

# A Survey on Machine Learning and Statistical Techniques in Bankruptcy Prediction

S. Sarojini Devi and Y. Radhika

**Abstract**—The bankruptcy prediction model (BPM) is eminently essential for financial institutions to verify the creditworthiness of companies or management. The inability to accurately predict the bankruptcy can destroy the effects of socio-economics. Hence, it is significant to offer financial decision makers with efficient bankruptcy prediction to forestall these loss states. This paper presents a comprehensive review based on various statistical and machine learning techniques to address the issue of bankruptcy prediction. The statistical techniques include, linear discriminant analysis (LDA), multivariate discriminant analysis (MDA) and logistic regression (LR), and machine learning techniques such as artificial neural networks (ANN), support vector machines (SVM) and decision trees. Traditional statistical techniques were used to handle large data sets without affecting the prediction performance. Furthermore, the machine learning techniques provide greater prediction accuracy than the traditional statistical techniques for smaller data sets. Besides, optimization techniques, such as genetic algorithm (GA) and particle swarm optimization (PSO), were integrated with machine learning techniques to further improve the prediction accuracy for large data sets. This paper conducts a comparative analysis of the various techniques used based on their corresponding benefits and limitations. In our future work, the prediction of bankruptcy may be improved by integrating other heuristic evolutionary algorithms with the machine learning techniques using the Apache Mahout tool.

**Index Terms**—Artificial neural networks (ANN), bankruptcy prediction model (BPM), optimization techniques and support vector machines (SVM).

## I. INTRODUCTION

In the recent past, bankruptcy prediction is one of the critical issues faced by decision makers in the field of finance [1]. Bankruptcy causes various failures affecting economics, stockholders, customers, management, and so on. The main focus of bankruptcy prediction is to find out whether a financial firm or organization will go bankrupt or not [2]. The bankruptcy or financial concern is a position that proceeds, when a financial firm verifies its financial obligations. The development of technology supports the acquisition of information on the risk condition of an organization in various manners, such as mass media and professional agencies [3].

The causes of bankruptcy and loss of business are based on factors, such as financial, economic, disaster, and fraud. In bankruptcy, the economic factors consist of poor location, weakness of industry and financial factors consist of

enormous obligations [4]. Therefore, the accurate prediction of bankruptcy is a more important problem in financial and management problem [5]. The prediction of bankruptcy is a binary classification problem, which includes two classes namely bankrupt and non-bankrupt [6]. Different elements of bankruptcy are secured obligations, repayment, property elimination, etc. Therefore, many of the current techniques used to predict financial failure and bankruptcy must be continuously improved.

The primary challenge of bankruptcy prediction starts with the selection of the prediction technique. Several recent and advanced techniques for predicting bankruptcy have been developed [7]. Statistical and machine learning techniques are the two broad categories used to predict bankruptcy [8], [9]. Statistical techniques include linear discriminant analysis (LDA), multi-discriminant analysis (MDA) [10], logistic regression (LR), etc. Machine learning techniques, such as artificial neural networks (ANN), SVM [11], Bayesian network and decision trees [12]. In addition, to improve the prediction, accuracy optimization algorithms, such as genetic algorithm (GA) [13] and particle swarm optimization (PSO), have been combined with machine learning techniques [14].

This research study conducts a review on the various techniques used to predict bankruptcy in financial and organization fields. Section II is a literature review, which discusses the various statistical and machine learning techniques that have been proposed by various researchers. Section III demonstrates the comparison of various techniques with their advantages and limitations. Finally, Section IV concludes the paper with the future scope for research in this area.

## II. LITERATURE REVIEW

There are several techniques proposed by various authors used to predict bankruptcy. Fig. 1 shows the different types of bankruptcy prediction techniques.

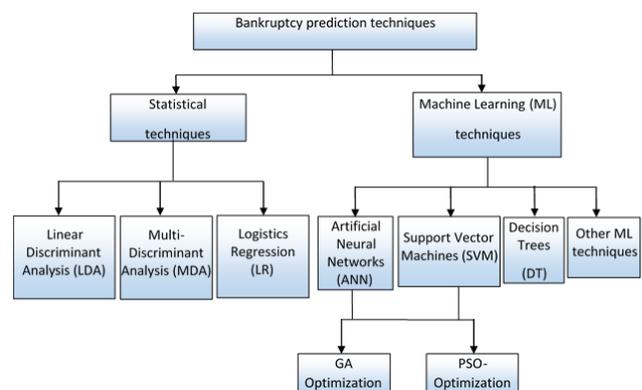


Fig. 1. Various techniques used for bankruptcy prediction.

Manuscript received February 12, 2018; revised April 4, 2018.

The authors are with the Dept. of CSE, GIT, GITAM University, Visakhapatnam, Andhra Pradesh, India (e-mail: ssarojini2012@gmail.com, radhika@gitam.edu).

## A. Statistical Techniques

### 1) Linear discriminant analysis (LDA)

A comparison between conventional statistical techniques for distress classification and prediction, namely linear discriminant (LDA) with an artificial intelligence algorithm, has been analyzed [15]. The experimental outcomes are a unit of a larger attempt involving separate designs for industrial, retail, and construction firms. The results demonstrate a balanced degree of accuracy and other advantageous properties between LDA and neural networks (NN). In addition, Altman et al. have pointed out the issues of 'black-box' NN systems, which consist of incorrect overfitting and weightings in the training phase, both of which have an impact on predictive accuracy. Both types of diagnostic approaches display acceptable classification performance.

### 2) Multivariate discriminant analysis (MDA)

Multivariate discriminant analysis (MDA) was a breakthrough in credit risk assessment that occurred when [16] proposed a study of bankruptcy between manufacturing firms that attained related classification results. The technique is based on a minimization in the variance between observations in the same group and a maximization distance between observations in different groups [17]. The technique produces a score and the observation is classified into a group depending on the score relative to an arbitrary cut-off value. The contrary considerations of MDA require the normally distributed variables and sensitivity to outliers that formed the LR, namely the multivariate model for credit risk modeling.

### 3) Logistics regression (LR)

A new model has been presented for the prediction of company failure in the French context and an improvement to the existing practical study by adding two major novel schemes [18]. The prediction of bankruptcy is achieved by the application of all the variables in LR, which are not affected by multicollinearity. However, it has the major challenge of convergence issues in maximum likelihood computation. Besides, this paper shows that the practical application of the partial least squares method to firms essential to several activity areas offers good results. The application of this method to two samples of healthy and failing companies permitted the researchers to obtain good results and to present a prediction model more appropriate than that acquired using the parametric method. This result provides bankers and investors an informational description of practices based on failure indicators. This permits them to achieve a comparative analysis of their individual models for the prediction of financial distress.

The integrated method used for bankruptcy prediction was proposed and validated for Malaysian firms [19]. The integrated method was varied from conventional researches by integrating the Ohlson logit method, Springate-Canadian method, and macroeconomic parameters. This method was improved based on the financial and macroeconomic constructs. The results showed that logistic regression achieves better performance and it was used to validate the method. Besides, the results demonstrate that the capacity of the method to predict correctly is 100% for both samples.

## B. Machine Learning Techniques

### 1) Artificial neural networks (ANN)

A hybrid discriminant neural network (HDNN) has been proposed based on discriminant diagnosis, self-organizing maps and back-propagation NN to predict risk failure [20]. This method considers the way firms move in failure space over a period of three years and the limitation of missing data to describe their risk failure. They implemented a hybrid model of NN to the risk failure of Moroccan firms considering the data reliability and availability. The implementation results demonstrate that the HDNN provides good accuracy when compared to discriminant analysis. Furthermore, it verifies the hypothesis that a method trained with suitable variables provides enhanced results than with generally used variables with the missing data. This hybrid method is a useful technique for investors and stakeholders to describe the risk profile.

In [21], an efficient hybrid approach using clustering algorithm and genetic algorithm with artificial neural networks (GA-ANN) algorithm has been proposed to equalize the proportion between the minority and majority classes. This paper provides a better appropriate data set for both reducing data imbalance and enhancing the classification accuracy. They extracted the suitably balanced data set collected from an optimal situation for the ANN approach. Besides, the approach extracted the knowledge-based prediction of the data set and classified the data through the clustering algorithm while employing concurrent optimization for the ANN approach. This approach was effectively applied to the bankruptcy prediction issue using financial data where the proportion of medium-sized and small-sized bankruptcy firms is very small compared to the non-bankruptcy firms in the manufacturing industry.

A prediction technique based on the application and advantages of data mining has been developed in the domain of corporate bankruptcy [22]. It evaluates a data set of 120 industries based on different data mining techniques. This neural network method achieves better results in the prediction of corporate bankruptcy. Hence, the appropriate selection of data mining techniques was used to improve the accuracy of the bankruptcy prediction model. Financial organizations can significantly gain advantage by using these various prediction models as they permit them to predict the condition of businesses ultimately and make decisions respectively.

In [23], a technique of NN was implemented to the bankruptcy of U.S. Banks, considering the particular characteristics of the recent financial crisis. They integrate multilayer perceptron and self-organizing maps to offer a technique that shows the probability of distress up to 3 years before bankruptcy occurred. Using data from the Federal Deposit Insurance Corporation between 2002 and 2012, the experimental results illustrate that unsuccessful banks are more focused in real estate loans. The situation is partially due to risky expansion that results in less balance and interest income. After drawing the profile of distressed banks, they implemented a technique to predict failures and a method to analyse bank risk in the short, medium, and long-term using bankruptcies generated from May 2012 to December 2013 in U.S. banks. The techniques can predict 96.15% of the failures

in this period and achieve better performance than conventional techniques used for bankruptcy prediction.

The modeling and performance evaluation of the neural network and particle swarm optimized fuzzy neural network (PSO-FNN) technique has been used to solve the bankruptcy prediction problem [24]. The technique uses a set of financial data chosen from listed corporations in the Chinese stock market. To compare the two models, the ST (special treated) corporations represent the bankrupt corporations and non-ST corporations represent the non-bankrupt corporations. The proportion of ST corporations in the entire sample was designed as 25% based on the actual ratio. The practical results represent that the performance of the PSO-FNN technique is much better than neural network techniques in the field of bankruptcy prediction.

A multi-industry examination of the bankruptcy of Korean industries has been presented using a back-propagation neural network (BNN) [25]. The industries consist of construction, retail, and manufacturing. It offers an industry related method to predict bankruptcy by choosing the appropriate variables. The prediction accuracy of the BNN is better when compared to that of MDA. The results represent that the prediction using an industry sample performs better than the prediction using the total sample, which is not classified based on the industry by 6–12%. The prediction accuracy of bankruptcy using BNN is greater than that of MDA.

Corporate failure prediction using previous financial data is a well-documented research method. Previously, the prediction of bankruptcy has been based on statistical techniques such as multiple discriminant analysis, logit, and probit. However, a number of techniques have shown that artificial intelligence, such as NNs, is another technique that can be used for classification issues to which conventional statistical techniques have long been employed. Also, numerous theoretical and experimental studies have confirmed the benefits of NNs in the classification method. However, the major drawback of NNs is that the user cannot realize the rules of the NNs technique. In [26], GA techniques were investigated and the results showed that GA could be employed to bankruptcy prediction modeling. The benefit of this model using GA was the possibility to extract rules that are easy to understand for users, such as expert systems. The preliminary results demonstrate that the rule extraction method using GAs for bankruptcy prediction modeling is promising.

### 2) *Support vector machines (SVM)*

The efficiency of SVM for predicting bankruptcy has been investigated [27]. However, although the back-propagation neural network works well in the pattern recognition tasks, the technique has a few drawbacks, thus it is necessary to determine a suitable technique and optimal solution. Besides, loading of the training set into the network is required to find the weights of the network. The results show that the SVM classifier performs better than the BPN for the prediction of corporate bankruptcy. Moreover, the bankruptcy prediction using SVM showed a better performance than BPN in terms of its accuracy. The authors have also tested the effect of the variability in performance based on various values of the metrics in SVM.

A hybrid switching particle swarm optimization (SPSO) and support vector machine (SVM) technique has been developed for bankruptcy prediction [28]. The major objective was to use the better stability and descriptive power of the SVM. In general, the established switching PSO algorithm was applicable to predicting the optimal parameter values of the radial basis function kernel of the SVM. Then, the SVM parameters were optimized by the switching PSO algorithm. Finally, the combined technique was effectively employed to identify bankruptcy. The experimental results show that the SPSO-SVM performs better than SVM, GA-SVM, and PSO-SVM.

In [29], several conventional statistical techniques used for financial distress prediction and a few unconventional techniques, such as NN, decision tree classification, and evolutionary estimation techniques, were compared. For this analysis, the data was collected from two hundred Taiwan stock exchange corporation listed companies. The experimentation was conducted based on the sum of 42 ratios consisting of 33 financial, eight non-financial and one integrated macroeconomic index using principle component analysis (PCA) to extract the appropriate variables. In this paper, the authors proposed four major contributions such that (i) nearly 80% fewer financial ratios have been obtained by the PCA algorithm, which provides high accuracy prediction, (ii) the conventional statistical techniques are better in the case of large data sets without satisfactory prediction performance and, on the other hand, data mining techniques obtain better performance with smaller data sets, (iii) the practical results demonstrate that C5.0 and CART offer the optimal prediction accuracy for bankruptcies, and (iv) SVMs with evolutionary estimation support a good balance of high accuracy prediction performance. Hence, the simulation results demonstrate that the PSO combined with SVM (PSO-SVM) technique can be used to recognize potential financial distress.

### 3) *Decision trees*

The data mining techniques of BPMs, such as NNs and SVM, are well suited for financial data. However, these algorithms are considered as black-box techniques due to their unambiguousness. In addition, decision trees are more accurate for users, however, the numerous number of conditions result in creating incomprehensibility. The number of conditions attained from decision tree algorithms is restricted to some degree through setting various minimum support levels. Decision tree algorithms used for bankruptcy data with the application of accuracy and a number of conditions have been proposed [30]. For bankruptcy data, the decision tree algorithm achieved relatively more accurate when compared to NN and SVM, but there were more condition nodes than required.

Data mining and machine learning techniques are important in bankruptcy prediction. The conventional techniques have been developed with various multiple-classifiers that outperform several single classifiers. There are several issues that affect the performance while constructing the classifiers. In [31], the method applied various classifier techniques, such as SVM, multilayer perceptron (MLP) neural networks and decision trees (DT) to

bankruptcy prediction. The experimental results demonstrate that the DT ensembles composed of 80–100 classifiers using the boosting method performed best when compared to the other classifiers.

#### 4) Other machine learning techniques

A genetic algorithm based two-step classification method has been developed that permits both choosing the significant factors and adapting the model itself to an application [32]. The various classifier techniques are trained in the first step and integrated into the voting ensemble in the second step. The integration of random sampling and feature selection techniques were used to ensure the necessary diversity level of the classifiers in the first step. Genetic algorithms are employed in the feature selection step and then the weights determined in the ensemble. The features of this method were verified on a balanced set of data, which consist of 912 observations of Russian companies and 55 features of micro-business environment factors. The results show that a highest accuracy of 0.934 was achieved. Besides, it also demonstrated the most balanced precision-recall ratio. It obtains the bankrupts and non-bankrupts (precision = 0.910) and (recall = 0.953) rather accurately when compared to the other tested methods. So, the method permits one to enhance the benefits and alleviate the weaknesses inherent in the ordinary classifiers, enabling business decision support with higher reliability.

Corporate bankruptcy prediction has become an increasingly important problem for financial institutions due to the advanced and recent financial crisis, and European debt crisis. A statistical and intelligent method has been developed for bankruptcy prediction [33]. Recent researches recommend ensemble learning techniques have potential applicability in corporate bankruptcy prediction. A novel and enhanced boosting, FS-boosting, was implemented to recognize corporate bankruptcy. By injecting a feature selection scheme into boosting, FS-boosting, better performance can be attained. The experimental results demonstrate the effectiveness and feasibility of FS-boosting. Besides, the results reveal that FS-boosting could be used as another method for corporate bankruptcy prediction.

The verification of prediction techniques needs both an in-sample and out-of-sample estimation of their performance. However, TOPSIS (technique for order preference by similarity to ideal solution) classifiers limit a framework for performing their out-of-sample evaluation. To address this issue, an instance of case-based reasoning methodology was developed in which the k-nearest neighbor was trained on the results of a TOPSIS classifier [34]. They evaluated the performance of this framework based on a UK data set for bankrupt and non-bankrupt firms. The simulation results demonstrate its outstanding prediction performance. Furthermore, the results of the proposed framework are robust to a variety of implementation decisions.

In [35], a probabilistic point-of-view by employing Gaussian processes (GP) was considered in the context of bankruptcy prediction, comparing it against LR and SVM. Based on real-world bankruptcy data, an in-depth analysis was implemented showing that, also a probabilistic interpretation, the GP can efficiently enhance the bankruptcy prediction performance with high accuracy when compared

to the other techniques. Furthermore, they created a complete graphical visualization to enhance the understanding of the different attained performances, effectively gathering all the conducted experiments in a certain manner. Finally, entropy-based evaluation was used to improve the uncertainty handling properties offered by GP, which is important for prediction processes under extremely competitive and volatile business environments.

A kernel extreme learning machine (KELM) parameter tuning scheme has been presented using a swarm intelligence algorithm, namely gray wolf optimization (GWO) [36]. GWO simulates the social hierarchy and hunting role of gray wolves in nature. This method was adopted to build an efficient KELM approach for bankruptcy prediction. The derived GWO-KELM model was precisely compared with three KELM algorithms, such as particle swarm optimization based KELM, a genetically algorithm based KELM, grid search technique based KELM and improved extreme learning machine on two real-time data sets through 10-fold cross-validation analysis. The simulation outcomes confirm the superiority of the implemented approach using the classification accuracy and computational time. Hence, the GWO-KELM prediction approach was confirmed as a robust early warning tool with superior performance for bankruptcy prediction.

A cost-effective bankruptcy prediction based on a hybrid algorithm and learning vector quantization (LVQ) technique has been proposed [37]. The asymmetric costs of two types of errors were combined into the evolutionary classification model used for the prediction of bankruptcy. The genetic algorithm is used to search the solution space for an optimal representation of the LVQ technique.

The simulation was conducted on real-life bankruptcy data and showed that the hybrid technique leads to a lower entire misclassification cost when compared to the Sole-LVQ, which does not include the costs and cost-LVQ, which uses a local search strategy.

### III. A COMPARISON OF THE VARIOUS TECHNIQUES

The prediction scheme uses financial and non-financial data to build a bankruptcy prediction model and analyse the performance of the scheme based on four performance metrics including the accuracy, precision, sensitivity, and specificity [29]. These metrics are estimated based on the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) so as to calculate the accuracy, precision, sensitivity, and specificity. In this, TP and TN denote the classifier attaining the correct classification and FP and FN represents the classifier obtaining the wrong classification results.

#### a) Accuracy

Accuracy is defined as the percentage of correctly classified cases and is used to estimate the classification performance.

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

#### b) Precision

Precision is defined as the number of positive cases

classified.

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

c) Sensitivity

Sensitivity is defined as the ratio of the number of true positive cases to the number of total positive cases.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

d) Specificity

Specificity is defined as the ratio of the number of true negative cases to the total number of negative cases.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

The performance analysis of various bankruptcy prediction algorithms are discussed based on the parameters, such as accuracy, precision, sensitivity, and specificity [29]. Machine learning algorithms, such as ANN and SVM achieve better performance in terms of the four performance metrics. The authors improved the performance by integrating SVM with optimization algorithms, such as GA and PSO. The decision tree algorithm obtained the highest sensitivity of 100% when compared to the other algorithms. From this overall analysis, the SVM-PSO prediction algorithm achieved the maximum accuracy, precision, and specificity when compared with the other statistical and machine learning algorithms. Table I and Table II show the performance evaluation charts of the bankruptcy prediction model. Fig. 2 and Fig. 3 show the performance comparison plots for accuracy and precision, and sensitivity and specificity, respectively.

A comparative analysis of the various bankruptcy prediction models is shown in Table III. This comparison table provides the advantages and research gaps of each technique.

TABLE I: PERFORMANCE ANALYSIS OF THE VARIOUS BANKRUPTCY PREDICTION ALGORITHMS IN TERMS OF THEIR ACCURACY AND PRECISION [29].

Algorithms	Accuracy (%)	Precision (%)
Linear discriminant analysis (LDA)	75.47	68.75
Logistic regression (LR)	79.25	71.87
Decision trees (DT)	86.79	78.12
Learning vector quantization (LVQ)	87.5	85
Artificial neural networks (ANN)	90	94.11
Support vector machines (SVM)	90	94.11
Support vector machines with genetic algorithm (SVM-GA)	92.5	94.44
Support vector machines with particle swarm optimization algorithm (SVM-PSO)	<b>95</b>	<b>94.73</b>

TABLE II: PERFORMANCE ANALYSIS OF THE BANKRUPTCY PREDICTION ALGORITHMS USING THE SENSITIVITY AND SPECIFICITY [29].

Algorithms	Sensitivity (%)	Specificity (%)
Linear discriminant analysis (LDA)	88	64.28
Logistic regression (LR)	92	67.85
Decision trees (DT)	<b>100</b>	75
Learning vector quantization (LVQ)	89.47	<b>95.23</b>
Artificial neural networks (ANN)	84.21	85.23
Support vector machines (SVM)	84.21	<b>95.23</b>
Support vector machines with genetic algorithm (SVM-GA)	89.47	<b>95.23</b>
Support vector machines with particle swarm optimization algorithm (SVM-PSO)	94.73	<b>95.23</b>

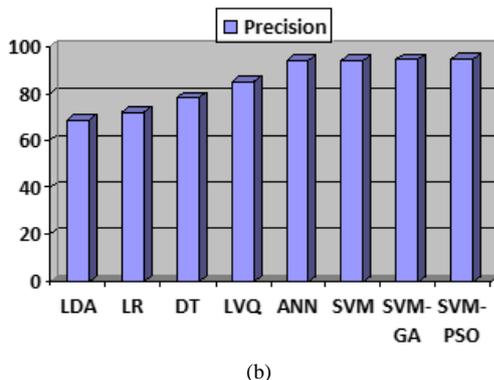
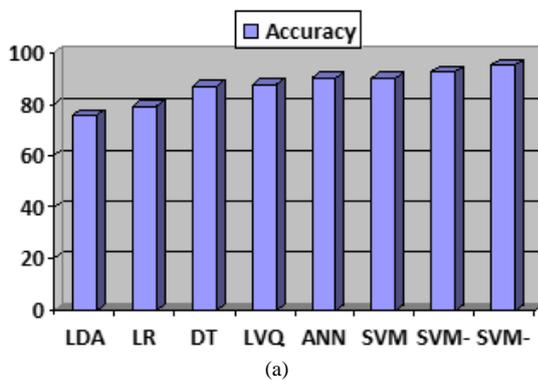


Fig. 2. Performance evaluation of the bankruptcy prediction models: (a) Accuracy and (b) precision.

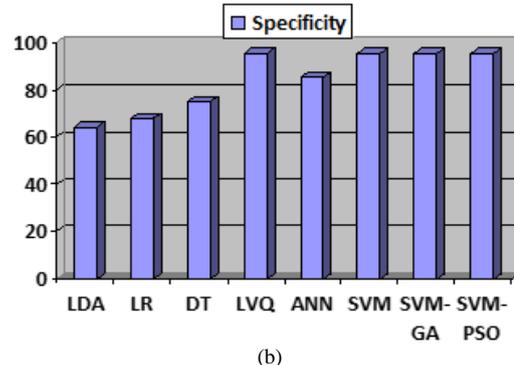
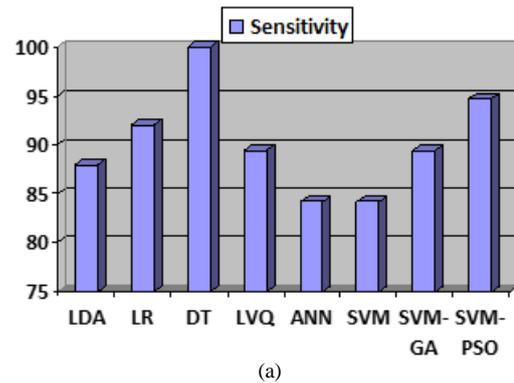


Fig. 3. Performance evaluation of the bankruptcy prediction models: (a) Sensitivity and (b) specificity.

TABLE III: A COMPARATIVE ANALYSIS OF THE VARIOUS BANKRUPTCY PREDICTION MODELS.

Paper Title	Technique used	Outcomes	Research Gap
Corporate distress diagnosis: Comparisons using linear discriminant analysis and NNs	Linear discriminant analysis (LDA)	Achieved the prediction accuracy of 90%.	More time consumption with moderate complexity.
Detecting credit card fraud by modified Fisher discriminant analysis	Multivariate discriminant analysis (MDA)	Better performance when compared to LDA.	More complexity.
Bankruptcy prediction using Partial Least Squares Logistic Regression	Logistic regression (LR)	More appropriate results are achieved.	Convergence issue in maximum likelihood computation.
Bankruptcy prediction using data mining techniques	Artificial neural networks (ANN)	Obtained good performance with better prediction accuracy.	It requires more training data sets.
A comparative study of classifier ensembles for bankruptcy prediction	Decision trees with boosting method	Higher accuracy performance when compared to ANN and SVM.	Decision trees without boosting method provide less accuracy.
A hybrid switching PSO algorithm and support vector machines for bankruptcy prediction	Particle swarm optimization-support vector machines (PSO-SVM)	PSO-SVM achieves better performance compared to SVM, GA-SVM.	It is very difficult when the number of data sets is increased.

#### IV. CONCLUSIONS

In this paper, the literature pertaining to bankruptcy prediction techniques has been rigorously explored and discussed. The performance of the various techniques used for predicting bankruptcy, such as statistical techniques and machine learning techniques, has been discussed [29]. Besides, the meta-heuristic optimization based machine learning algorithms are discussed, which are used to improve the prediction accuracy. The evaluation of these techniques was considered under various performance metrics, such as the accuracy, precision, sensitivity, and specificity. Furthermore, a comparative analysis of these techniques has been performed describing their benefits and limitations. From this inclusive review, the SVM-PSO prediction algorithm obtained the highest accuracy of 95%, precision (94.73%) and specificity (95.23%) when compared with the traditional statistical analyses and other machine learning algorithms. The Apache Mahout tool is an open source software used for data mining that provides a robust, scalable implementation of machine learning algorithms used for clustering, classification, and heuristic evolutionary

algorithms. Therefore, our future work will include the improvement of machine learning with other heuristic evolutionary optimization techniques used for the prediction of bankruptcy using the Apache Mahout tool with the aim to enhance the prediction accuracy.

#### REFERENCES

- [1] Z. S. Hua *et al.*, "Predicting corporate financial distress based on integration of support vector machine and logistic regression," *Expert Systems with Applications*, vol. 33, no. 2, pp. 434-440, 2007.
- [2] G. Kumar and S. Roy, "Development of hybrid boosting technique for bankruptcy prediction," in *Proc. International Conference on Information Technology*, 2016, pp. 248-253.
- [3] L. Becchetti, and J. Sierra, "Bankruptcy risk and productive efficiency in manufacturing firms," *Journal of Banking & Finance*, vol. 27, no. 11, pp. 2099-2120, November 2003.
- [4] W.-S. Chen and Y.-K. Du, "Using neural networks and data mining techniques for the financial distress prediction model," *Expert Systems with Applications*, vol. 36, no. 2, pp. 4075-4086, 2009.
- [5] M. A. Aziz and H. A. Dar, "Predicting corporate bankruptcy: Where we stand?" *Corporate Governance: The International Journal of Business in Society*, vol. 6, no.1, pp. 18-33, 2006.
- [6] W. Y. Lin, Y. H. Hu, and C. F. Tsai, "Machine learning in financial crisis prediction: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 421-436, November 2011.
- [7] P. R. Kumar and V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques — A review," *European Journal of Operational Research*, vol. 180, no. 1, pp. 1-28, July 2007.
- [8] M. Y. Chen, "Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches," *Computers & Mathematics with Applications*, vol. 62, no. 12, pp. 4514-4524, December 2011.
- [9] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent financial statements," *Expert Systems with Applications*, vol. 32, no. 4, pp. 995-1003, May 2007.
- [10] S. Lee and W. S. Choi, "A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis," *Expert Systems with Applications*, vol. 40, no. 8, pp. 2941-2946, June 2013.
- [11] B. E. Erdogan, "Prediction of bankruptcy using support vector machines: an application to bank bankruptcy," *Journal of Statistical Computation and Simulation*, vol. 83, no. 8, pp. 1543-1555, August 2013.
- [12] S. Cho, H. Hong, and B. C. Ha, "A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: For bankruptcy prediction," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3482-3488, April 2010.
- [13] K. S. Shin and Y. J. Lee, "A genetic algorithm application in bankruptcy prediction modeling," *Expert Systems with Applications*, vol. 23, no. 3, pp. 321-328, October 2002.
- [14] S. H. Min, J. Lee, and I. Han, "Hybrid genetic algorithms and support vector machines for bankruptcy prediction," *Expert Systems with Applications*, vol. 31, no. 3, pp. 652-660, October 2006.
- [15] E. I. Altman, G. Marco, and F. Varetto, "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)," *Journal of Banking & Finance*, vol. 18, no. 3, pp. 505-529, 1994.
- [16] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589-609, 1968.
- [17] N. Mahmoudi, and E. Duman, "Detecting credit card fraud by modified Fisher discriminant analysis," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2510-2516, 2015.
- [18] S. B. Jabeur, "Bankruptcy prediction using partial least squares logistic regression," *Journal of Retailing and Consumer Services*, vol. 36, pp. 197-202, May 2017.
- [19] S. Bakar, I. T. W. Kiong, and A. M. Nassir, "Examining and validating a bankruptcy prediction models in Malaysia," in *Proc. 2012 International Conference on in Statistics in Science, Business, and Engineering*, pp. 1-5, September 2012.
- [20] F. Z. Azayite and S. Achhab, "Hybrid discriminant neural networks for bankruptcy prediction and risk scoring," *Procedia Computer Science*, vol. 83, pp. 670-674.
- [21] H. Kim, N. O. Jo, and K. S. Shin, "Optimization of cluster-based evolutionary undersampling for the artificial neural networks in

- corporate bankruptcy prediction,” *Expert Systems with Applications*, vol. 59, pp. 226-234, 2016.
- [22] M. Wagle, Z. Yang, and Y. Benslimane, “Bankruptcy prediction using data mining techniques,” in *Proc. 8th International Conference on Information and Communication Technology for Embedded Systems*, May 2017, pp. 1-4.
- [23] F. J. L. Iturriaga and I. P. Sanz, “Bankruptcy visualization and prediction using neural networks: A study of US commercial banks,” *Expert Systems with Applications*, vol. 42, no. 6, 2857-2869, 2015.
- [24] L. Rui, “A particle swarm optimized fuzzy neural network for bankruptcy prediction,” in *Proc. International Conference on Future Information Technology and Management Engineering*, October 2010, vol. 2, pp. 557-560.
- [25] S. Lee and W. S. Choi, “A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis,” *Expert Systems with Applications*, vol. 40, no. 8, pp. 2941-2946, 2013.
- [26] K. S. Shin and Y. J. Lee, “A genetic algorithm application in bankruptcy prediction modeling,” *Expert Systems with Applications*, vol. 23, no. 3, pp. 321-328, 2002.
- [27] K. S. Shin, T. S. Lee, and H. J. Kim, “An application of support vector machines in bankruptcy prediction model,” *Expert Systems with Applications*, vol. 28, no. 1, pp. 127-135, 2005.
- [28] Y. Lu, J. Zhu, N. Zhang, and Q. Shao, “A hybrid switching PSO algorithm and support vector machines for bankruptcy prediction,” in *Proc. 2014 International Conference on Mechatronics and Control*, IEEE, July 2014, pp. 1329-1333.
- [29] M. Y. Chen, “Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches,” *Computers & Mathematics with Applications*, vol. 62, no. 12, pp. 4514-4524, 2011.
- [30] D. L. Olson, D. Delen, and Y. Meng, “Comparative analysis of data mining methods for bankruptcy prediction,” *Decision Support Systems*, vol. 52, no. 2, 464-473, 2012.
- [31] C. F. Tsai, Y. F. Hsu, and D. C. Yen, “A comparative study of classifier ensembles for bankruptcy prediction,” *Applied Soft Computing*, vol. 24, 977-984, 2014.
- [32] Y. Zelenkov, E. Fedorova, and D. Chekrizov, “Two-step classification method based on genetic algorithm for bankruptcy forecasting,” *Expert Systems with Applications*, vol. 88, pp. 393-401, 2017.
- [33] G. Wang, J. Ma, and S. Yang, “An improved boosting based on feature selection for corporate bankruptcy prediction,” *Expert Systems with Applications*, vol. 41, no. 5, pp. 2353-2361, 2014.
- [34] J. Ouenniche, B. Pérez-Gladish, and K. Bouslah, “An out-of-sample framework for TOPSIS-based classifiers with application in bankruptcy prediction,” *Technological Forecasting and Social Change*, 2017.
- [35] F. Antunes, B. Ribeiro, and F. Pereira, “Probabilistic modeling and visualization for bankruptcy prediction,” *Applied Soft Computing*, 2017.
- [36] M. Wang, H. Chen, H. Li, Z. Cai, X. Zhao, C. Tong, and X. Xu, “Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction,” *Engineering Applications of Artificial Intelligence*, vol. 63, pp. 54-68, 2017.
- [37] N. Chen, B. Ribeiro, A. S. Vieira, J. Duarte, and J. C. Neves, “Hybrid genetic algorithm and learning vector quantization modeling for cost-sensitive bankruptcy prediction,” in *Proc. 2010 Second International Conference on Machine Learning and Computing*, February 2010, pp. 213-217.



**Saladi Sarojini Devi** is a research scholar in the Dept. of Computer Science and Engineering, GIT, GITAM University, Visakhapatnam. She holds a B.Tech(IT) from Pondicherry University and Master 's degree from Acharya Nagarjuna University in Computer Science Engineering. Her research interests are in data mining and big data.



**Y. Radhika** is currently a professor in the Dept. of Computer Science and Engineering, GIT, GITAM University, Visakhapatnam. Her educational qualifications are BE, M.Tech and Ph.D. She is currently engaged in the research project “Application of Data mining Techniques for weather data analysis” by the Sanctioning Agency, Department of Science and Technology (DST), Govt. of India.