# Feature Selection Optimization Model for Business Risk Assessment Model

**Noura Metawa[1], Mohamed Elhoseny[2]**

[1]College of Business Administration, American University in the Emirates, Dubai, UAE

[2]College of Computer and Information Technology, American University in the Emirates, Dubai, UAE

Emails: Noura.metawa@aue.ae; Mohamed.elhoseny@aue.ae

**Abstract**

Financial risk assessment becomes a hot research topic among financial firms or companies to assess the financial status and thereby avoid future crises. Earlier studies have focused on statistical models for the assessment of financial risks and the recently developed machine learning (ML) models find useful to improve the assessment performance. In this aspect, this study introduces a novel Butterfly Optimization based Feature Selection with Classification Model for Financial Risk Assessment (BOFS-CFRA) technique. The proposed BOFS-CFRA technique involves pre-processing at the primary stage to get rid of unwanted data. In addition, K-means clustering approach is developed to group the financial data into clusters. Then, the BOFS technique is applied to choose the subset of features from the clustered data. Finally, the classification of financial risks takes place by the use of functional link neural network (FLNN). In order to ensure the enhanced performance of the BOFS-CFRA technique, a series of simulations were carried out and the results are inspected under various measures. The simulation outcome portrayed the supremacy of the BOFS-CFRA technique over the other financial risk assessment models in terms of several performance measures.

**Keywords:** Financial risk assessment, Classification, Feature selection, Butterfly optimization algorithm, FLNN.

## 1. Introduction

As the worldwide financial crisis in recent times, risk assessment in banks has acquired noticeable quality, and there has been a consistent spotlight on how risks are being recognized, estimated, announced and oversaw. Significant exploration [1-3], both in scholarly world and industry, has zeroed in on the improvements in banking and risk the executives and the current and arising difficulties. Pair, there has been a developing impact of AI in business applications, with numerous arrangements previously carried out and a lot more being investigated. McKinsey and Co featured that risk capacities in banks, by 2025, would should be in a general sense not quite the same as what they are today. The widening and developing of guidelines, advancing client assumptions and the advancement of risk types are relied upon to drive the change inside risk the executives [4]. New items, administrations and risk the board procedures are being empowered through the utilization of developing advances and progressed investigation. AI, recognized as one of the advancements with significant ramifications for risk the executives, can empower the structure of more exact risk models by distinguishing complicated, nonlinear examples inside huge datasets. The prescient force of these models can develop with all of data added, subsequently improving prescient control after some time. It is normal that AI will be applied



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

across different regions inside a bank's risk association. AI has additionally been suggested as a drive that could help in the change of the risk the executives work at banks.

The bank administration aim is to construct returns for its administrators arises at the expenses of extended risk. Bank administration is facing various risks—financing cost risk, market risk, credit risk, wobbly sheet risk, innovation and functional risk, unfamiliar trade risk, sovereign/nation risk, bankruptcy risk and liquidity risk. Successful management of such risks is vital to a bank presentation. Moreover, provided this risk and the task that bank plays in financial architecture, they are depending on administration concern [5]. The controller expects bank administrating to maintain capital for the many risks which appear and are transported owing to a bank fluctuated workloads. The Basel guiding principle for the guarantee of capital necessities have been invented since 1998. Capital is needed for each person of the principle risk categories. Generally, Credit risk is the dangerous risk challenging bank today, and typically the one requiring the most capital. Market risk emerges fundamentally from the interchanging activities of a banks, whereas functional risk is the risk of misfortune from outside occasions/interior architecture disappointments. As well as working out administration capital, many huge banks additionally compute financial capital, that is depending on a bank model contrasted with on solutions from the managers [6]. The essential risks that bank administration faces are market, credit, and functional risks, with distinct types of risk involving reputational, liquidity, and business risk. Bank administration is occupied efficiently with risk the executives to oversee, measure and screen this risks [7].

Artifcial knowledge (AI), and the AI strategies are fluctuating, and would disappoint, how we approach financial risk the executives [8]. With controlling and understanding risk is accessible for anyone over the development of AI-driven arrangement: from selecting how well a banks provide loan to the customer, notice signs to financial marketing merchants regarding positional risk, to recognizing insider and customer misrepresentations, and further developing lessening and consistence risk models. In this section we detail current AI procedures being utilized and current uses of those methods. We further conceive the future job for completely AI arrangements as the normal following stage after the far reaching reception of AI in assisting the association with overseeing risk [9, 10]. An illustration of ZestFinance serves to delineate the potential for AI and AI in risk the executives. ZestFinance was established by a previous Chief Information Offcer of Google and in 2016 collaborated with Baidu, the prevailing web crawler in China, to further develop Baidu's loaning choices in the Chinese market. Baidu was especially keen on making little credit offers to retail clients purchasing items from their foundation. In contrast to most created nations, the risk with loaning in the Chinese market is that under 20% of individuals have credit profiles or FICO assessments. Loaning to individuals who have either 'dainty' credit profles, or no credit profiles, is innately risky as there is no set of experiences to attract on to actually look at borrower dependability. ZestFinance (with authorization) takes advantage of the enormous volume of data on individuals held by Baidu, for example, their inquiry or buy narratives to assist Baidu with concluding whether to loan. They use great many information focuses per client are as yet ready to settle on loaning choices on new applications in a flash.

### 1.1 Prior Financial Risk Assessment Models

Le et al. [11] undergone the comparison study of the accurateness of 2 methods: ML procedures and conventional statistical techniques, that try to forecast the failure of banks. The example of 3000 US banks (1562 active banks and 1438 failures) is inspected by 3 ML methods (ANN, SVM, and KNN) and 2 classical statistical approaches (LR and Discriminant analysis). Dumitrescu et al. [12] present a higher-efficiency and credit scoring technique named penalized logistic tree regression (PLTR) that employs data from DT to enhance the performances of LR method. Generally, rules extracted from numerous short-depth DTs constructed by unique prediction variable is utilized as predictor in a PLTR method. The presented method enables to capture of nonlinear effects which could arise in credit scoring information when conserving the inherent interpretability of the LR method.

Rao [13] explores distinct communications among users within a Super-App provides a novel resource of data to forecast borrower behaviour. Then, 2 researches with distinct graph-based methods were introduced, initially employing graph based features as input in a classifier method and next utilizing graph NN method. Chang et al. [14] aim to utilize the classification, XGBoost, to create a credit risk valuation method for financial organizations. Cluster-based undersampling is positioned for processing



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

imbalanced information. Lastly, area under the accuracy of classification and receiver operative curve is the calculation indicator, compared to commonly employed single-phase classifier.

Lee et al. [15] introduce financial network indicator which is employed for global stock market investment. They present a directed and undirected volatility network of global stock markets-based system-wide connectedness and pair-wise correlation of national stock indices with a vector auto-regressive method. They investigate the usefulness and effect of network indicators by using them as input to determine approaches through numerous ML models (LR, SVM, and RF). Munkhdalai et al. [16] established a novel standard with actual user information and offer ML methods which serve as a baseline on this standard. They implemented wide-ranging comparisons among the ML methods and a human experts-based method—FICO credit scoring scheme—through a Survey of Consumer Finances (SCF) information. Since the SCF information is non-synthetic and includes huge quantity of real parameters, they employed 2 parameter-selection approaches: initially employed correlation, hypothesis test, and RF based feature measure and then NAP method used for selecting the illustrative feature for efficient modelling. Then, constructed regression method based on numerous ML model ranges from LR and SVM to an ensemble of GBT and DNN.

Zhu et al. [17] develops an improved hybrid ensemble ML method named RS-MultiBoosting by integrating 2 traditional ensemble ML methods, MultiBoosting and random subspace (RS), for improving the exactness of prediction SME credit risk. The research sample originated from information on quoted core enterprises (CEs) and 46 quoted SMEs as well as seven in the Chinese security markets amid 31 March 2014 and 31 December 2015, were gathered for testing the effectiveness and of feasibility the RS-MultiBoosting method. Bao et al. [18] present a combinational approach of unsupervised learning with supervised learning method for credit risk calculation. The distinction between this study and prior studies on unsupervised learning is that they employ unsupervised integrating method at 2 distinct phases: dataset clustering phase and the consensus phase. Comparison of system efficiency are implemented according to the 3 credit data sets in 4 sets: individual model + consensus model, individual model, clustering + individual model + consensus models, clustering + individual model.

### 1.2 Paper Contribution

This study introduces a novel Butterfly Optimization based Feature Selection with Classification Model for Financial Risk Assessment (BOFS-CFRA) technique. The proposed BOFS-CFRA technique involves pre-processing at the primary stage to get rid of unwanted data. In addition, K-means clustering approach is developed to group the financial data into clusters. Then, the BOFS technique is applied to choose the subset of features from the clustered data. Finally, the classification of financial risks takes place by the use of functional link neural network (FLNN). In order to ensure the enhanced performance of the BOFS-CFRA technique, a series of simulations were carried out and the results are inspected under various measures.

## 2. The Proposed Financial Risk Assessment Model

In this study, a novel BOFS-CFRA technique is developed for financial risk assessment. The proposed BOFS-CFRA technique comprises pre-processing, K-means clustering, BOA based feature selection, and FLNN based Classification. The detailed working of these processes is elaborated in the succeeding sections.

### 2.1 K-Means Clustering Process

Flow level clustering needs effectively subdividing gathered flows for all applications into groups according to the transferred NetFlow attribute. The most leading unsupervised clustering methods are the $k$- means clustering method chosen on another approaches like hierarchical clustering, because of its improved computation efficacy. $k$-means minimize a provided amount of vectors through selecting $k$ random vector as initial cluster center and assign all vectors to clusters as defined as a distance metrics compared to the cluster centers (a squared error function) as shown in (1). Then, the Cluster center is recalculated as the average (or mean) of the cluster member. This process repeats continuously, terminating whether the cluster converges or a certain amount of processes have been delivered [19]:



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^j - c_j \right\|^2 . \tag{1}$$

In (1), $cj$ represent cluster centers, $n$ equal the sizes of sample space (gathered flow), as well $k$ represent the selected values for amount of single cluster (flow class). Therefore, with $k$- means, $n$ flow is separated as $k$ class. $k$ value is very important since it influences directly the amount of flow class that affects overfitting.

**2.2 BOA based Feature Selection Process**

The nature stimulated metaheuristic method has been presented, called BOA [20], that inspires the mating and foraging behaviors of the butterfly. The key feature of BOA distinct from another metaheuristic is that all the butterflies have their individual scent. The fragrance is expressed by:

$$f_i = cI^a \tag{2}$$

whereas $c$ denotes the sensory modality, $f_i$ represent the perceived magnitude of fragrance, and $I$ indicates the stimulus intensity, besides $a$ signifies the power exponent-based amount of fragrance absorption. In theoretic, several values of sensory morphology coefficient $c$ in the interval of $[0, \infty]$ could be considered. But the values are defined by the particularity of optimization issues in the iteration method. The sensory modality $c$ in the optimum searching region of the method is represented by:

$$c_{t+1} = c_t + [0.025/(c_t \cdot T_{\max})] \tag{3}$$

In which $T_{\max}$ represent the maximal amount of iterations, and $c$ represent the primary value fixed to 0.01.

Additionally, it consists of local search phase and global search phase respectively. The arithmetical expression of the butterfly's global search movement is given by:

$$x_i^{t+1} = x_i^t + (r^2 \times g_{best} - x_i^t) \times f_i \tag{4}$$

whereas $x_i^t$ signifies the solution vector $x_i$ of $ith$ butterfly in $t$ iteration and $r$ implies an arbitrary value in the range of zero and one. Now, $g_{best}$ denotes the present optimal solution established between each solution in the present phase. Mainly $f_i$ signifies the fragrance of ith butterfly. The local search stage is expressed by:

$$x_i^{t+1} = x_i^t + \left(r^2 \times x_i^k - x_j^t\right) \times f_i \tag{5}$$

In which $x_j^t$ & $x_i^k$ denotes $j$th and $k$th butterflies selected arbitrarily from the solution space. When $x_j^t$ and $x_i^k$ belongs to similar iterations, it implies that the butterflies become a local arbitrary walk. Then, this type of arbitrary motion would vary the solutions.

The local search and global search for food and mating partners through the butterfly in nature might take place. Thus, a $p$ switch probability is fixed to change the intensive local search and the standard global search. In all the iterations, the BOA arbitrarily creates in the range of zero and one, i.e., related to $p$ switch probability for deciding either to perform a local/global search. The process involved in BOA is shown in Fig. 1 [21].



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

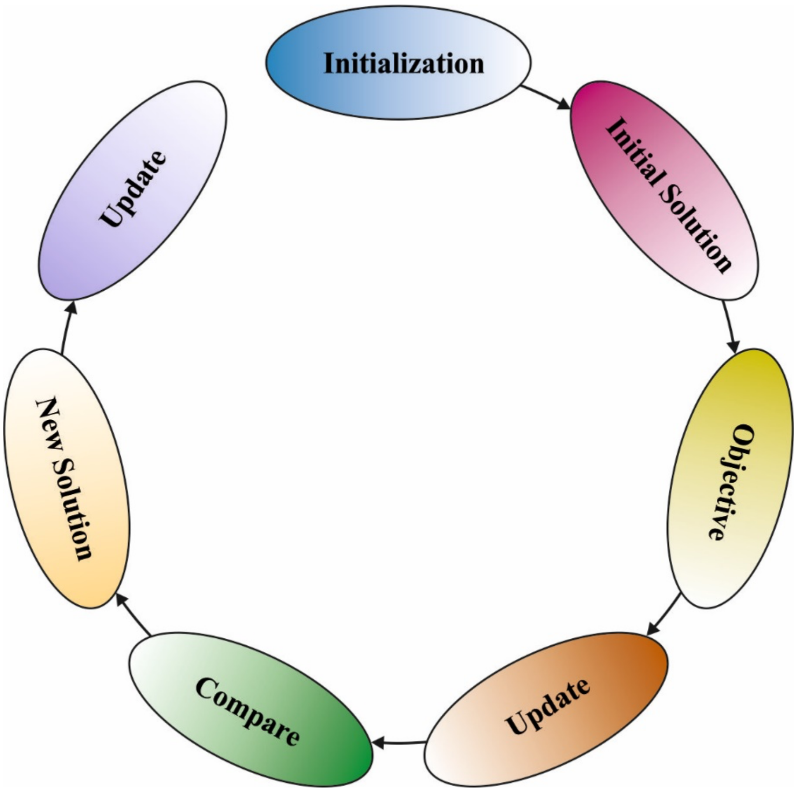


**Fig. 1. Steps involved in BOA**

Unlike the typical BOA, where the solution is upgraded in the searching space toward continuous valued position, in the BBOA, the search space is modeled as $n$ dimension Boolean lattice. Also, the solution is upgraded over the corner of a hypercube. Additionally, to solve the problem of whether to elect or not, a provided parameter and binary solution vectors are applied, in which 1 corresponds to a parameter being elected to comprise the novel datasets, and 0 corresponds to else.

In binary algorithms, one uses the step vectors to evaluate the likelihood of changing position, the transfer function considerably impacts the balance between exploitation and exploration. In FS method, when the size of feature vectors represent $N$, the amount of distinct feature combination tend to be $2^N$, i.e., a massive space for comprehensive search. The proposed hybrid algorithms are employed for this purpose for searching the feature space vigorously also generating the correct combinations of features. The FS falls within multiobjective challenges since it needs to fullfill several purposes for getting an optimal solution, which minimizes the subsets of FS and simultaneously, maximizing the accuracy of output to provided classifiers.

Based on the abovementioned, the fitness function (FF) to determine solution in this condition made to attain a balance among the two objectives as:

$$fitness = \alpha \Delta_R(D) + \beta \frac{|Y|}{|T|} \quad (6)$$

$\Delta_R(D)$ is the classifier error rate. $|Y|$ denotes the size of subsets that the technique chooses and $|T|$ overall quantity of features contained in the present datasets. $\alpha$ denotes a parameter $\in [0, 1]$ associated with the weight of error rate of classifications, correspondingly also $\beta = 1 - \alpha$ represents the importance of reduction feature. The classification performance is permitted an important and weight instead of the number of selected features. When the estimation function only considered the classification accuracy, the effects would be the neglect of solution which might contain similar accuracy, however, have lesser selected feature which serves as the main aspect in reducing the dimensionality problem.



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

### 2.3 FLNN based Classification Process

Generally, the functional link-based NN is single-layer ANN framework having minimum computation load and maximum rate of convergence when compared to MLP framework. The mapping and behaviour capacity of a PPN and the applications to channel equalization are stated. The computational calculation and arithmetical expression are estimated in accordance with MLP. Patra initially projected FLNN, and it is a new single-layer ANN model which can form randomly complicated decision regions by making non-linear decision limits. In FLNN, the hidden layer is detached. Furthermore, the FLNN architecture provides maximum convergence speed and minimum computation difficulty when compared to MLP due to its single-layer architecture. Now, the functional extension block utilizes functional models including a subsection of orthogonal cos and sin basis function and the unique pattern as well as its outer product. The BP model, i.e., utilized for training the network, becomes easy owing to nonappearance of some hidden layers. The explanation for using trigonometric function in the FLANN architecture is given in [22] and the structure is shown in Fig. 2 [23].

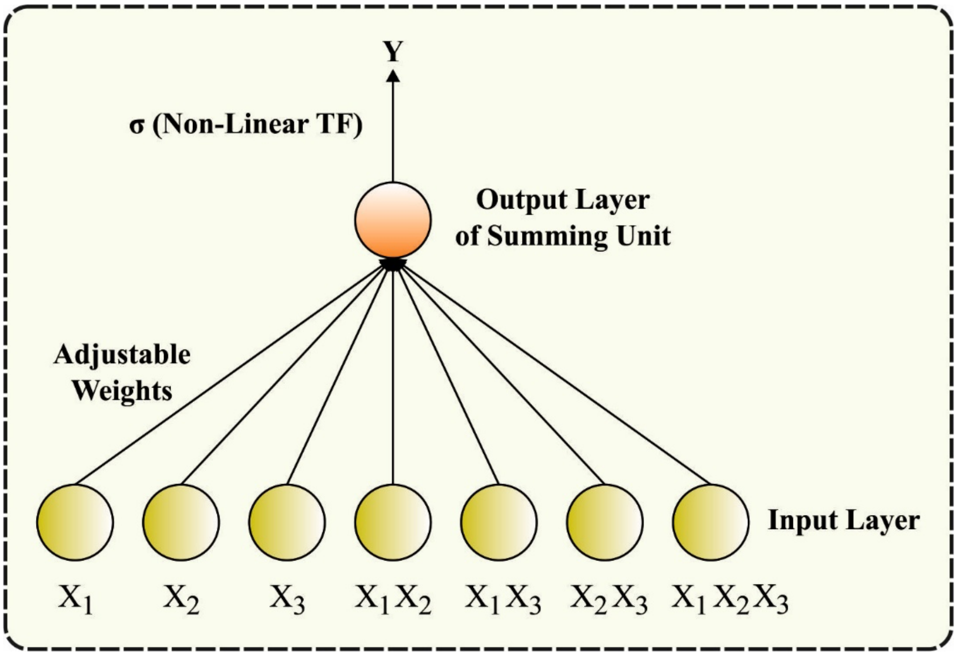


**Fig. 2. Strucutre of FLNN model**

### 3. Performance Validation

This section investigates the financial risk assessment performance of the BOFS-CFRA technique on two benchmark dataset namely Australian Credit and German credit dataset. The details related to the dataset is offered in Table 1 and the selected ten ranked features are offered in Table 2.

**Table 1 Dataset Description**

| Dataset | Source | # of instances | # of attributes | # of class | B/ NB |
|---|---|---|---|---|---|
| Australian-Credit | UCI | 690 | 14 | 2 | 383/307 |
| German-Credit | UCI | 1000 | 20 | 2 | 300/700 |



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

**Table 2 Top 10 Ranked Features**

| Dataset | Features |
|---|---|
| Australian Credit | A2, A14, A8, A3, A13, A7, A10, A9, A5 |
| German Credit | Credit amount, Status of existing checking account, Duration in months, Age in years, Credit history, Savings account/bonds, Purpose, Property, Present employment since, Housing |

### 3.1 Results analysis on Australian Credit Dataset

Fig. 3 demonstrates the confusion matrices of the BOFS-CFRA technique on the test Australian Credit dataset with five different runs. On the applied run-1, the BOFS-CFRA technique has classified 371 instances into bankrupt and 293 instances into non-bankrupt. Likewise, On the applied run-2, the BOFS-CFRA technique has classified 372 instances into bankrupt and 293 instances into non-bankrupt. Similarly, On the applied run-3, the BOFS-CFRA technique has classified 374 instances into bankrupt and 295 instances into non-bankrupt. Moreover, on the applied run-1, the BOFS-CFRA technique has classified 371 instances into bankrupt and 296 instances into non-bankrupt. Furthermore, on the applied run-1, the BOFS-CFRA technique has classified 372 instances into bankrupt and 295 instances into non-bankrupt.

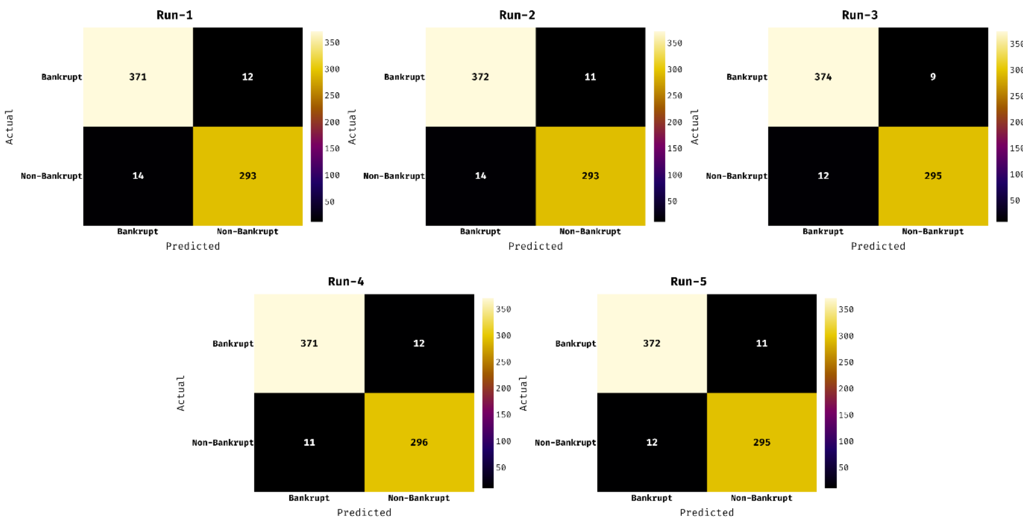


**Fig. 3. Confusion matrix of BOFS-CFRA technique on Australian Credit dataset**

Table 3 provides a detailed classification performance of the BOFS-CFRA technique under five runs. With run-1, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9636, 0.9687, 0.9623, 0.9661, and 0.9237 respectively. Also, With run-1, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9637, 0.9713, 0.9638, 0.9675, and 0.9266. In addition, With run-3, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9689, 0.9765, 0.9696, 0.9727, and 0.9384. Followed by, With run-4, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9712, 0.9687, 0.9667, 0.9699, and 0.9325. Finally, With run-5, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9688, 0.9713, 0.9667, 0.9700, and 0.9325.



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

**Table 3 Overall classification results of BOFS-CFRA technique on Australian Credit dataset**

| No. of Runs | Precision | Recall | Accuracy | F-Score | MCC |
|---|---|---|---|---|---|
| Run-1 | 0.9636 | 0.9687 | 0.9623 | 0.9661 | 0.9237 |
| Run-2 | 0.9637 | 0.9713 | 0.9638 | 0.9675 | 0.9266 |
| Run-3 | 0.9689 | 0.9765 | 0.9696 | 0.9727 | 0.9384 |
| Run-4 | 0.9712 | 0.9687 | 0.9667 | 0.9699 | 0.9325 |
| Run-5 | 0.9688 | 0.9713 | 0.9667 | 0.9700 | 0.9325 |
| **Average** | **0.9672** | **0.9713** | **0.9658** | **0.9692** | **0.9307** |

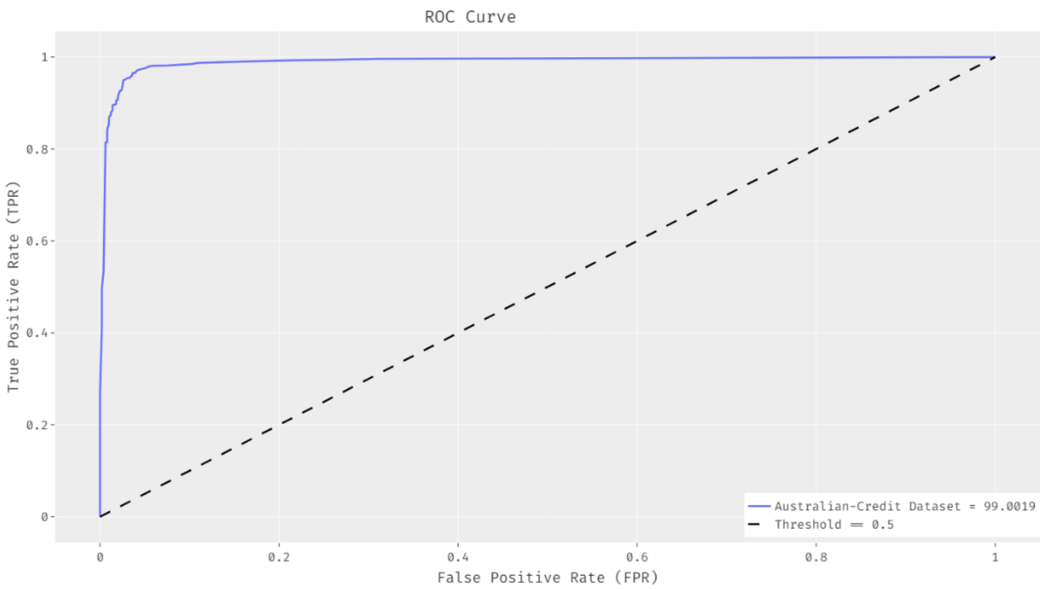


**Fig. 4. ROC analysis of BOFS-CFRA technique on Australian Credit dataset**

Fig. 4 showcases the ROC analysis of the BOFS-CFRA technique on the Australian Credit dataset. The figure demonstrated that the BOFS-CFRA technique has accomplished improved performance with the maximum ROC of 99.00193

### 3.2 Results analysis on German Credit Dataset

Fig. 5 demonstrates the confusion matrices of the BOFS-CFRA technique on the test German Credit Dataset with five different runs. On the applied run-1, the BOFS-CFRA technique has classified 289 instances into bankrupt and 677 instances into non-bankrupt. Likewise, on the applied run-2, the BOFS-CFRA technique has classified 292 instances into bankrupt and 683 instances into non-bankrupt. Likewise, on the applied run-3, the BOFS-CFRA technique has classified 291 instances into bankrupt and 680 instances into non-bankrupt. Besides, on the applied run-4, the BOFS-CFRA technique has classified 289 instances into bankrupt and 679 instances into non-bankrupt. Additionally, on the applied run-5, the BOFS-CFRA technique has classified 291 instances into bankrupt and 680 instances into non-bankrupt.



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

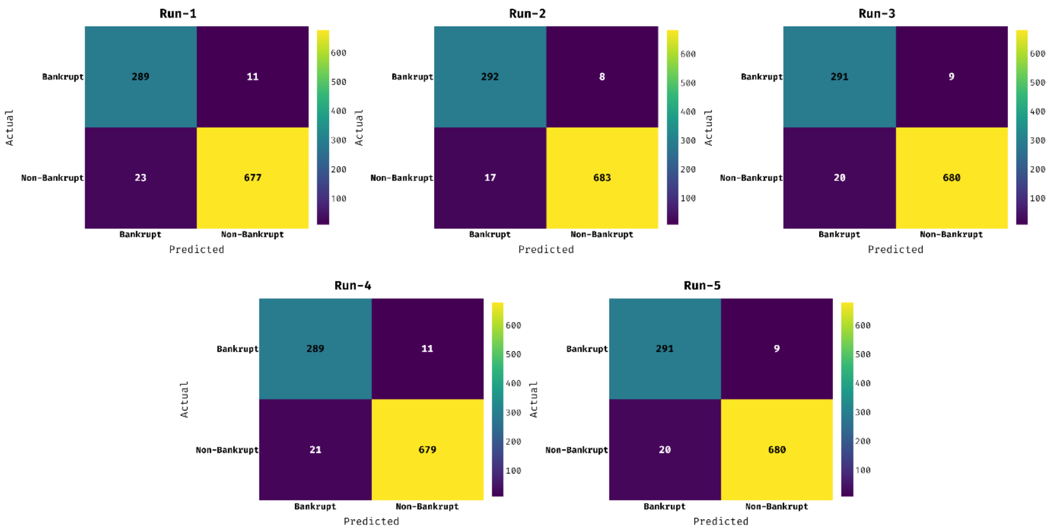


**Fig. 5. Confusion matrix of BOFS-CFRA technique on German Credit dataset**

**Table 4 Overall classification results of BOFS-CFRA technique on German Credit dataset**

| No. of Runs | Precision | Recall | Accuracy | F-Score | MCC |
|---|---|---|---|---|---|
| Run-1 | 0.9263 | 0.9633 | 0.9660 | 0.9444 | 0.9203 |
| Run-2 | 0.9450 | 0.9733 | 0.9750 | 0.9589 | 0.9412 |
| Run-3 | 0.9357 | 0.9700 | 0.9710 | 0.9525 | 0.9320 |
| Run-4 | 0.9323 | 0.9633 | 0.9680 | 0.9475 | 0.9248 |
| Run-5 | 0.9357 | 0.9700 | 0.9710 | 0.9525 | 0.9320 |
| **Average** | **0.9350** | **0.9680** | **0.9702** | **0.9512** | **0.9301** |

Table 4 provides a detailed classification performance of the BOFS-CFRA technique under five runs. With run-1, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9263, 0.9633, 0.9660, 0.9444, and 0.9203 respectively. Followed by, With run-2, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9450, 0.9733, 0.9750, 0.9589, and 0.9412. Then, With run-3, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9357, 0.9700, 0.9710, 0.9525, and 0.9320. Afterward, With run-4, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9323, 0.9633, 0.9680, 0.9475, and 0.9248. At last, With run-5, the BOFS-CFRA technique has gained $PRE_N$, $REC_L$, $ACC_Y, F_{SCORE}$, and MCC of 0.9357, 0.9700, 0.9710, 0.9525, and 0.9320.



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

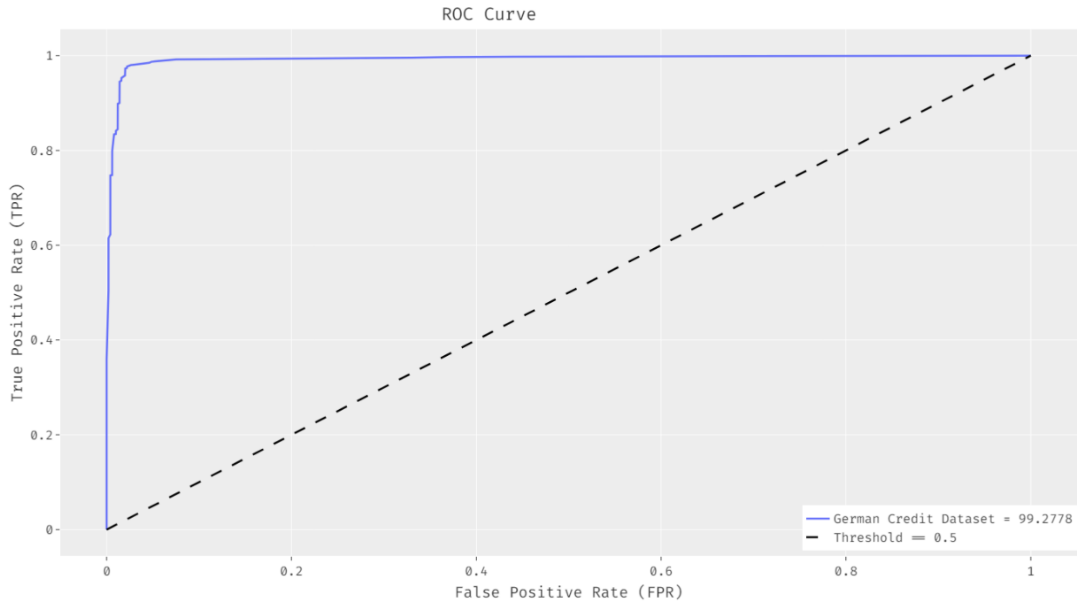


**Fig. 6. ROC analysis of BOFS-CFRA technique on German Credit dataset**

Fig. 6 showcases the ROC analysis of the BOFS-CFRA technique on the German Credit Dataset. The figure demonstrated that the BOFS-CFRA technique has accomplished improved performance with the maximum ROC of 99.2778.

### 3.3 Discussion

Table 5 demonstrates the detailed results analysis of the BOFS-CFRA technique with other existing techniques on the Australian and German Credit dataset.

**Table 5 Comparative study of BOFS-CFRA technique with traditional methods**

| Methods | Australian Credit | German Credit |
|---|---|---|
| BOFS-CFRA | 0.9658 | 0.9702 |
| FRA Model | 0.9568 | 0.9476 |
| SVM | 0.8492 | 0.7320 |
| Bagging | 0.8405 | 0.7460 |
| DT | 0.8347 | 0.7240 |
| LR | 0.8463 | 0.7530 |
| RBF | 0.8347 | 0.7120 |

In order to showcase the better analysis of the BOFS-CFRA technique on the test Austrailan credit dataset, a comparative analysis is shown in Fig. 7. The figure portrayed that the SVM, Bagging, DT, LR, and RBF techniques have accomplished reduced accuracy of 0.8492, 0.8405, 0.8347, 0.8436, and 0.8347 respectively. At the same time, the FRA model has resulted to a moderately considerable outcome with the accuracy of 0.9658. However, the proposed BOFS-CRFA technique has accomplished improved outcome with the accuracy of 0.9658.



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

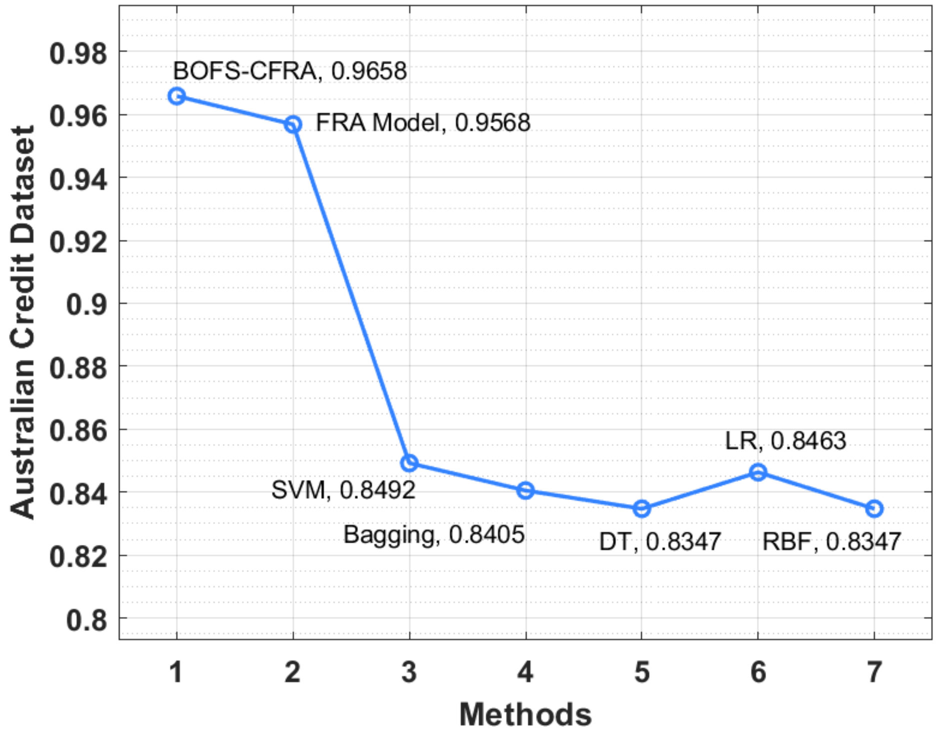


**Fig. 7. Comparative results analysis of BOFS-CFRA technique with traditional approaches on Australian Credit dataset**

A brief comparative result analysis of the BOFS-CFRA technique on the test German credit dataset is shown in Fig. 8. The figure portrayed that the SVM, Bagging, DT, LR, and RBF techniques have accomplished reduced accuracy of 0.7320, 0.7460, 0.7240, 0.7530, and 0.7120 respectively. In addition, the FRA model has resulted to a moderately considerable outcome with the accuracy of 0.9476. However, the proposed BOFS-CRFA technique has accomplished improved outcome with the accuracy of 0.9702.

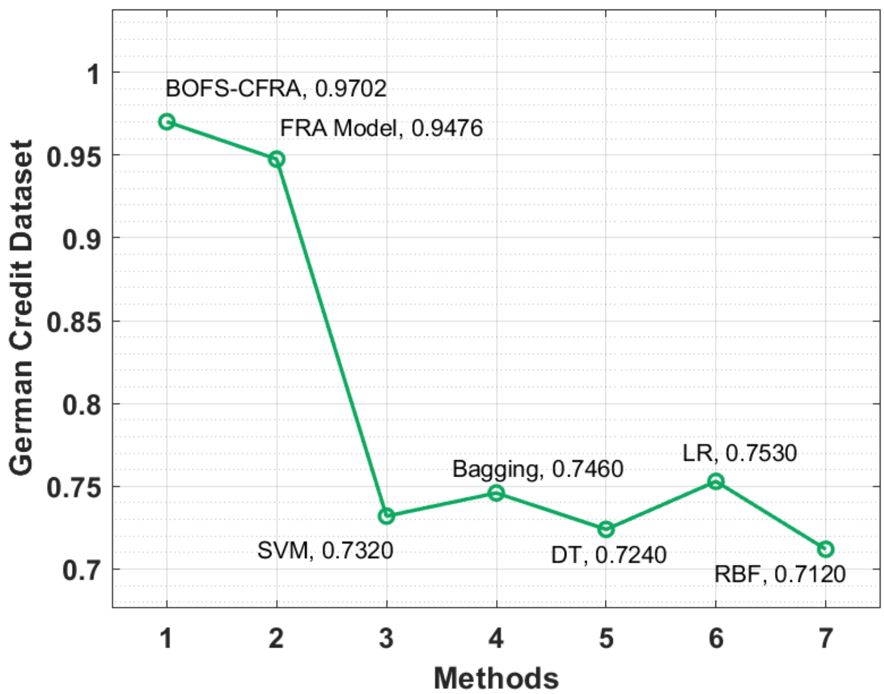


**Fig. 8. Comparative results analysis of BOFS-CFRA technique with traditional approaches on German Credit dataset**



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

Another comparative analysis with recent methods [24] are made in Table 6. A comprehensive result analysis of the BOFS-CFRA technique on the test Australian credit dataset is shown in Fig. 9. The figure portrayed that the SVM-IGDFS, SVM-GAW, KNN-IGDFS, KNN-GAW, and NB-IGDFS techniques have accomplished reduced accuracy of 0.9075, 0.8901, 0.8675, 0.8565, and 0.8797 respectively. Besides, the FRA model has resulted to a moderately considerable outcome with the accuracy of 0.9568. However, the proposed BOFS-CRFA technique has accomplished improved outcome with the accuracy of 0.9658.

**Table 6 Comparative study of BOFS-CFRA technique with recent methods**

| Methods | Australian Credit | German Credit |
|---|---|---|
| BOFS-CFRA | 0.9658 | 0.9702 |
| FRA Model | 0.9568 | 0.9476 |
| SVM-IGDFS | 0.9075 | 0.8280 |
| SVM-GAW | 0.8901 | 0.8040 |
| KNN-IGDFS | 0.8675 | 0.7020 |
| KNN-GAW | 0.8565 | 0.7580 |
| NB-IGDFS | 0.8797 | 0.7730 |

A brief comparative result analysis of the BOFS-CFRA technique on the test German credit dataset is shown in Fig. 10. The figure portrayed that the SVM-IGDFS, SVM-GAW, KNN-IGDFS, KNN-GAW, and NB-IGDFS techniques have accomplished reduced accuracy of 0.8280, 0.8040, 0.7020, 0.7580, and 0.7730 respectively. Also, the FRA model has resulted to a moderately considerable outcome with the accuracy of 0.9476. However, the proposed BOFS-CRFA technique has accomplished improved outcome with the accuracy of 0.9702.

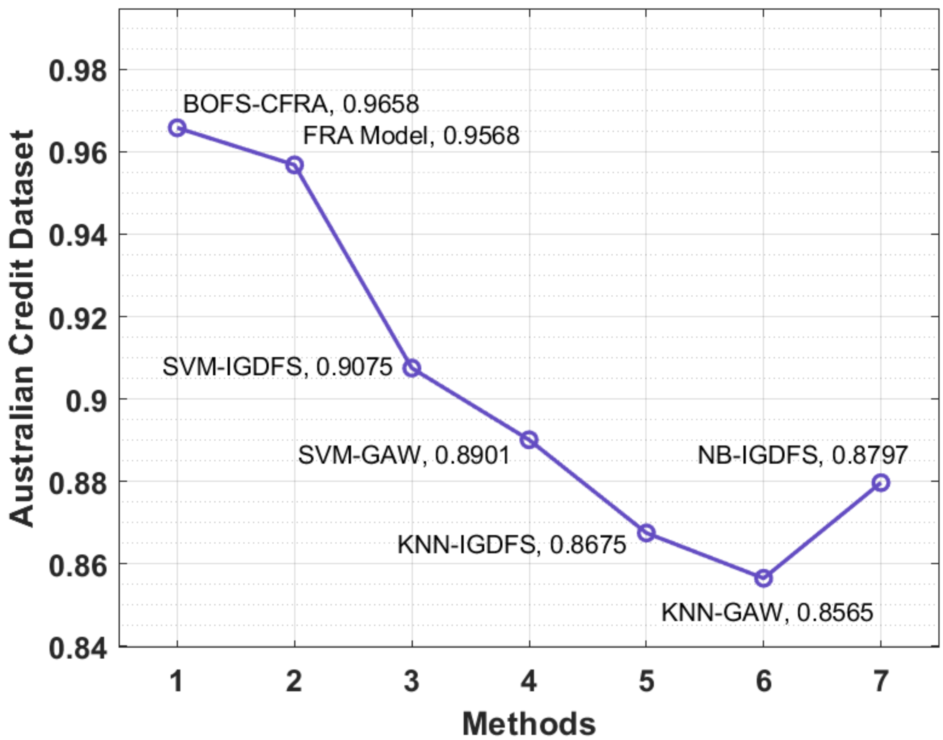


**Fig. 9. Comparative results analysis of BOFS-CFRA technique on Australian Credit dataset**



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

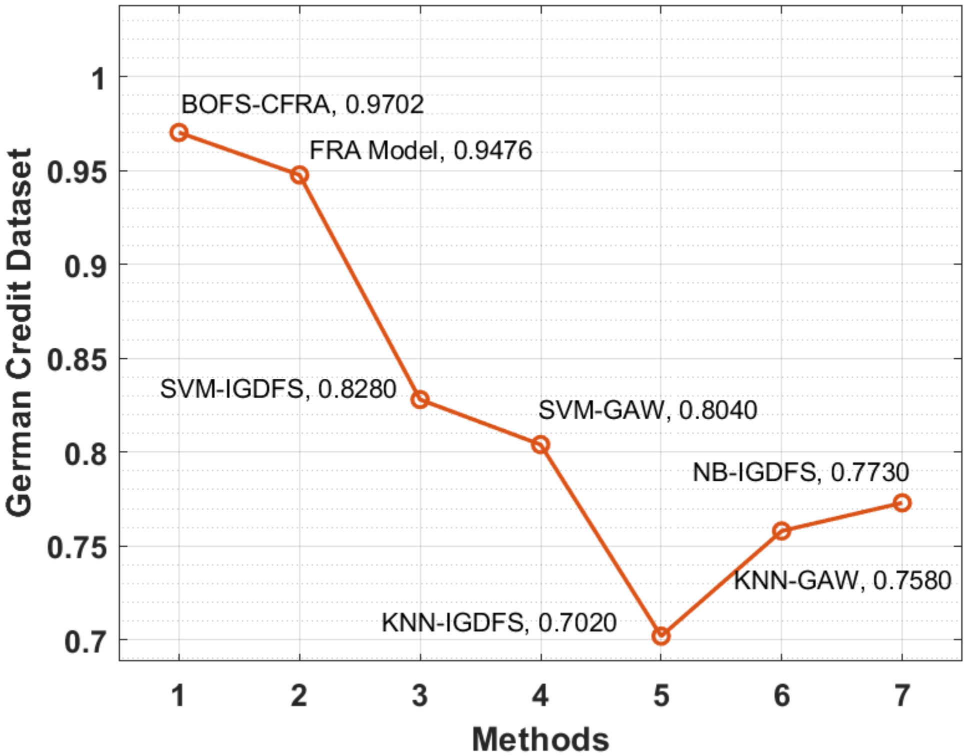


**Fig. 10. Comparative results analysis of BOFS-CFRA technique on German Credit dataset**

## 4. Conclusion

In this study, a novel BOFS-CFRA technique is developed for financial risk assessment. The proposed BOFS-CFRA technique comprises pre-processing, K-means clustering, BOA based feature selection, and FLNN based Classification. The proposed BOFS-CFRA technique involves pre-processing at the primary stage to get rid of unwanted data. In addition, K-means clustering approach is developed to group the financial data into clusters. Then, the BOFS technique is applied to choose the subset of features from the clustered data. Finally, the classification of financial risks takes place by the use of FLNN. In order to ensure the enhanced performance of the BOFS-CFRA technique, a series of simulations were carried out and the results are inspected under various measures. The simulation outcome portrayed the supremacy of the BOFS-CFRA technique over the other financial risk assessment models interms of several performance measures. In future, outlier detection and feature reduction approaches can be devised to enhance the assessment outcomes.

## References

[1] Lee, I. and Shin, Y.J., 2020. Machine learning for enterprises: Applications, algorithm selection, and challenges. Business Horizons, 63(2), pp.157-170.

[2] Gupta, A. and Lohani, M.C., 2022. Comparative Analysis of Numerous Approaches in Machine Learning to Predict Financial Fraud in Big Data Framework. In Soft Computing: Theories and Applications (pp. 107-123). Springer, Singapore.

[3] Gupta, A. and Lohani, M.C., 2022. Comparative Analysis of Numerous Approaches in Machine Learning to Predict Financial Fraud in Big Data Framework. In Soft Computing: Theories and Applications (pp. 107-123). Springer, Singapore.

[4] Gu, S., Kelly, B. and Xiu, D., 2018. Empirical asset pricing via machine learning (No. w25398). National bureau of economic research.

[5] Chen, Y., Zheng, W., Li, W. and Huang, Y., 2021. Large group activity security risk assessment and risk early warning based on random forest algorithm. Pattern Recognition Letters, 144, pp.1-5.

[6] Addo, P.M., Guegan, D. and Hassani, B., 2018. Credit risk analysis using machine and deep learning models. Risks, 6(2), p.38.



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021

[7] Paiva, F.D., Cardoso, R.T.N., Hanaoka, G.P. and Duarte, W.M., 2019. Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. Expert Systems with Applications, 115, pp.635-655.
[8] Kunnathuvalappil Hariharan, N., 2017. Predictive model building for driver-based budgeting using machine learning. Predictive Model Building for Driver-Based Budgeting Using Machine Learning (June 5, 2017).
[9] Polak, P., Nelischer, C., Guo, H. and Robertson, D.C., 2020. "Intelligent" finance and treasury management: what we can expect. AI & SOCIETY, 35(3), pp.715-726.
[10] Henckaerts, R., Côté, M.P., Antonio, K. and Verbelen, R., 2021. Boosting insights in insurance tariff plans with tree-based machine learning methods. North American Actuarial Journal, 25(2), pp.255-285.
[11] Le, H.H. and Viviani, J.L., 2018. Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. Research in International Business and Finance, 44, pp.16-25.
[12] Dumitrescu, E., Hue, S., Hurlin, C. and Tokpavi, S., 2021. Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. European Journal of Operational Research.
[13] Roa, L., Rodríguez-Rey, A., Correa-Bahnsen, A. and Valencia, C., 2021. Supporting Financial Inclusion with Graph Machine Learning and Super-App Alternative Data. arXiv preprint arXiv:2102.09974.
[14] Chang, Y.C., Chang, K.H. and Wu, G.J., 2018. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. Applied Soft Computing, 73, pp.914-920.
[15] Lee, T.K., Cho, J.H., Kwon, D.S. and Sohn, S.Y., 2019. Global stock market investment strategies based on financial network indicators using machine learning techniques. Expert Systems with Applications, 117, pp.228-242.
[16] Munkhdalai, L., Munkhdalai, T., Namsrai, O.E., Lee, J.Y. and Ryu, K.H., 2019. An empirical comparison of machine-learning methods on bank client credit assessments. Sustainability, 11(3), p.699.
[17] Zhu, Y., Zhou, L., Xie, C., Wang, G.J. and Nguyen, T.V., 2019. Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. International Journal of Production Economics, 211, pp.22-33.
[18] Bao, W., Lianju, N. and Yue, K., 2019. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. Expert Systems with Applications, 128, pp.301-315.
[19] Na, S., Xumin, L. and Yong, G., 2010, April. Research on k-means clustering algorithm: An improved k-means clustering algorithm. In 2010 Third International Symposium on intelligent information technology and security informatics (pp. 63-67). Ieee.
[20] Arora, S. and Singh, S., 2019. Butterfly optimization algorithm: a novel approach for global optimization. Soft Computing, 23(3), pp.715-734.
[21] https://transpireonline.blog/2020/04/15/butterfly-optimization-algorithmboa-to-solve-engineering-problems/
[22] Dehuri, S., Roy, R., Cho, S.B. and Ghosh, A., 2012. An improved swarm optimized functional link artificial neural network (ISO-FLANN) for classification. Journal of Systems and Software, 85(6), pp.1333-1345.
[23] Al-Jumeily, D., Ghazali, R. and Hussain, A., 2014. Predicting physical time series using dynamic ridge polynomial neural networks. PLOS one, 9(8), p.e105766.
[24] Acharya, S., Pustokhina, I.V., Pustokhin, D.A., Geetha, B.T., Joshi, G.P., Nebhen, J., Yang, E. and Seo, C., 2021. An improved gradient boosting tree algorithm for financial risk management. Knowledge Management Research & Practice, pp.1-12.



DOI: https://doi.org/10.54216/AJBOR.020104
Received: August 28, 2020 Accepted: May 01, 2021