the highest accuracy for unbalanced dataset, followed by ANN (91.667% with RoS and 89.035% with SMOTE). On the balanced dataset, AI based approaches considerably outperform the Linear Regression model (83.90% with RoS and 83.30% with SMOTE). It was observed that after balancing the data for all approaches, the specificity rose significantly. NN performed second best after SVM. Further improvement can be done by finding variables that have a significant influence on the probability of default.

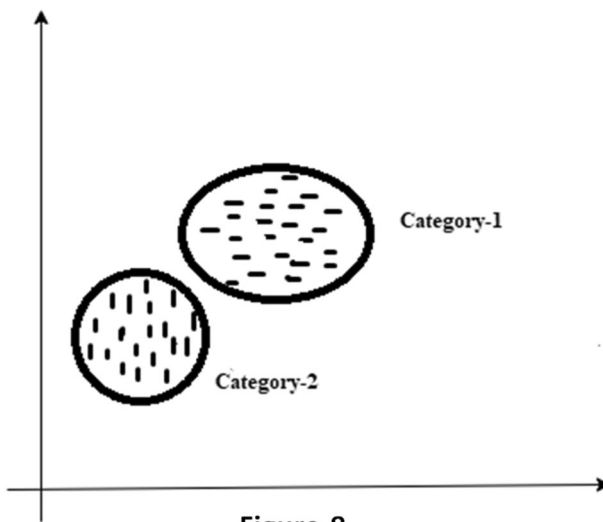


Fig. 8 An example of k-means



Moula et al. [61] in their research, compared a SVM-based Credit default prediction (CDP) algorithm with other statistical and intelligent approaches like- DA, LR, CART, SVM, MLP, and RBF using six different types of databases - Australian, German, and Japanese credit datasets from UCI machine learning database repository and The Chinese credit, a project dataset, provided by a leading Chinese commercial bank, and two other real-life datasets, PKKDD and Kaggle credit dataset. They utilized 10-fold cross-validation in their study. The radial basis function (RBF) was used as the kernel in the SVM credit prediction model. According to the findings, SVM has the greatest predictive performance in terms of average accuracy (85.32%), specificity (68.37%), precision (95.24%), and F-measure (89.12%).

Danenas and Garsva [28] offered a Particle Swarm Optimization (PSO)-based method for finding the optimal linear SVM classifier, which was tested against other classifiers in terms of accuracy and class identification. Experiments utilizing a real-world financial dataset from the SEC EDGAR database revealed that, despite its high average classification accuracy (over 90%), the suggested approach may produce results that are equivalent to those produced by other classifiers such as logistic regression and RBF networks. The performance of PSO-LinSVM was found to be less stable than that of its competitors. Linear SVM beats others in terms of average accuracy.

The major issue in credit risk research is determining which variables have a substantial impact on the chance of default. As a consequence, credit risk prediction accuracy is enhanced by identifying the most important financial and non-financial characteristics that can be utilized to better design the credit score model and assess company solvency. Few articles have arrived at the conclusion that some parameters are the most beneficial, but the list of parameters changes when other types of datasets are evaluated. Another source of worry for the researchers is class inequality. The credit decision is particularly tough in an incomplete information frame, and the bank might give credit to a bad borrower while refusing to fund a good borrower. In a balanced class distribution, linear SVM works well, but in the event of nonlinearity, choosing the right kernel function is crucial for achieving greater levels of performance.

#### 4.5 Credit risk assessment using Bayesian Network

Another widely accepted example of supervised machine learning technique Bayesian Network (BN), which utilizes a directed acyclic graph to express probabilistic connections. In BN, there exist methods for inference and learning. Each node in the network has a probability function that accepts an input and assigns a probability to the value associated with the node. Figure 7 showed an example of BN. Following are the few research studies that employed BN algorithm in evaluating credit risk.

Pavlenko and Chernyak [66] showed how probabilistic graphs may be used to predict and analyze credit concentration risk. The study suggested that Bayesian networks offer an appealing answer to these difficulties, and they demonstrated how to use them in describing, measuring, and managing the uncertain information in credit risk concentrations. Finally, they proposed tree-augmented Bayesian network which was also appropriate for stress-testing analysis, in particular, it can offer estimates of the posterior risk of losses associated with negative changes in the financial situations of a group of connected borrowers.

Customer default payments and the risks associated with loan allocation are the focus of this study by Triki and Boujelbene [75]. The outcomes of this study assist to create an effective decision support system for banks in recognizing and reducing the rate of problematic

borrowers by using a Bayesian Network model. The findings showed that, as compared to older applicants (76.88%), younger households had more creditworthiness since they have fewer obligations. In comparison to males, females had a lower risk of default (95.11% for females to 4.89% for males). As per the study, gender had a role in the categorization of problematic default borrowers. Another interesting fact is that the majority of borrowers who defaulted on their payments had loans with terms ranging from 0 to 84 months. This explained whenever the payback period was shorter, the probability of default decreased. BNs are a graph-based structure of a joint-multivariate probability distribution that describes how an expert determines the relationship between variables. As a result, the BN model provides a consistent framework for dealing with uncertainty and risk.

Anderson [5] described an empirical reject inference approach that used a Bayesian network. The suggested technique was superior to existing reject inference methods. The study was conducted utilizing two real-world credit score data sets from a German bank, as well as a public data set available from Lending Club. For the training and test data sets, each data set was randomly partitioned into ten folds with 70–30 train-test split. The rejected applicants were categorized as excellent or bad loans based on a cut-off predicted the high possibility of default. To distinguish between good and bad loans, ROC and AUC were used. The accuracy and AUC obtained using the German bank and Lending Club were (72%, 0.77) and (68%, 0.88), respectively.

To explain the payment default of loan customers, Masmoudi et al. [57] employed a discrete Bayesian network with a latent variable in this article. A latent variable with two classes was used to create BN. The Directed Acyclic Graph (DAG) was created by combining past expert knowledge with the Hill-climbing method. Second, estimation of the model parameters using the suggested Expectation-Maximization approach was considered. There were two classifications identified, each with a different credit risk profile. Its performance is compared to that of three other common classifiers, the decision tree, discrete BN, and Radial SVM, in terms of accuracy, precision, recall, and F1 Score. With 94.77% accuracy, BN with latent variable surpassed other classifiers. For both groups, the gender variable appears to have a minor impact on the likelihood of payment default. The model may be used to assess the likelihood of payment default while taking into consideration a variety of criteria and handling a multi-class situation.

#### 4.6 Credit risk assessment using k-means algorithm

K-means clustering is a widely recognized unsupervised machine learning approach that derives inferences from datasets using just input vectors and no labelled outputs. By grouping similar data points together, K-means tries to discover underlying patterns. To do this, it looks for a specific number ‘k’ of clusters in a dataset. A cluster is a group of data objects that have been grouped together owing to shared characteristics. To put it another way, the K-means algorithm calculates k-centroids and then allocates each data point to the cluster with the fewest centroids possible. Category wise data points were distributed in Fig. 8. Very few credit risk assessment studies were concluded with k-means algorithm. Following section discussed few of them.

Gavira-Durón et al. [33] searched for any inflection points when a company’s credit risk increases from low to high. They utilized simulations with homogeneous Markov chains and the k-means clustering method to calculate thresholds and migration within clusters to predict the chance of minimal credits migrating to medium and then to high. To demonstrate this, they

examined quarterly financial data from a sample of 35 public companies listed on the USA, Mexico, Brazil, and Chile stock markets. They used the k-mean approach to analyze financial data for each of the four quarters. As per the experimental result, it's been observed that high-risk organizations had a 0.79 chance of default, medium-risk firms had a 0.28 probability of default, and low-risk companies had a 0.009 likelihood of default. The major shortcoming of their research was that they didn't examine the entities' behaviour while issuing credits in different conditions.

The efficiency of generating credit rating transition matrices using sequence-based clustering on historical credit rating sequences using the PCA-guided K-means method was investigated by Le et al. [50]. The data set utilized in this study included monthly credit rating sequences from 1899 Korean firms. The credit rating sequences were first transformed into sequence matrices and grouped with PCA-guided K-means. The proposed clustering model was tested in three distinct long-term classification scenarios: 7-class credit rating prediction, credit rating transition direction prediction, and default behavior prediction. A 5-fold cross-validation method was considered. As per the study, the clustering model appeared to beat the benchmark model in terms of both the AA and F1 -measures. The author proposed to include characteristics in the model that account for Non-Markovian impacts of credit ratings, such as rating drift.

## 5 Optimization Technique based methods

Genetic Algorithms, Ant Colony Optimization, and Simulated Annealing are the most often used optimization techniques. Feature selection methods were created to discover acceptable feature subsets in order to increase the performance of a data categorization process. The most common use of optimization or heuristic strategies is to discover relevant feature subsets, which improves the model's accuracy.

### 5.1 Credit risk assessment using Genetic Algorithm

The GA is a classical meta-heuristic technique for solving NP-hard problems that replicate the process of biological genetic evolution, often known as Darwinian evolution [34, 39]. It keeps track of chromosomes, each of which may be a solution to the problem we're trying to solve. In the domain of genetic algorithms, chromosomes are frequently referred to as strings. A string, in turn, is made up of a number of genes, each of which can have any number of alleles. Each string is assigned a fitness value, which defines how 'excellent' it is on the overall scale. To identify the next parent, the fitness function is utilized to evaluate each chromosome, and then crossover and mutation are employed to produce the next generation population. Major operations related to GA are chromosome encoding, fitness function and genetic operations-(crossover, mutation, and selection).

To find the optimal feature subset and improve classification accuracy and scalability in credit risk assessment, Oreski and Oreski [64] presented a unique Hybrid Genetic Algorithm based Neural Networks (HGA-NN) approach. An essential characteristic in the high-dimensional input feature space is difficult to locate, as per the observation by the author. To reduce the feature space, additional improvements were made to the initial population and through the genetic algorithm. Earlier experience, Information gain, Gain ratio, Gini index/impurity, and correlation were used as filter approaches for generating initial solutions. The

proposed HGA-NN classifier was trained and tested using a real-world credit dataset gathered from a Croatian bank, and the results are then confirmed using a real-world German credit dataset taken from a UCI database. The classification accuracy was compared to GA-NN, SVM + GA [42], GP [41], NN [46], and SVM [79]. HGA-NN had an average prediction accuracy of 82.88% for the Croatian dataset and 78.90% for the German credit dataset, which can be considered at per level. As compared with the other methods, HGA-NN also used lowest number features to predict.

A credit crunch situation frequently triggered by a prolonged period of reckless and improper lending, resulting in severe losses for lending institutions and debt investors. In a highly competitive market with a credit crunch limitation, Metawa et al. [58] offered an intelligent model based on the GA to organize bank lending choices (GAMCC). In the quest for a dynamic lending decision, GAMCC provides a framework for optimizing bank objectives while creating loan portfolios by maximizing bank profit and reducing the likelihood of bank failure. GA was used to find the best consumers based on a variety of criteria like loan age, credit limit, loan amount, loan interest rate, loan type, and borrower credit rating. All data was collected from the World Bank public database for the year 2016, and the second dataset-Southern Louisiana Credit Union was used for further confirmation. GAMCC reduced loan screening time by between 36 and 50% when compared to TLP. When compared to a Multi-Objective Evolutionary algorithm (MODE-GL), the improvement ratio was between 12 and 28%. Furthermore, it significantly boosts bank profit from 3.9% to 8.1%.

Random Forest optimized using a genetic algorithm with profit score (RFoGAPS) was proposed by Ye et al. [81] to increase the loan evaluation effect and lender profits. After filtering the data, the experiment employed roughly 36 K borrow records from the Lending Club1 data. Finally, throughout the trials, 33 characteristics were employed. As per the findings, RFoGAPS, with a profit score of 787.20, was the most profitable technique, and the only one that was higher than Actual profit. In particular, LR was smaller than Actual Profit. The profit score of RFoGAPS improved by 7.73% when compared to the actual profit. This demonstrates that the RFoGAPS was the most effective and may assist lenders in obtaining 5.65 million more returns than the Actual profit in the years 2014–2016. In a separate experiment using 35 loan records from the 2016 dataset, it was observed that Random Forest had a predictive accuracy of 77.14% (27/35) and RFoGAPS had a predictive accuracy of 71.43% (25/35). However, the lender's real profit using RFoGAPS stands at \$59,205.87, which was significantly greater than the lender's actual profit from Random Forest (\$27,125.325). Finally, the author came to the conclusion that a high degree of anticipated accuracy does not necessarily translate into a higher profit.

Genetic Algorithm based models were utilized to extract the features from the feature subsets in order to increase the performance of a data categorization process. It was observed that when comparing the performance with the original methods, GA-based approaches performed superior in terms of all aspects. Other methods have been also considered and it was observed that in all cases, GA-based methods came out on top.

## 5.2 Credit risk assessment using Ant Colony Optimization

The Ant Colony Optimization (ACO) is a meta-heuristic approach for tackling combinatorial optimization problems that are heavily based on Marco Dorigo's [29] Ant System (AS) meta-heuristic. ACO is a system made up of agents that mimic the natural behaviour of ants, including cooperation and adaptability processes. It utilizes visual

**Table 9** Comparison of existing models based on k-NN

| Paper | Evaluation Model   | Dataset                  | Input / Output | Dataset Year | Size of dataset             | predictor | Accuracy/ Error rate  | Software |
|-------|--------------------|--------------------------|----------------|--------------|-----------------------------|-----------|---|----------|
| [1]   | K-Nearest Neighbor | Tunisian commercial bank | 24             | 2003 to 2006 | 924 credit applications     | 24        | 88.63%  | NA       |
| [38]  | K-Nearest Neighbor | Littlewoods organization | NA             | NA           | 19,186 Applications (80:20) | 16        | The k-NN had the average lowest bad rate (43.09%) among other 5 methods | NA       |

**Table 10** Different Kernel functions

| Kernels                                     | Formula  |
|---|--|
| Linear kernel [25, 73, 83]                  | $k(x, y) = x^T y + c$  |
| Polynomial kernel [25, 73, 83]              | $k(x, y) = (\alpha x^T y + c)^d$   |
| Gaussian Radial basis kernel [25, 73, 83]   | $k(x, y) = \exp\left\{-\frac{\ x-y\ ^2}{2\sigma^2}\right\}$  |
| Exponential RBF [83]                        | $k(x, y) = \exp\left\{-\frac{\ x-y\ }{2\sigma^2}\right\}$  |
| Sigmoid kernel function [25, 73, 83]        | $k(x, y) = \tanh(\alpha x^T y + c)$  |
| Cauchy kernel [73]                          | $k(x, y) = \frac{1}{1 + \frac{\ x-y\ ^2}{\sigma^2}}$   |
| Chi-Square kernel [73]                      | $k(x, y) = 1 - \sum_{i=1}^n \frac{(x_i - y_i)^2}{\frac{1}{2}(x_i + y_i)}$  |
| Laplace RBF kernel [73]                     | $k(x, y) = \exp\left\{-\frac{\ x-y\ }{\sigma}\right\}$   |
| Bessel Kernel [73]                          | $k(x, y) = \exp\left\{-\frac{J_{\nu+1}(\sigma \ x-y\ )}{\ x-y\ ^{-\nu(\nu+1)}}\right\}$  |
| ANOVA radial basis kernel [73]              | $k(x, y) = \sum_{k=1}^n \exp(-\sigma (x^k - y^k)^2)^d$   |
| Poly-Gaussian kernel [25]                   | $k(x, y) = \exp\left(-\frac{\ x-y\ ^2}{2\sigma^2} + c\right)^d$  |
| UKF [83]                                    | $k(x, y) = \frac{L}{\ x-y\ ^2 + \sigma^2}$<br>where L is a normalization constant.   |
| Linear splines kernel in one-dimension [73] | It comes from Lorentzian function.<br>$k(x, y) = 1 + xy + xy \min(x, y) - \frac{x+y}{2} \min(x, y)^2 + \frac{1}{3} \min(x, y)^3$ |

information to simulate the tactics used by actual ants to quickly determine the quickest route from a source to their destination and vice versa. Ants begin examining the area surrounding their source at random, and as they travel, a bit of pheromone falls to the ground, leaving a trail of the material. When a solitary ant walks at random and detects a laid pheromone, it is quite likely that it will choose to follow the laid pheromone's course. This ant will deposit a particular quantity of pheromone on the route, therefore enforcing the pheromone trail. More ants' follows a track, the more appealing that trail becomes and the more likely it is to be followed by more ants. The pheromone, on the other hand, begins to fade over time. The longer it takes an ant to walk down and back up the trail, the longer the pheromone will have to dissipate, and the path will become less visible. A shorter path, on the other hand, will be visited by more ants and can be described as a positive feedback loop in which the probability that an ant will choose a path is proportional to the number of ants who have already passed by that path, resulting in a higher pheromone density for a longer period of time. Because ants prefer to follow trails with more pheromones, all of the ants will ultimately converge on the shorter path. The ACO's core idea is to describe the issue as a graph search for the lowest cost path. Artificial ants explore this network in search of promising pathways. Because each ant has a rather basic behaviour, it will often only find low-quality trails on its own. The emergent effect of the colony's worldwide collaboration is the discovery of better pathways. An ACO algorithm is made up of a series of solution creation cycles (iterations). During each iteration, a group of ants (which is a parameter) develop entire solutions based on heuristic knowledge and prior ant groups' experiences [53, 54].

Nearest neighbour classifier was integrated with the proposed Ant Colony Optimization(ACO) algorithm by Marinakis et al. [54]. The ACO algorithm was utilized to identify credit risk level by using data from 1411 enterprises collected for the period of 1994-1997 from a large Greek commercial bank. In this scope, two approaches was presented- ACO-

**Table 11** Comparison of existing models based on Support Vector Machine

| Paper | Evaluation Model   | Dataset  | Input / Dataset Output Year | Size of dataset                       | predictor | Accuracy/ Error rate  | Software  |
|-------|--|--|-----------------------------|---------------------------------------|-----------|---|---|
| [36]  | LR, K- means + LR, CSVM-RBF, K means + SVM-RBF, SVM-RBF, CSVM-linear, K means + SVM-linear, and SVM-linear | German credit dataset  | NA NA                       | 1000 applicants(700 good and 300 bad) | NA        | accuracy: 84.525%   | NA  |
| [78]  | Linear Regression, Classical SVM, FOA-SVM and IFOA-SVM.  | RenRenDai  | NA 2010                     | 1000 samples                          | NA        | Accuracy for Linear Regression Model: 77.6%<br>Accuracy for Classical SVM Model: 82.4%<br>Accuracy for FOA-SVM Model: 90.8%<br>Accuracy for IFOA-SVM Model: 93%   | IFOA-SVM model used LIBSVM software package MATLAB R2011 b. |
| [48]  | LR, ANN, SVM. [ROS and SMOTE]  | Tunisian commercial bank   | 25/1 2011–2012              | 408 Com                               | 25        | LR: a)unbalanceddata:81.15%,<br>b)ROS:83.90%<br>c) SMOTE: 83.30%<br>ANN:<br>a)unbalanceddata:81.967%,<br>b)ROS: 91.667%<br>c) SMOTE: 89.035%<br>SVM: a)unbalanceddata:82.787%,<br>b)ROS: 92.222%<br>c) SMOTE: 92.105% |   |
| [61]  | SVM, DA,LR, CART, MLP,RBF.   | Australian, German, Japanese, The Chinese credit, a project dataset- provided by a leading Chinese commercial bank, PKKDD and Kaggle credit dataset. | NA NA                       | Total Cases: 205,491                  | NA        | For SVM average accuracy (85.32%), specificity (68.37%), precision (95.24%), and F-measure (89.12%).  | WEKA 3.8.0 and KNIME 2.12.0 credit prediction model         |
| [28]  | PSO-LinsVM   | SEC-EDGAR database.  | NA NA                       | NA                                    | NA        | Classification accuracy 90%.  | NA  |

1nn was the first approach, while the ACO-wknn was the second. The value of  $k$  in the ACO-wknn was dynamically adjusted based on the number of iterations. The proposed algorithm's results were then compared with SVM, CART, and two other meta-heuristic algorithms. The Root Mean Squared Error (RMSE) and classification accuracy were used to assess the quality of the model. Using 7.9 features, ACO-1nn determined to deliver the best results in terms of overall Accuracy with 97.52%.

The selection of appropriate features that are relevant to the situation is a key step toward the construction of any categorization model. Feature selection frequently relies on the subjective assessment of specialists in financial choices. Automated feature selection techniques, on the other hand, maybe of considerable assistance to decision-makers by allowing them to effectively explore the solution space. For this challenge, Marinakis et al. [55] employed two nature-inspired methods: Ant Colony Optimization and Particle Swarm Optimization. It's worth noting that ACO and PSO-based meta-heuristic methods both performed better with the  $k$ -nn classification method than the other variations, which use 1-nn and wk-nn. Using 10 folds cross-validation analysis, PSO  $k$ -nn offered 10.4 features, the best result in terms of the number of features utilized in the final model.

Credit risk models generally fall under supervised learning, and they must be simple to comprehend and transparent. But most of the machine learning techniques are of black-box nature, hence methods like neural networks and support vector machines are less suited. Martens et al. [56] employed AntMiner+ to create internal credit risk rating systems. Using Ant Colony Optimization concepts, AntMiner+ was able to infer a propositional rule set from a given data set. Three real-life credit risk data sets were used in the experiments: one for retail (German dataset), one for SMEs, and one for bank ratings (Bankscope database). AntMiner+ was shown to extract a strong and compact rule-set for each of these data sets. The extracted rule sets were found to be powerful in terms of discriminating power as well as comprehensibility. In addition, a concept detailing how AntMiner+ fits into a worldwide Basel II [11] credit risk management system was given. In terms of accuracy, non-linear SVM classifiers outperform linear SVM classifiers. We can see extremely competitive accuracies when comparing the rule- and tree-based classifiers AntMiner+ and C4.5, but when considering the amount of rules as well, AntMiner+ comes out on top with on average 4.9 rules.

### 5.3 Credit risk assessment using Simulated Annealing

Simulated Annealing (SA) [14] is a probabilistic search strategy that converts the chemical annealing process into an algorithm. The annealing process begins at a high temperature and then rapidly cools to create crystals. Identically, one object from the partitioning problem's solution space is used as an initial state, and a certain number of iterations are performed. At every temperature, the process maintains its equilibrium at that temperature. During the process, new solutions are developed and reviewed to see whether a better solution has come; if one has, it is maintained; otherwise, a selection method is carried out. The approach for solving the problem is based on probabilistic criteria. If a solution fails to meet the requirements, it is discarded; otherwise, it is given the opportunity to realign itself. As the switching occurrences continue at each step, there is a good chance of jumping over the local optima and into another. The beginning temperature, initial permutation, cooling rate, and equilibrium state are all significantly dependent on the process [15].



**Table 12** Comparison of existing models based on Bayesian Network

| Paper | Evaluation Model                      | Dataset                            | Input / Output | Dataset Year  | Size of dataset  | predictor | Accuracy/ Error rate  | Software |
|-------|---------------------------------------|------------------------------------|----------------|---------------|--|-----------|---|----------|
| [75]  | Bayesian network                      | commercial banks in Tunisia        | 9              | 2013–2014     | 6240   | 9         | NA  | Matlab   |
| [5]   | Bayesian network                      | i. German bank<br>ii. Lending Club | i. 19<br>ii. 5 | ii. 2007–2018 | i. 4500 (3000 accepted applicants and 1500 rejected applicants.)<br>ii. 798031(42,539 accepted applicants and 755,492 rejected applicants) | NA        | i. 72%<br>ii. 68%   | SAS      |
| [57]  | Bayesian network with latent variable | Tunisian Central bank              | 9              | 1990–2012     | 106,298  | NA        | Accuracy:94.77%, Precision: 85.31%,<br>Recall: 81.31%,<br>F1: 83.26%. | NA       |

**Table 13** Comparison of existing models based on K-Means classifier

| Paper | Evaluation Model | Dataset  | Input / Output | Dataset Year              | Size of dataset | predictor | Accuracy/ Error rate   | Software |
|-------|------------------|--|----------------|---------------------------|-----------------|-----------|--|----------|
| [33]  | K Means          | 35 public enterprises                                | NA             | 1 July 2006–28 March 2020 | NA              | NA        | High risk: 0.79<br>Medium Risk: 0.28<br>Low Risk: 0.009  | R        |
| [50]  | K Means          | National Information & Credit Evaluation Inc., Korea | 22 cluster     | 1986-09-01 to 2018-09-01  | 1899 firms      | 7 cluster | Input date: 2003-01-01<br>Clustering model performance: 0.9264<br>Benchmark model performance: 0.9179<br>Clustering models F1 score: 0.7425<br>Benchmark models F1 score: 0.7128 | NA       |

**Table 14** Comparison of existing works based on Genetic Algorithm based models

| Paper | Evaluation Model              | Dataset  | Input / Output | Dataset Year                     | Size of dataset  | predictor | Accuracy/ Error rate   | Software |
|-------|-------------------------------|--|----------------|----------------------------------|--|-----------|--|----------|
| [64]  | HGA-NN                        | Croatian credit dataset, German credit dataset                   | 12, 20         | September 2004 to September 2011 | i.1000 (750 good, 250 bad)<br>ii.1000 (700 good, 300 bad)                      | 12, 20    | HGA-NN: 82.88% for the Croatian dataset and 78.90% for the German credit dataset.                                | NA       |
| [58]  | Genetic Algorithm(GA)         | i.World Bank public database, ii.Southern Louisiana Credit Union | 10             | 2016                             | i.2000<br>ii.1000  | 10        | reduce the loan screening time by a range of 12% to 50%.<br>increases the bank profit by a range of 3.9% to 8.1% | C#       |
| [81]  | RF, RFoGAPS, SVM, DT, KNN, LR | Lending Club   | 17             | January 2014 and December 2016   | Before Filtering: 110 K borrow records;<br>After Filtering:36 K borrow records | 33        | 71.43%   | NA       |

**Table 15** Comparison of existing works based on Ant Colony Optimization based models

| Paper | Evaluation Model  | Dataset  | Input / Output            | Dataset Year                           | Size of dataset               | predictor  | Accuracy/ Error rate  | Software   |
|-------|---|--|---------------------------|--|-------------------------------|--|---|------------|
| [54]  | Ant Colony Optimization using Nearest Neighbor Classifier                     | Greek commercial bank  | 16                        | 1994–1997                              | 1411 firms                    | 16   | ACO-1nn:<br>Overall Accuracy- 0.9752<br>RMSE- 0.181<br>ACO-wknn:<br>Overall Accuracy- 0.9660<br>RMSE- 0.156<br>ACO k-nn: 72.85<br>PSO k-nn: 73.30 | Fortran 90 |
| [55]  | Ant Colony Optimization and Particle Swarm Optimization based meta-heuristics | UK non-financial firms   | 26                        | 1999–2001                              | 1330 firms                    | 26   |   | Fortran 90 |
| [56]  | Ant Colony Optimization, SVM, Decision Tree                                   | i. German Credit data set<br>ii.SME<br>iii. Bankscope database | i.20<br>ii. 40<br>iii. 37 | i.NA<br>ii. 1989–1997<br>iii.1998–2003 | i.1000<br>ii.422<br>iii. 2996 | i.RetailBanking: 20 but only 5 used<br>ii. SME Bankruptcy Prediction: 40 but only 5 used<br>iii. Rating Prediction: 37 | i. Retail Banking: C4.5 74.2<br>ii. SME: SVM 86.3<br>iii. Rating: SVM 87.7<br>AntiMiner+4.9   | Java       |

**Table 16** Comparison of existing works based on Simulated Annealing based models

| Paper | Evaluation Model | Dataset                    | Input / Output         | Dataset Year | Size of dataset | predictor | Accuracy/ Error rate  | Software |
|-------|------------------|----------------------------|------------------------|--------------|-----------------|-----------|---|----------|
| [45]  | SA and DT        | The German Credit database | Input: 20<br>Output: 1 | NA           | 1000 records    | 20        | Error rate: 34%   | NA       |
| [40]  | SA and GA        | a small credit platform    | 300                    | NA           | 5000            | NA        | Simulated annealing<br>Sensitivity-63.78%<br>Specificity-66.47%<br>AUC-0.7604 | R        |

**Table 17** Comparison of existing works based on Hybrid/Ensemble based models

| Paper | Evaluation Model   | Dataset  | Input / Output  | Dataset Year | Size of dataset  | predictor   | Accuracy/ Error rate   | Software           |
|-------|--|--|---|--------------|--|---|--|--------------------|
| [70]  | Random Forest and SVM.                                   | German Credit Data Set                                 | NA  | NA           | 1000   | NA  | Accuracy for Decision Tree: 0.72%, Random Forest: 0.78% and 10 fold Cross Validation:0.77% | NA                 |
| [84]  | EMPNGA and HEMPNGA                                       | Australian, German, Japanese, PPDai, and GMSC dataset. | Australian: 14, German: 20, Japanese: 15, PPDai: 29, GMSC: 10 | NA           | Australian: 690, German: 1000, Japanese: 690, PPDai: 55,596, GMSC: 120,269 | Australian: 14, German: 20, Japanese: 15, PPDai: 29, GMSC: 10 | Australian (0.8754), German (0.7682), Japanese (0.8720), PPDai (0.8728), GMSC (0.9352)     | Python Version 3.6 |
| [80]  | HMPGA, MPGA and GA                                       | German and Australian datasets.                        | German: 20, Australian: 14                                    | 2007         | German: 1000, Australian: 690  | NA  | German: 78.53% and Australian: 86.96%  | NA                 |
| [27]  | Feature Selection-based Hybrid-Bagging algorithm (FS-HB) | German dataset   | German: 20  | NA           | 1000 datasets  | NA  | NA   | Weka software      |
| [74]  | GBDT, LR, SVM, CART, MLP, AdaBoost, and RF               | Credit Assessment Company                              | NA  | NA           | 51,400 rows and 345 columns  | 80–90   | 90.99%(For GBDT)   | NA                 |

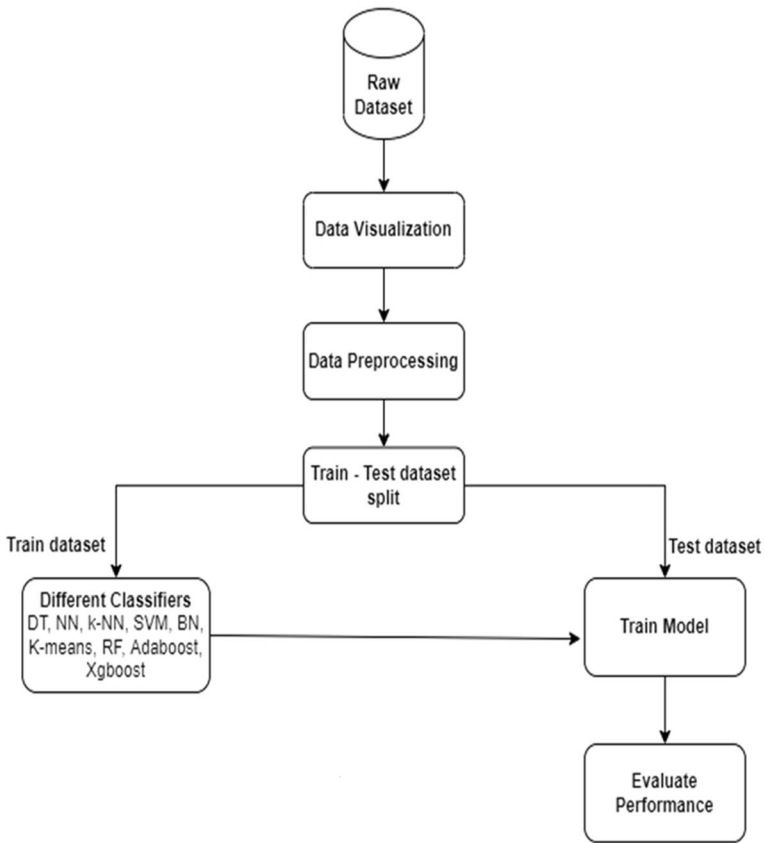


Fig. 9 Flowchart of the experimental work

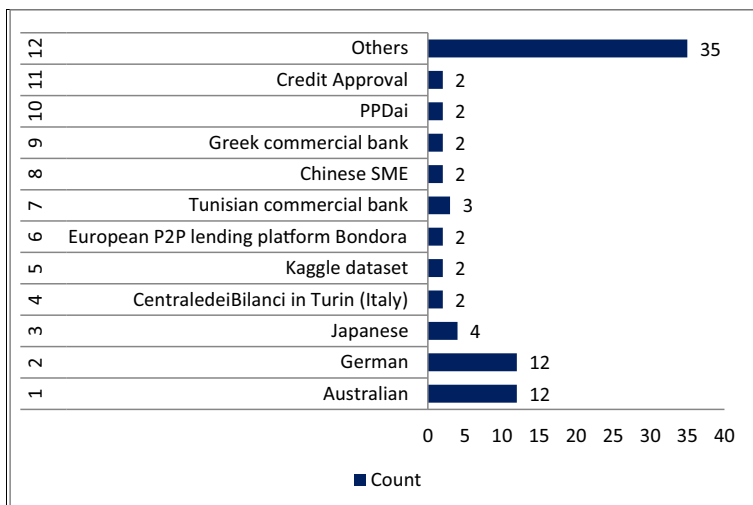


Fig. 10 Frequency of different datasets utilized by researchers

**Table 18** Different evaluation models tested on the available datasets

| Sl | Dataset                               | Evaluation Model  |
|----|---------------------------------------|---|
| 1  | Australian                            | HMM, LR, k-NN, MLPNN, BPNN, CRJ, SVM, EmNNs, Conventional NN, RxNCM, RxREN, RE-Re-RX, ERENNR, Re-RX, RxREN, PTVPSO-FKNN, TVPSO-FKNN, DA, DT, ANFIS, DNN, CART, ACO, GA, PSO, GWO, EMPNGA and HEMPNGA.   |
| 2  | German                                | HMM, LR, k-NN, MLPNN, BPNN, LR, DA, DT, ANFIS, DNN, CRJ, SVM, RxNCM, RxREN, RF, SVM, CART, LR, K- means + LR, CSVM-RBF, K means + SVM-RBF, SVM-RBF, CSVM-linear, K means + SVM-linear, SVM-linear, BN, ACO, GA, PSO, GWO, SA and DT based Hybrid Credit Scoring Model, EMPNGA and HEMPNGA |
| 3  | Japanese                              | LR, MLPNN, BPNN, DA, DT, ANFIS, DNN, SVM, DA, LR, CART, EMPNGA and HEMPNGA  |
| 4  | CentraledeiBilanci in Turin (Italy)   | Feed Forward Multi Layer Neural network, LDA and ANN  |
| 5  | Kaggle dataset                        | CRJ, SVM, DA, LR, CART, MLPNN   |
| 6  | European P2P lending platform Bondora | NN, LR, DT and k-NN   |
| 7  | Tunisian commercial bank              | LR,ANN,SVM, k-NN, BN  |
| 8  | Chinese SME                           | BPNN, RBF, GRNN, PNN, LR, MLPNN, DA, DT, ANFIS, DNN   |
| 9  | Greek commercial bank                 | LDA, Ant Colony Optimization using Nearest Neighbor Classifier  |
| 10 | PPDai                                 | BPNN, EMPNGA and HEMPNGA.   |
| 11 | Credit Approval                       | RE-Re-RX, ERENNR, Re-RX, RxREN  |

The Simulated Annealing Algorithm was initially utilized to construct the discriminant function for credit scoring problems. A top-down recursive approach was utilized by Jiang [45] using Decision Tree learning. It compared the value of an attribute in an inner node before generating the branch that follows it. Finally, from the leaf node, extraction was made to get the outcome. As a result, a conjunction rule corresponds to a path from the root to a leaf node, while the whole decision tree corresponds to a set of alternative expression rules. Quinlan [67] created the notion of Decision Tree Algorithm C4.5 by expanding and enhancing the ID3 algorithm. C4.5 can handle continuous and default characteristics in addition to the ID3 functions. Additionally, when cross certification was enabled, imbalanced trees were avoided using trimming technology. It was a non-parameter approach that makes no assumptions about the data distribution. The German credit database was utilized in the study, and it was discovered that a hybrid algorithm combining decision trees with the Simulated Annealing approach outperforms their separate results.

Feature selection was done using simulated annealing and a genetic algorithm by Hu and Cai [40] in this research work. Using data from a microfinance online platform, this research conducts an empirical investigation of individual credit value. Following feature selection and adequate data pre-processing procedures, the 5000 record dataset was split in a 70–30 ratio for model training and validation. In comparison to standard full variable logistic regression, the experimental results demonstrated that both the logistic model based on simulated annealing and the logistic model based on genetic algorithm had improved prediction performance and model interpretability. Observing the empirical results, it was revealed that the logistic model based on simulated annealing performed better than the logistic model based on genetic algorithm.



**Table 19** Best Performer on each category

| Credit Risk Evaluation Approach                      | Methodology  | Reference Model           | Best Accuracy(%)                              | Dataset   |
|--|--|---------------------------|---|---|
| Statistical Methods                                  | LA   | Logit Model               | 86.10   | Seven Asia Pacific Capital Markets                                    |
|  | HMM  | Hidden Markov Model       | 71.66,<br>84.33                               | German,<br>Australia  |
| Machine Learning Based Methods                       | DT   | Decision Tree based Model | 83.60   | Pre loan survey   |
|  | NN<br>Conventional NN                                | MLPNN                     | 79.00,<br>89.28,<br>89.71                     | German,<br>Japanese,<br>Australian                                    |
| Hybrid Models based on NN                            | Hybrid Models based on NN                            | MLP4                      | 99.73,<br>99.73,<br>96.51,<br>96.34,<br>98.07 | Chinese SME,<br>Chinese Farmer,<br>Australian,<br>German,<br>Japanese |
|  |  | EmNN<br>ERENNR            | 88.84<br>86.62,<br>85.51,<br>73.10            | Australian<br>Credit Approval,<br>Australian,<br>German               |
| Emotional Neural Network<br>Rule Extraction from ANN | Emotional Neural Network<br>Rule Extraction from ANN | k-NN                      | 88.63   | Tunisian Commercial Bank  |
|  |  | SVM                       | 93.00   | RenRenDai   |
| BN   | BN   | IFOA-SVM                  | 94.77   | Tunisian Central Bank   |
|  |  | BN with latent variable   | 92.64   | National Information & Credit Evaluation Inc., Korea                  |
| K-means  | K-means  | PCA-guided K-means        | 92.64   | National Information & Credit Evaluation Inc., Korea                  |
|  |  | HGA-NN                    | 82.88,<br>78.90                               | Croatian,<br>German   |
| Optimization Technique based methods                 | GA   | ACO-1NN                   | 97.52   | Greek Commercial Bank   |
|  |  | SA based DT               | 66.00   | German  |
| Hybrid and Ensemble methods                          | Hybrid   | HEMPNGA                   | 87.54,<br>76.82,<br>87.20,<br>87.28,<br>93.52 | Australian,<br>German,<br>Japanese,<br>PPDai,<br>GMSC                 |
|  |  | GBDT                      | 90.99   | Credit Assessment Company   |

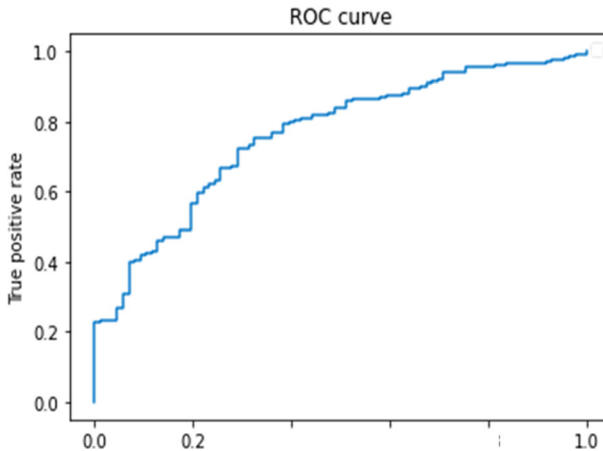


Fig. 11 RoC curve of XgBoost using German credit dataset

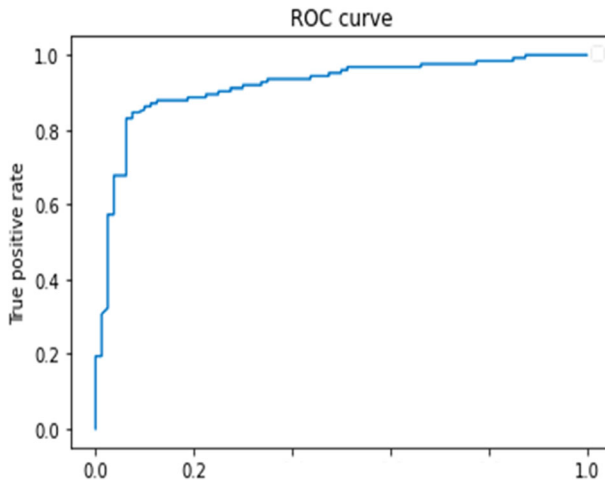
Simulated Annealing techniques haven't been used extensively in credit risk analysis purposes. It's been applied as a feature selection technique by the researchers in all the research works. Since it's a probabilistic searching technique, in most of the research works it came out as the best tool. In comparison to other meta-heuristic methods, simulated annealing performed exceptionally well but still, research areas are open to make it more transparent.

## 6 Hybrid/Ensemble methods

Hybrid techniques were developed to improve the original methods' performance and minimise Type-I or Type-II error. Roy and Urolagin [70] offered an approach that uses Random Forest and Support Vector Machine to perform two-level data processing to properly detect creditworthiness of the clients. RF was used to construct a precise credit scoring model, which

**Table 20** Performance comparison table for German and Australian credit datasets

| Machine Learning Type | Classifier Type            | Name of Classifier | German Credit Data    |               | Australian Credit Data |               |
|-----------------------|----------------------------|--------------------|-----------------------|---------------|------------------------|---------------|
|                       |                            |                    | Accuracy (10-fold CV) | RoC-AUC Score | Accuracy (10-fold CV)  | RoC-AUC Score |
| Supervised            | <b>Single Classifier</b>   | <b>DT</b>          | 70.10                 | 0.7404        | 84.30                  | 0.8878        |
|                       |                            | <b>k-NN</b>        | 66.00                 | 0.5154        | 62.80                  | 0.9076        |
|                       |                            | <b>SVM</b>         | 69.20                 | 0.5946        | 81.40                  | 0.9252        |
|                       |                            | <b>GNB</b>         | 71.30                 | 0.7165        | 86.10                  | 0.9216        |
|                       |                            | <b>MLPNN</b>       | 66.30                 | 0.4157        | 86.10                  | 0.9151        |
|                       | <b>Ensemble Classifier</b> | <b>RF</b>          | 74.30                 | 0.7554        | <b>86.20</b>           | <b>0.9349</b> |
|                       |                            | <b>Adaboost</b>    | 74.00                 | 0.7182        | 81.90                  | 0.9172        |
|                       |                            | <b>Xgboost</b>     | <b>75.60</b>          | <b>0.7618</b> | 85.20                  | 0.9326        |
| Unsupervised          | <b>Single Classifier</b>   | <b>Kmeans</b>      | 31.30                 | –             | 37.60                  | –             |



**Fig. 12** RoC curve of Random Forest using Australian credit dataset

will then be improved further with the SVM. They used the Gaussian Radial Basis Function (RBF) kernel for the SVM classification in this study. The German dataset was divided into 80:20 ratios for training and testing. According to the study, decision tree accuracy achieved 0.72 and for random forest accuracy attained 0.78 and with 10 fold cross validation accuracy level reached 0.77.

Based on the two optimum subsets the Zhang et al. [84] presented an enhanced multi-population niche genetic algorithm (EMPNGA), which enhances the selection, crossover, and mutation phases, and adds speciality and migration steps to increase optimization effectiveness. In order to verify the performance of the model, they employed five real-world credit datasets: the PPDai and GMSC datasets, as well as the Australian, German and Japanese credit datasets. Four assessment measures- accuracy, AUC, H-measure and Brier scores were utilised to provide the comprehensive prediction of the proposed model. As per the findings, the model ranked number one out of all classifiers. Ensemble classifiers (such as RF or GBDT) outperform individual classifiers in terms of average ranking in each evaluation metric. Consequently, the average ranking of the classifiers using HEMPNGA in feature selection was higher than the classifiers without HEMPNGA. Accuracy of the proposed model in several datasets: Australian (0.8754), German (0.7682), Japanese (0.8720), PPDai (0.8728), GMSC (0.9352). The paper finally concludes that research should focus on improving the efficiency of the heuristic method used in feature selection to improve the accuracy level and also points on the multi-classification work.

To address the feature selection problem, an adaptive evolutionary algorithm was employed to intelligently optimize MPGA based technique. Wang et al. [80] offered a hybrid method based on filtering and the HMPGA. It integrates the wrapper method into three filter approaches to obtain some critical prior information for MPGA initial population's configuration and then it uses the MPGA's global optimization and rapid convergence features to determine the best feature subset. This article compares HMPGA, MPGA, and GA using two genuine credit scoring data sets: Australian and German datasets. It shows that the accuracies of feature subsets obtained via

HMPGA, MPGA, and GA are superior to those obtained from three filter methods. On the German and Australian datasets, the suggested method-HMPGA had an average accuracy of 78.53% and 86.96%, respectively. Furthermore, the author's conclusion was also supported by a nonparametric Wilcoxon signed rank test.

Hybrid models, which are based on the combination of information from multiple approaches, can significantly improve the performance of the standalone models. The main goal of this research carried out by Dahiya et al. [27] was to improve the accuracy of a credit risk hybrid assessment model based on bagging technique. The benefits of integrating feature selection approach with an ensemble learning method were addressed by the suggested FS-HB algorithm. When qualitative and numeric German credit datasets were utilized, combined with two FS methods—chi-square test and PCA with bagging classifier—using the unstable base classifiers C4.5 and MLP, respectively, FS-HB algorithm outperformed others. Using FSHB, the percentage accuracy of bagging was increased by about 2%, while the accuracy of base classifiers was increased by around 6–8%.

Gradient Boosting Decision Tree was used to estimate credit risk in this research by Tian et al. [74]. A credit assessment firm provided a large quantity of data, which included 50,000 rows and 350 columns with discrete integer values. After cleaning the data, the dataset included about 51,400 rows and 345 columns. The Pearson Correlation Coefficient as a technique to minimize the total number of features was included in the model. It showed how closely the two variables were related. For proper classification of dataset, Tian et al. [74] utilized the six methods- LR, SVM, CART, MLP, AdaBoost, and Random Forest to evaluate the efficacy of Gradient Boosting Decision Tree. Gradient Boosting Decision Tree(90.99%, 90.37%, 0.97) outperforms four base learners & two ensemble learning methods in terms of three indicators (accuracy, f1score and AUC): LR(74.43%, 74.37%, 0.84), SVM(77.64%, 77.94%, 0.87), CART(84.68%, 84.71%, 0.85), MLP(84.61%, 83.45%, 0.93), AdaBoost(87.67%, 87.37%, 0.95) and Random Forest(88.96%, 88.45%, 0.96). Although the performance of the model appeared promising, its applicability may be challenged due to a lack of application on varied datasets. Data preparation and feature selection methods may want extra care.

As per the study, overall in almost all cases, a hybridized algorithm outperforms its standard version. Although it surpasses its counterpart in the vast majority of situations, it fails to explain why. In real-world circumstances, determining the underlying cause may be important in order to comprehend the problem's fundamentals. Despite the fact that several academics pointed out these flaws in detail, these algorithms failed to attain transparency. More research is needed to fully comprehend these techniques and determine the optimal method for precisely measuring the parameters so that the generalized algorithm produces the best possible outcome. Because these findings have such a direct impact on the financial industry, a better methodology might aid financial institutions in identifying credit risks and, as a consequence, save huge amounts of money. Despite the fact that this study used data from a variety of nations and sectors, the models' classification accuracy may have been skewed because of the impact specific factors like country, industry etc. Aside from this, the sample selection procedure may add a survivorship bias, because most businesses fail in their early years. Finding the right parameters might be a challenge for achieving optimality also. Future studies in these areas may improve in order to increase our understanding (Tables 9, 10, 11, 12, 13, 14, 15, 16 and 17).

## 7 Experimental work

Any learning methods employed in this domain are data-driven and computational-based, and they rely less on assumptions about the data, particularly the distribution. While they are seen to be more resilient and better at dealing with complicated non-linear interactions, they are also thought to be difficult to comprehend. Machine learning provides the capacity to find significant patterns in data, and it has become a standard tool for practically any activity requiring the extraction of relevant data from large datasets. Several stages must be completed in order to construct a machine learning model. Following the selection of the problem, appropriate datasets must be gathered from various sources or via surveys. Following that, datasets must be pre-processed by means of filtering out missing data, and/or encoding, and/or normalizing to improve interpretation (feature selection, outlier detection). To train the model and validate it using the dataset, a train test split must be performed.

Depending on the kind of problem, the appropriate machine learning model must be selected. Following training and testing, performance measures must be used to assess the degree of performance. For credit risk analysis, machine learning methods such as Decision Tree, Artificial Neural Network, k-Nearest Neighbor, Support Vector Machine, Bayesian Network, K-means, Random Forest, Adaboost, and Xgboost have been addressed in the experimental setup. 70:30 train test split has been considered. For validation, 10-fold cross validation is been employed for better insight. Finally, accuracy as well as Roc-AUC score have been recorded. Figure 9 depicts the framework of our model.

## 8 Results & discussion

In this study, we have encountered with 46 datasets, among them 35 datasets have been utilized for once. Among other 11 datasets, Australian, German and Japanese are the three most frequently utilized datasets by the researchers. Figure 10 shows the use of different datasets used by researchers in the past studies. 28 different methods including six rule extraction techniques have utilized Australian dataset. For German dataset, 32 different methods were tested. Table 18 showed the details briefly. Comparison among all these models wasn't possible due to its heterogeneity. Different types of data pre-processing method implementation, train-test splitting, different cross validation and different use of implementation platform are the key issues in this regard. Sixteen methods with their hybrid implementation have been studied in this work. Table 19 shows models with highest accuracy on different datasets. Significantly, the performance of hybrid or ensemble classifiers was exceptionally better than other single classifiers. To establish this hypothesis, six standalone classifiers – DT, k-NN, SVM, GNB, MLPNN, and K-means along with three ensemble classifiers – RF, Adaboost, and Xgboost have been tested in terms of accuracy and RoC-AUC score for two benchmark datasets – German and Australian credit data. As per the experimental result, Xgboost outperformed all other eight classifiers in terms of accuracy (75.60%) as well as RoC-AUC score (0.7618) for German dataset; whereas the performance of Adaboost and RF followed very closely. Figure 11 showed the RoC curve of Xgboost for German dataset. For the second Australian dataset, RF ranked on top for both accuracy (86.20%) and

RoC-AUC score (0.9349). It's been observed that the performances of all the ensemble classifiers are very much significant in contrast to other standalone classifiers for both datasets. The comparison has been demonstrated in Table 20. Figure 12 showed the RoC curve of RF for Australian dataset.

## 9 Conclusion

The subject of credit risk assessment using statistical, machine learning and heuristic methods were discussed in this article. The goal of this research is to analyze and examine the most recent machine learning algorithms and other techniques for credit risk analysis, as well as classify them based on their performance. As per the study shows that MLP's are typically better than alternative approaches, although the model of choice is dependent on the dataset provided. In general, credit risk analysis is categorized as classification problem but in few studies, it was observed that level of risk also concluded to make it a regression problem. In this work, 46 different datasets have been observed but Australian, German and Japanese dataset were the most utilized. Data pre-processing was carried out in different stages. For few models, Chi-square method was employed for validating the parameters; whereas few employed Pearson correlation for feature selection. For the reduction of high correlation and dimensionality of data, Principal Component Analysis (PCA) was utilized in few articles. Various performance methods including Accuracy, Precision, FP-Rate, Recall, f-Measure, MCC etc. were used to access the model's performance. Among all the performance metrics- the accuracy, of a machine learning model was utilized in most of the studies and when compared with the other models, it may not be as much to get noticeable but still, the margin of significance is quite influential for the financial institutes. Performance of a model was also increased by reducing the Type-I and Type-II errors. According to the findings, the accuracy of the machine learning technique in dealing with financial issues is superior to traditional statistical methods, particularly when dealing with nonlinear patterns. In most of the cases, Hybrid or Ensemble algorithms have shown to beat conventional algorithms in terms of performance. Optimization based methods mostly helped to reduce the size of the feature set and the errors to get more accurate result. In most of the cases the models with high performance were trained and tested utilizing very limited datasets. Although some of these methods are highly accurate, further research is needed to determine the best parameters and techniques for achieving better results in a transparent manner.

Finally, the article conducts an analytical study on the cited publications in order to draw conclusions and determines the main tendencies of future research. The experimental setup was carried out for two benchmark datasets. Since most of the earlier research works utilized accuracy as base performance metric, the experiment incorporates accuracy as well as RoC-AUC score as performance metric for comparison. It's been found that ensemble classifiers performed better than the standalone classifiers for both the datasets. It's been validated with the prior research also. Although the performance of all the classifiers were very much significant but scope of improvement still exists. With proper data pre-processing, feature engineering, feature selection and suitable classifier based on the available dataset can significantly improve the performance as well as reliability of the model.

### 10 Performance metrics

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP};$$

$$\text{Precision} = \frac{TP}{TP + FP};$$

$$\text{Sensitivity/True Positive Rate (TPR)/Recall} = \frac{TP}{TP + FN};$$

$$\text{Specificity/True Negative Rate (TNR)} = \frac{TN}{TN + FP};$$

$$\text{Youden's Index (J)} = \text{Sensitivity} + \text{Specificity} - 1.$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{TN + FP} = 1 - \text{Specificity};$$

$$\text{False Negative Rate (FNR)} = \frac{FN}{TP + FN} = 1 - \text{Sensitivity};$$

$$\text{F1 Score} = 2 * \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})};$$

$$\text{Matthews Correlation Coefficient (MCC)} = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}};$$

$$\text{Kappa score} = K = 1 - \frac{1 - p_o}{1 - p_e}$$

where,  $p_o$  is the observed value, and  $p_e$  is the expected value.

|                          |                          |                          |
|--------------------------|--------------------------|--------------------------|
| Confusion Matrix         | <b>Predicted Good(1)</b> | <b>Predicted Bad (0)</b> |
| <b>Actually Good (1)</b> | True Positive (TP)       | False Negative (FN)      |
| <b>Actually Bad (0)</b>  | False Positive (FP)      | True Negative (TN)       |

**Abbreviations** *AdaBoost*, Adaptive Boosting; *ANFIS*, Adaptive Neuro-Fuzzy Inference System; *ANN*, Artificial Neural Network; *AUC*, Area Under Curve; *BPNN*, Back-Propogation Neural Network; *CART*, Classification And Regreesion Tree; *CCR*, Candidate Classifier Repository; *CGD*, Conjugate Gradient Desecent; *CNN*, Convolutional Neural Network; *ConsA*, Consensus Approach; *CRJ*, Cycle Reservoir with Regular Jump; *CSVM*, Clustered Suport Vector Machine; *DA*, Discriminant Networks; *DAG*, Directed Acylic Graph; *DNN*, Deep Neural Network; *DP*, Discriminate Power; *DT*, Decision Tree; *EAD*, Exposure At Default; *EmNN*, Emotional Neural Network; *EMPNGA*, Enhanced Multi-Population Niche Genetic Algorithm; *FKNN*, Fuzzy K-Nearest Neighbour; *FNN*, Feedforward Neural Network; *GA*, Genetic Algorithm; *GBDT*, Gradient Boosting Decision Tree; *GD*, Gradient Descent; *GNG*, Gabriel Neighbourhood Graph; *GRNN*, General Regession Neural Network; *GWO*, Grey Wolf Optimization; *HMM*, Hidden Markov Model; *IFOA*, Improved Fruit Fly Optimization Algorithm; *IMF*, International Monetary Fund; *KDD*, Knowledge Discovery in Data; *KNN*, K- Nearest Neighbour; *LDA*, Linear Discriminant Anaysis; *LGD*, Loss Given Default; *LM*, Levenberg – Marquadt; *LR*, Logistic Regression; *MARS*, Multivariate Adaptive Regression Splines; *MLP*, Multilayer Perception; *MLPNN*, Multilayer Perception Neural Network; *MODE-GL*, Multi-Objective Evolutionary Algorithm; *MPGA*, Multiple Population Genetic Algorithm; *MSE*, Mean Squared Error; *NB*, Naïve Bayes; *NN*, Neural Network; *OS*, One-step Secant; *P2P*, Peer To Peer; *PD*, Probability of Default; *PNN*, Probalistic Neural Network; *PSO*, Particle Swarm Optimization; *PTVPSO*, Parallel TVPSO; *RBF*, Radial Basis Function; *RF*, Random Forest; *RFoGAPS*, Random Forest optimized by genetic algorithm with profit score; *RNN*, Recurrent Neural Network; *ROC*, Receiver operating Characteristic; *RoS*, Random Over Sampling; *SME*, Small- and Medium-sized Enterprises; *SMOTE*, Synthetic Minority Over-Sampling Technique; *SVM*, Support vector Machine; *TLP*, Traditional Linear Programming; *TVPSO*, Time Variant Particle Swarm Optimization; *UNCTAD*, UN Conference on Trade and Development

**Acknowledgements** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Data availability** All data generated or analysed during this study are included in this article.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Abdelmoula AK (2015) Bank credit risk analysis with k-nearest-neighbor classifier: Case of Tunisian banks. *Account Manag Inf Syst* 14(1):79
2. Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J Financ* 23(4):589–609
3. Altman EI, Saunders A (1997) Credit risk measurement: developments over the last 20 years. *J Bank Financ* 21(11–12):1721–1742
4. Anagnostou I, Kandhai D (2019) Risk factor evolution for counterparty credit risk under a hidden markov model. *Risks* 7(2):66
5. Anderson B (2019) Using Bayesian networks to perform reject inference. *Expert Syst Appl* 137:349–356
6. Atiya AF (2001) Bankruptcy prediction for credit risk using neural networks: a survey and new results. *IEEE Trans Neural Netw* 12(4):929–935
7. Augasta MG, Kathirvalavakumar T (2012) Reverse engineering the neural networks for rule extraction in classification problems. *Neural Process Lett* 35(2):131–150
8. Ayodele OE (2021) “Development of credit risk prediction model using support vector machine technique,” PhD Thesis, Federal University of Technology Akure
9. Back B, Laitinen T, Sere K, van Wezel M (1996) Choosing bankruptcy predictors using discriminant analysis, logit analysis, and genetic algorithms. *Turku Centre Comput Sci Tech Rep* 40(2):1–18



10. Bahrammirzaee A (2010) A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Comput & Applic* 19(8):1165–1195
11. Balin BJ (2008) “Basel I, Basel II, and emerging markets: a nontechnical analysis,” Available at SSRN 1477712
12. Baum LE, Eagon JA (1967) An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull Am Math Soc* 73(3):360–363
13. Baum LE, Petrie T (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Ann Math Stat* 37(6):1554–1563
14. Bertsimas D, Tsitsiklis J (1993) Simulated annealing. *Stat Sci* 8(1):10–15
15. Bhattacharya A, Ghatak S, Ghosh S, Das R (2014) “Simulated annealing approach onto VLSI circuit partitioning,”
16. Biswas SK, Chakraborty M, Purkayastha B, Roy P, Thounaojam DM (2017) Rule extraction from training data using neural network. *Int J Artif Intell Tools* 26(03):1750006
17. Chakraborty M, Biswas SK, Purkayastha B (2018) Recursive rule extraction from NN using reverse engineering technique. *N Gener Comput* 36(2):119–142
18. Chakraborty M, Biswas SK, Purkayastha B (2019) Rule extraction from neural network using input data ranges recursively. *N Gener Comput* 37(1):67–96
19. Chang Y-C, Chang K-H, Chu H-H, Tong L-I (2016) Establishing decision tree-based short-term default credit risk assessment models. *Commun Stat Theory Methods* 45(23):6803–6815
20. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357
21. Chen H-L, Yang B, Wang G, Liu J, Xu X, Wang SJ, Liu DY (2011) A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method. *Knowl-Based Syst* 24(8):1348–1359
22. Chen N, Ribeiro B, Chen A (2016) Financial credit risk assessment: a recent review. *Artif Intell Rev* 45(1): 1–23
23. Chi L-C, Tang T-C (2006) Bankruptcy prediction: application of logit analysis in export credit risks. *Aust J Manag* 31(1):17–27
24. Chi G, Uddin MS, Abedin MZ, Yuan K (2019) Hybrid model for credit risk prediction: an application of neural network approaches. *Int J Artif Intell Tools* 28(05):1950017
25. Chidambaram S, Srinivasagan KG (2019) Performance evaluation of support vector machine classification approaches in data mining. *Clust Comput* 22(1):189–196
26. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
27. Dahiya S, Handa SS, Singh NP (2017) A feature selection enabled hybrid-bagging algorithm for credit risk evaluation. *Expert Syst* 34(6):e12217
28. Danenas P, Garsva G (2015) Selection of support vector machines based classifiers for credit risk domain. *Expert Syst Appl* 42(6):3194–3204
29. Dorigo M, Di Caro G (1999) “Ant colony optimization: a new meta-heuristic,” in Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406), vol. 2, pp. 1470–1477
30. Estrella A (2000) “Credit ratings and complementary sources of credit quality information,”
31. Fatemi A, Fooladi I (2006) “Credit risk management: a survey of practices,” *Managerial Finance*
32. From global pandemic to prosperity for all: avoiding another lost decade. (2020)
33. Gavira-Durón N, Gutierrez-Vargas O, Cruz-Aké S (2021) Markov Chain K-Means Cluster Models and Their Use for Companies’ Credit Quality and Default Probability Estimation. *Mathematics* 9(8):879
34. Goldberg DE, Holland JH (1988) “Genetic algorithms and machine learning,”
35. Gyamfi K, Abdulai J-D (2018) “Bank fraud detection using support vector machine,” in 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 37–41
36. Harris T (2015) Credit scoring using the clustered support vector machine. *Expert Syst Appl* 42(2):741–750
37. He J, Liu X, Shi Y, Xu W, Yan N (2004) Classifications of credit cardholder behavior by using fuzzy linear programming. *Int J Inf Technol Decis Mak* 3(04):633–650
38. Henley WE (1997) Construction of a k-nearest-neighbour credit-scoring system. *IMA J Manag Math* 8(4): 305–321
39. Holland JH (1992) Genetic algorithms. *Sci Am* 267(1):66–73
40. Hu J, Cai J (2017) “Internet Credit Risk Scoring Based on Simulated Annealing and Genetic Algorithm,” in 2017 International Conference on Applied Mathematics, Modelling and Statistics Application (AMMSA 2017), pp. 373–377
41. Huang J-J, Tzeng G-H, Ong C-S (2006) Two-stage genetic programming (2SGP) for the credit scoring model. *Appl Math Comput* 174(2):1039–1053
42. Huang C-L, Chen M-C, Wang C-J (2007) Credit scoring with a data mining approach based on support vector machines. *Expert Syst Appl* 33(4):847–856

43. Huang X, Liu X, Ren Y (2018) Enterprise credit risk evaluation based on neural network algorithm. *Cogn Syst Res* 52:317–324
44. Imandoust SB, Bolandrafar M (2013) Application of k-nearest neighbor (knn) approach for predicting economic events: theoretical background. *Int J Eng Res Appl* 3(5):605–610
45. Jiang Y (2009) “Credit scoring model based on the decision tree and the simulated annealing algorithm,” in 2009 WRI world congress on computer science and information engineering, vol. 4, pp. 18–22
46. Khashman A (2010) Neural networks for credit risk evaluation: investigation of different neural models and learning schemes. *Expert Syst Appl* 37(9):6233–6239
47. Khashman A (2011) Credit risk evaluation using neural networks: emotional versus conventional models. *Appl Soft Comput* 11(8):5477–5484
48. Khemakhem S, Said FB, Boujelbene Y (2018) “Credit risk assessment for unbalanced datasets based on data mining, artificial neural network and support vector machines,” *J Modell Manag*
49. Konglai ZHU, Jingjing LI (2011) Studies of discriminant analysis and logistic regression model application in credit risk for China’s listed companies. *Manag Sci Eng* 4(4):24–32
50. Le R, Ku H, Jun D (2021) Sequence-based clustering applied to long-term credit risk assessment. *Expert Syst Appl* 165:113940
51. Leo M, Sharma S, Maddulety K (2019) Machine learning in banking risk management: A literature review. *Risks* 7(1):29
52. Lileikienė A (2008) “Analysis of chosen strategies of asset and liability management in commercial banks,” *Eng Econ*, vol. 57, no. 2
53. Marinakis Y, Marinaki M, Doumpos M, Matsatsinis N, Zopounidis C (2008) Optimization of nearest neighbor classifiers via metaheuristic algorithms for credit risk assessment. *J Glob Optim* 42(2):279–293
54. Marinakis Y, Marinaki M, Zopounidis C (2008) Application of ant colony optimization to credit risk assessment. *New Math Natural Comput* 4(01):107–122
55. Marinakis Y, Marinaki M, Doumpos M, Zopounidis C (2009) Ant colony and particle swarm optimization for financial classification problems. *Expert Syst Appl* 36(7):10604–10611
56. Martens D, Van Gestel T, De Backer M, Haesen R, Vanthienen J, Baesens B (2010) Credit rating prediction using ant colony optimization. *J Oper Res Soc* 61(4):561–573
57. Masmoudi K, Abid L, Masmoudi A (2019) Credit risk modeling using Bayesian network with a latent variable. *Expert Syst Appl* 127:157–166
58. Metawa N, Hassan MK, Elhoseny M (2017) Genetic algorithm based model for optimizing bank lending decisions. *Expert Syst Appl* 80:75–82
59. Miller LH, LaDue EL (1988) “Credit assessment models for farm borrowers: a logit analysis,”
60. Mohammadi N, Zangeneh M (2016) Customer credit risk assessment using artificial neural networks. *IJ Information Technol Comput Sci* 8(3):58–66
61. Moula FE, Guotai C, Abedin MZ (2017) Credit default prediction modeling: an application of support vector machine. *Risk Manag* 19(2):158–187
62. Nazari M, Alidadi M (2013) Measuring credit risk of bank customers using artificial neural network. *J Manag Res* 5(2):17
63. Ogun HT, Gurgun FS (2008) “Credit risk analysis using hidden markov model,” in 2008 23rd International Symposium on Computer and Information Sciences, pp. 1–5
64. Oreski S, Oreski G (2014) Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Syst Appl* 41(4):2052–2064
65. Pacelli V, Azzollini M (2011) An artificial neural network approach for credit risk management. *J Intell Learn Syst Appl* 3(02):103
66. Pavlenko T, Chernyak O (2010) Credit risk modeling using bayesian networks. *Int J Intell Syst* 25(4):326–344
67. Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1(1):81–106
68. Regulation R-BB (2009) “Foundations of banking risk,”
69. Rodan A, Faris H (2016) “Credit risk evaluation using cycle reservoir neural networks with support vector machines readout,” in Asian Conference on Intelligent Information and Database Systems, pp. 595–604
70. Roy AG, Urolagin S (2019) “Credit risk assessment using decision tree and support vector machine based data analytics,” in Creative Business and Social Innovations for a Sustainable Future, Springer, pp. 79–84
71. Satchidananda SS, Simha JB (2006) Comparing decision trees with logistic regression for credit risk analysis. International Institute of Information Technology, Bangalore
72. Setiono R, Baesens B, Mues C (2008) Recursive neural network rule extraction for data with mixed attributes. *IEEE Trans Neural Netw* 19(2):299–307
73. Souza CR (2010) Kernel functions for machine learning applications. Creative Commons Attribution-Noncommercial-Share Alike 3:29

74. Tian Z, Xiao J, Feng H, Wei Y (2020) Credit risk assessment based on gradient boosting decision tree. *Procedia Comput Sci* 174:150–160
75. Triki MW, Boujelbene Y (2017) “Bank credit risk: evidence from Tunisia using Bayesian networks.”
76. Uddin MS (2021) “Machine learning in credit risk modeling: empirical application of neural network approaches,” *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success*, pp. 417–435
77. Wang Y, Duan D (2021) Research on risk assessment of clients before loan based on decision tree algorithm. *J Phys Conf Ser* 1774(1):012056
78. Wang T, Li J (2019) An improved support vector machine and its application in P2P lending personal credit scoring. *IOP Conf Series: Mater Sci Eng* 490(6):062041
79. Wang S, Mathew A, Chen Y, Xi L, Ma L, Lee J (2009) Empirical analysis of support vector machine ensemble classifiers. *Expert Syst Appl* 36(3):6466–6476
80. Wang D, Zhang Z, Bai R, Mao Y (2018) A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring. *J Comput Appl Math* 329:307–321
81. Ye X, Dong L, Ma D (2018) Loan evaluation in P2P lending based on random forest optimized by genetic algorithm with profit score. *Electron Commer Res Appl* 32:23–36
82. Yurynets R, Yurynets Z, Dosyn D, Kis Y (2019) “Risk Assessment Technology of Crediting with the Use of Logistic Regression Model.,” in *COLINS*, pp. 153–162
83. Zhang R, Wang W (2011) Facilitating the applications of support vector machine by using a new kernel. *Expert Syst Appl* 38(11):14225–14230
84. Zhang W, He H, Zhang S (2019) A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: an application in credit scoring. *Expert Syst Appl* 121:221–232

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.