



Bankruptcy Prediction Using Stacked Auto-Encoders

Makram Soui, Salima Smiti, Mohamed Wiem Mkaouer & Ridha Ejbali

To cite this article: Makram Soui, Salima Smiti, Mohamed Wiem Mkaouer & Ridha Ejbali (2020) Bankruptcy Prediction Using Stacked Auto-Encoders, Applied Artificial Intelligence, 34:1, 80-100, DOI: [10.1080/08839514.2019.1691849](https://doi.org/10.1080/08839514.2019.1691849)

To link to this article: <https://doi.org/10.1080/08839514.2019.1691849>



Published online: 14 Nov 2019.



Submit your article to this journal [↗](#)



Article views: 1332



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 10 View citing articles [↗](#)



Bankruptcy Prediction Using Stacked Auto-Encoders

Makram Soui^a, Salima Smiti^b, Mohamed Wiem Mkaouer^c, and Ridha Ejbali^d

^aCollege of computing and informatics, Saudi Electronic University, Saudia; ^bNational School of Computer Science, Manouba, Tunisia; ^cRochester Institute of Technology, Rochester, NY, USA;

^dNational School of Engineers of Gabès, Gabès, Tunisia

ABSTRACT

Bankruptcy prediction is considered as one of the vital topics in finance and accounting. The purpose of predicting bankruptcy is to build a predictive model that combines several econometrics parameters, which allow evaluating the firm financial status either bankrupt or non-bankrupt. In this field, various machine learning algorithms such as decision tree, support vector machine, and artificial neural network have been applied to predict bankruptcy. However, deep learning algorithms are experiencing a resurgence of interest. To this end, we propose a novel deep learning-based approach which includes both feature extraction and classification phase into one model for predicting bankruptcy of financial firms. Our approach combines Stacked Auto-Encoders (SAE) with softmax classifier. In the first stage, the stacked auto-encoders are employed to extract the best features from the training dataset. Second, a softmax classification layer is trained to predict the class label. We evaluate our proposed approach on the base of Polish and Darden datasets. The obtained results confirm the efficiency of the SAE with softmax classifier compared to other existing works to accurately predict corporate bankruptcy.

Introduction

In the actual economic situation, an early warning system for bankruptcy prediction become important tools that afford timely warning to employees, management, creditors, shareholders, and other interested parties. The main aim of the early warnings is to ensure economic stability and to take timely and effective strategic decisions to avert corporate bankruptcy. Hence, these tools may enhance the performance of commercial credit assignment. Furthermore, the number of corporate bankruptcies is essential to a country's economy, and it can be considered as an indicator of economic development (Van Gestel et al. 2003).

In addition, the economic, social costs and high individual encountered bankruptcies have prompted searches to better understand the bankruptcy

CONTACT Salima Smiti  salima.smiti92@gmail.com  National School of Computer Science, Manouba, Tunisia

This article has been republished with minor changes. These changes do not impact the academic content of the article. Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uaai.

risk (McKee and Lensberg 2002). Therefore, the risk of bankruptcy as an essential economic phenomenon and also impact the country's economy. Thus, the prediction of corporate bankruptcy has gained an increasingly important role for the economy and society since it has a crucial influence on the profitability of financial institutions and lending decisions. The main objective of the bankruptcy prediction is to assess a company's financial health and its future perspectives (Constand and Yazdipour 2011). In fact, bankruptcy prediction has become more critical since the emergence of the Basel II requirement. This accord emphasizes the importance of bankruptcy prediction and makes the need for accurate decision-making model. In addition, Basel II allows banks to measure the company's risk of bankruptcy by using their own internal models, the probability of default and the capital required to meet that loss (Andrés, Landajo, and Lorca 2012).

At present, most of the existing works treat the bankruptcy prediction as binary classification problems (Barboza, Kimura, and Altman 2017; Min and Jeong 2009; Yeh, Chi, and Lin 2014), divide the firms into two classes: bankrupt and non-bankrupt firms. For this reason, several classification methods have been used to predict the risk of bankruptcy. These methods can be divided into two main categories: (1) statistical methods and (2) artificial intelligence methods. The first category has been largely applied to develop models for predicting bankruptcy. It was beginning with (Altman 1968; Beaver 1966) who used multivariate discriminant analysis (MDA), linear probit (Zmijewski 1984) and logit (Ohlson 1980). There are, however, certain limitations inherent in the statistical techniques that decrease the algorithm's usability in the bankruptcy prediction. The biggest limitation lies in the assumption on linear separability, independence of the predictive variables, and multivariate normality (Karels and Prakash 1987; Ohlson 1980). However, these assumptions are violated by many common financial ratios. The second limitation is that the traditional technique uses a fixed function, and these assumptions of statistical method make them difficult to model complex financial systems (Yu, Yang, and Tang 2015). To tackle this drawback, several artificial intelligence has been suggested for building bankruptcy prediction models. The most popular artificial intelligence tools for bankruptcy prediction can be referred to support vector machine (SVM) (Barboza, Kimura, and Altman 2017; Erdogan 2013; Xie, Luo, and Yu 2011), artificial neural network (ANN) (Callejón et al. 2013; Jardin 2010), rough sets (McKee and Lensberg 2002; Sanchis et al. 2007), decision tree (DT) (Gepp and Kumar 2015; Gepp, Kumar, and Bhattacharya 2009; Ocal, Ercan, and Kadioğlu 2015), genetic algorithm (Shin and Lee 2002). The results obtained when using these techniques in building bankruptcy prediction model outperform the traditional statistical techniques (Chen 2011; Nguyen 2005; Ocal, Ercan, and Kadioğlu 2015). However, the techniques mentioned above have

some limitations, such as the time-consuming process (it is very expensive) and local maximum or minimum problems (Yu, Yang, and Tang 2015).

As for artificial intelligence technique is an umbrella term that includes many subsets such as Machine and Deep learning (DL). These two subsets represent the actual complexity of the human brain than others. Currently, DL is a new sub-field of machine learning research. It is also one of the most popular scientific research trends nowadays (Minar and Naher 2018). Deep learning has proved its effectiveness and its superiority over artificial intelligence techniques to solve the problems in recent years (Dey 2017). Moreover, deep learning is being increasingly popular and outperforms all traditional methods and machine learning methods in various field including language processing (Collobert and Weston 2008; Sutskever, Vinyals, and Le 2014), speech recognition (Bouallégue et al. 2016; Dahl et al. 2012; Hinton et al. 2012), computer vision (Krizhevsky, Sutskever, and Hinton 2012; Srinivas et al. 2016), and many other fields. However, there were few studies on bankruptcy prediction by using the deep learning algorithms. Therefore, we focus in this paper on building the stacked auto-encoders (SAE) deep learning algorithm accompanying with a softmax classifier to improve the bankruptcy prediction accuracy by correctly classify firms as bankrupt or non-bankrupt.

Actually, the SAE has been used in the various fields. It has been widely applied to solve the intractable problem in natural language processing (NLP), including document clustering, word embeddings, machine translation, paraphrase detection, and word embeddings (Jang, Seo, and Kang 2018). Also, the SAE is an effective technique successfully applied in the domain of human activity recognition (Mbarki, Ejbali, and Zaied 2017). However, there are few propositions and research studies about the use of the SAE in the economic or the financial field. Thus, we aim in this work to find out if they still work well in the economical domain by applying it with a real dataset in the task of bankruptcy prediction. In this paper, we assess the performance of the stacked auto-encoders with softmax classifier ones of the deep learning-based classification method to predict the bankruptcy of financial firms. This proposed model is divided into two processing stage: feature extraction stage and classification stage. First, the stacked auto-encoders (SAE) are used to reduce the dimensionality of the feature set to improve the accuracy of bankruptcy prediction model. Second, we apply the softmax layer to classify the samples. In this stage, firms have been classified into two classes: bankrupt or non-bankrupt. The results confirm the efficiency of the proposed algorithm compared with similar works used others classifiers through two performance indicator: accuracy and area under the ROC curve (AUC).

The rest of this paper is structured as follows: section 2 reviews the related work on the machine and deep learning algorithms applied to the bankruptcy classification problem. Section 3 describes the bankruptcy prediction model examined in the paper. Section 4 presents the conducted empirical analysis. Finally, we conclude the paper with some pointers toward future work in section 5.

Related Work

In literature, machine and deep learning algorithms have been effectively employed to solve the bankruptcy classification problem. Both of them are classified into artificial intelligence models and provide investigators with more accurate predictions. Here, we are going to brief some outstanding papers on the machine and deep learning algorithms for predicting bankruptcy.

Machine Learning Algorithms

Many machine-learning techniques have been successfully used to develop accurate bankruptcy prediction models such as DT, NN, and SVM, etc.

(Ocal, Ercan, and Kadioğlu 2015) used CHAID (Chi-square Automatic Interaction Detector) and C5.0 decision tree algorithms for predicting financial failure. Then, they compared the results of their proposed models. They have concluded that both CHAID and C5.0 decision tree models have an acceptable accuracy classification rates. Despite the C5.0 algorithm outperforms the CHAID algorithm for predicting unhealthy firms, it has lower result than the CHAID decision tree algorithm for predicting unhealthy firms. Moreover Fan, Liu, and Chen (2017) implemented and compared the Isolation forest, one-class SVM and multivariate Gaussian distribution techniques to predict bankruptcy for a real-world dataset of Polish companies. Considering the AUC (area under curve) criterion, the Isolation forest provided the best results compared to one-class SVM, multivariate Gaussian distribution, and other classification estimators. The result confirms that Isolation forest can successfully resolve the effect imbalanced learning problem.

Although the decision tree algorithms have gained popularity for the bankruptcy prediction because it's able to generate understandable rules and it's capable of dealing with continuous and categorical variables. DTs are prone to errors in bankruptcy classification problem based on multi-class. Furthermore, the process of growing and pruning a DT makes them computationally expensive to train. In this way, ANN and SVM were applied due to its high capability to solve a complex problem.

For instance, Jardin (2010) trained the neural network (NN) to propose a bankruptcy prediction model. This proposed model is implemented and validated using French bankruptcy dataset. The results confirm the efficiency of the proposed model regarding classifying and predicting corporate bankruptcy. Also, the result shows that the NN based on a set of attributes chosen with appropriate variable selection technique yields good results than a set selected with methods applied in the financial literature. In a continuous work, Kasgari et al. (2012) suggested the use of MLP for modeling bankruptcy prediction. As input, the MLP used four effective predictive financial

ratios: total liability to total assets, quick assets to total assets, operational income to sales, and ratios of sales to current assets ratio. Then, they have compared the obtained results with the MLP model with those obtained from Probit and with other existing techniques (logit, gene expression programming, and linear regression). The performance of this proposed method is experimented using Iranian dataset. The results show that the proposed MLP model provided the best results compared to other classifiers.

Despite the high accuracy of the neural network compared to other classification techniques, Ahn and Kim (2011) mentioned that there are some difficulties in their using. These are arising from the fact that several parameters to be set by heuristics and hence the model is exposed to overfitting. Thereby leading to poor performance of the model. Hence, a lot of research has been used in the SVM to construct an accurate bankruptcy prediction model, because the SVM can enhance the performance of a model by maximizing the margin to avoid overfitting (Erdogan 2013; Santoso and Wibowo 2018; Xie, Luo, and Yu 2011).

Santoso and Wibowo (2018) applied the SVM and linear discriminant analysis (LDA) combined with a variable selection technique for predicting bankruptcy of financial firms in Indonesia. The experimental results showed that the hybrid stepwise-SVM model achieved high performance compared with another model in practice. They concluded that SVM could be a powerful management tool that allows building an advanced decision support systems for corporate loans. In a previous study, Zhang et al. (2017) proposed an improved sequential minimal optimization (SMO) based on four-variable named FV-SMO. The main idea of this algorithm is to select four variables into the working set at each iteration. This proposed model is implemented and validated using China, German and Darden credit datasets. The results show that the FV-SMO algorithm achieved the highest accuracy for predicting bankruptcy compared with the five popular classification methods in credit risk assessment including MLP, DT, Logistic regression (LR), Baysenet, radial basic function (RBF).

The SVMs, however, are not perfect and not suitable for bankruptcy prediction. The major obstacle of SVM and ANN is their black box nature. This obstacle makes them less suited for predicting bankruptcy and they can lose a small percentage in terms of accuracy. Moreover, the black box models cannot adequately reveal information that maybe hidden in the data and the bankruptcy prediction remains difficult.

Recently, various studies have been applied ensemble techniques to predict corporate bankruptcy. (Zieba, Tomczak, and Tomczak 2016) proposed a new technique for financial distress prediction that makes use of eXtreme Gradient Boosting (XGB) for learning an ensemble of decision trees. In addition, the authors proposed a new approach named synthetic features which is a combination of the econometric measures and the arithmetical operation.

The purpose of synthetic features is to improve the overall performance. This proposed model is implemented and validated using real-life dataset of Polish companies from 2007 to 2013 (bankrupt) and from 2000 to 2012 for (still operating). This shows that boosting has successfully applied to predict corporate bankruptcy as compared to other techniques. Similarly, Wang, Ma, and Yang (2014) build a new and enhanced Boosting method called FS-Boosting based on feature selection to predict bankruptcy. The performance of this technique is analyzed using two real-world bankruptcy datasets based on three evaluation criteria: accuracy rate criterion, type I and II error criterion. This study highlighted the high performance of FS-Boosting for building an accurate and efficient bankruptcy model. However, this approach was effectively applied to solve the problem of financial distress using a small and balanced dataset.

More machine learning algorithms have failed to extract non-linear and complex patterns from large heterogeneous data. However, DL permits the use of linear models for a large quantity of data analysis tasks by extracting such features, such as prediction and classification; which is important when building models to transact with the large amount of data (Najafabadi et al. 2015). In this paper, we explore the application of DL algorithms such as stacked auto-encoders plus softmax classifier to improve the accuracy of bankruptcy prediction model.

Deep Learning Algorithms

In the last few years, deep learning has drawn much research attention, by outperforms machine learning techniques such as kernel machine, in several important applications (Addo et al., 2018). Furthermore, deep learning has achieved superior results in a wide variety of applications such as question answering (Bordes, Chopra, and Weston 2014), natural language understanding (Collobert et al. 2011), particularly topic classification, sentiment analysis, language translation (Jean, Cho, and Memisevic 2014; Sutskever, Vinyals, and Le 2014; Wang 2017) and image classification (Ejbali and Zaied 2018; Said et al. 2016). Deep learning is also known as representation learning, it is a new branch of machine learning. However, deep learning algorithms are rarely applied in the field of bankruptcy prediction. Thus, in this section, we review and discuss related work on the deep learning-based classification method to predict bankruptcy.

For instance, Yeh and Wang (2014) trained Deep Belief Network (DBN) to propose a bankruptcy prediction model. In addition, they considered the stock returns of both bankrupt and non-bankrupt firms as input, and use DBN with the Restricted Boltzmann Machine (RBM) to build the bankruptcy prediction models. This proposed model is implemented and validated using American dataset. The result confirms the efficiency of the proposed model

regarding classifying and predicting bankruptcy firm. Also, it showed that DBN outperforms the support vector machine. In addition, Lee, Jang, and Park (2017) addressed the bankruptcy prediction problem using DBN based on the RBM. The proposed model has two phases; (1) unsupervised learning phase and (2) a fine-tuning phase. The first phase is the pre-training step, applying an RBM with the full training dataset. The second phase is the fine-tunes, which is based on the back-propagation algorithm and a recent dataset which reflect the recent relations between predictors and firm performance. They compared the prediction performance of this proposed method with existing algorithms including support vector regression (SVR), feed-forward neural network (FNN) and deep belief networks (DBN) based on the evaluation criteria: root mean squared error (RMSE). The results showed that the DBN algorithm is an effective tool for predicting bankruptcy and outperformed the other used techniques. Similarly, Ribeiro and Lopes (2011) proposed the DBN for bankruptcy classification problem using a real French dataset. The prediction performance of this proposed technique is compared with the SVM-based prediction model and single RBM. The results confirm that the proposed DBN able to provide an accurate decision-making model and outperform the other used classifiers.

More recently, Chaudhuri and Ghosh (2018) implemented a complex hierarchical deep architecture (HAD) for predicting bankruptcy. However, HDA is composed of Hierarchical Rough Bayesian (HRB) with Fuzzy Rough Tensor Deep Stacking Networks (FRTDSN) models. The performance of this technique is experimented using Korean, American and European datasets and based on the evaluation criteria: misclassification error. Their results showed that FRTDSN-HRB is an effective tool for predicting bankruptcy and outperformed the soft computing and statistical models (Fuzzy-SVM, modified fuzzy-SVM, hazard, rough Bayesian, Bayesian, and mixed logit). In another study, Lanbouri and Achchab (2015) proposed a hybrid technique to predict bankruptcy which is divided into two processing stage: pre-training stage and classification stage. First, the nodes for each layer of the deep network are selected using the Local Receptive Field (LRF). Then, they applied the stacked RBM to form a DBN as pre-training. Second, the SVM technique is used to build the bankruptcy prediction model. In this stage, firms have been classified into two classes, bankrupt and non-bankrupt firm.

This paper contributes to the demonstrations of the applicability of deep learning algorithms in the bankruptcy prediction domain to enhance the prediction accuracy. Meanwhile, we witness the more applicability of the machine learning algorithms such as DT, SVM, and ANN, etc. All of these methods are already popular in both areas of data mining and bankruptcy prediction. However, the deep learning is an advanced machine learning approach applied in the bankruptcy prediction problem. To this end, we implement a deep neural networks algorithm called stacked auto-encoders

(SAE) with the softmax classifier. First, we used the SAE to reduce the dimensionality of the data. Then, we applied the softmax classifier layer to classify firm's as bankrupt or non-bankrupt.

Methodology

Stacked Auto-encoder with Softmax Classifier

An auto-encoder is introduced by Hinton in 1980, and the PDP group to solve the problem of unsupervised learning, by using the input as learning targets (Rumelhart, Hinton, and Williams 1986). In addition, an Auto-encoder is a kind of neural network where the input and the output are the same. It is an unsupervised task which uses a back-propagation algorithm for training. The auto-encoder contains three layers: (1) an input layer, (2) one or more hidden layers (encoding layers), and (3) an output layer (decoding layer). Certain variations on the auto-encoder exist to force the hidden layer to learn better representations of the input features.

Stacked auto-encoders (SAE) are, as the name implies, a stack of single-level auto-encoders. It is a deep learning model. The SAE consists of several layers of sparse auto-encoders in which the outputs of each layer are fed into the input of the successive layer. SAE stacking two auto-encoders in succession and use the greedy layer-wise for training.

In this paper, we constructed a stacked auto-encoders with two hidden layers and a softmax layer. In addition, an auto-encoder is the basic unit of stacked auto-encoders. It is consists of two parts, an encoder (from layer 1 to layer 2 in Figure 1) and a decoder (from layer 2 to layer 3 in Figure 1). The encoder compresses the input and generates the code; the decoder then reconstructs the input using only the encoding of the input. This process can be formulated as (1) and (2):

$$y = s(Wx + b) \quad (1)$$

$$z = s(W^t y + b') \quad (2)$$

where W and W^T correspond to the weight matrices; b and b' are the bias vectors; x is the input layer and y is called latent variable which represents this input; s represents the non-linearity function and in our study is the sigmoid function; however, z represents a prediction of x when the value of y is given.

In a stacked auto-encoders, auto-encoders are stacked, so they transform the output of one hidden layer into an input of its successive auto-encoder. However, given the input vectors, the objective of the auto-encoder is to minimize the difference between the input and the output. The cross-entropy function is used to calculate the reconstruction error, as given in Eq. (3), where x_k and z_k represent the k_{th} element of x and z , respectively.

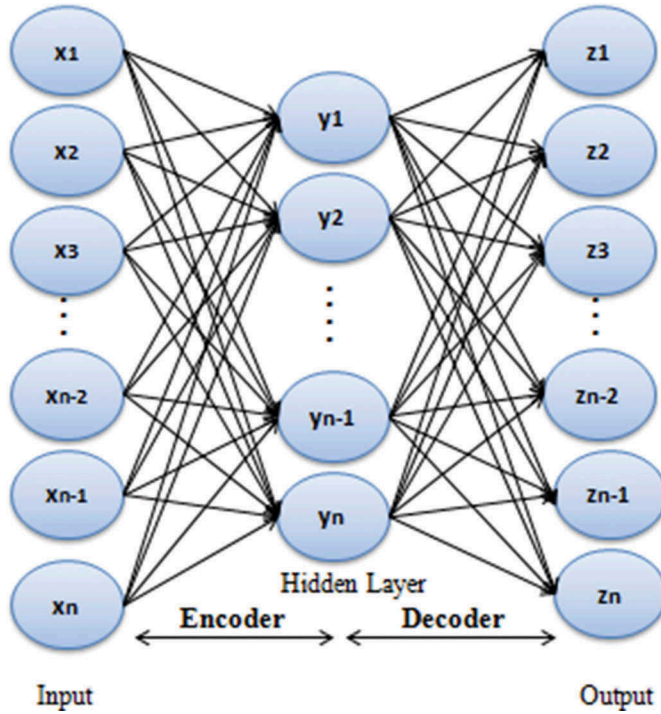


Figure 1. The structure of an auto-encoder.

$$L(x, z) = - \sum_{k=1}^d [x_k \ln z_k + (1 - x_k) \ln(1 - z_k)] \quad (3)$$

In addition, we can reduce the reconstruction error using the Gradient Descent method (Bottou, 2010). The weights in Eq. (1) and Eq. (2) must be updated according to Eqs. (4)-Eq. (5) and Eq. (6), where α represents the learning rate.

$$W = W - \alpha \frac{\partial L(x, z)}{\partial W} \quad (4)$$

$$b = b - \alpha \frac{\partial L(x, z)}{\partial b} \quad (5)$$

$$b' = b' - \alpha \frac{\partial L(x, z)}{\partial b'} \quad (6)$$

After unsupervised pre-trained, we start a second stage known as fine-tuning where we fine-tune the whole network. This stage allows minimizing the prediction error on supervised task by using the back propagation in order to enhance the results by fine-tuning the parameters of all layers that are modified at the same time. The probability, given by our model, that an input vector x (layer 2 in Figure 2) belongs to class i is defined in Eq. (7),

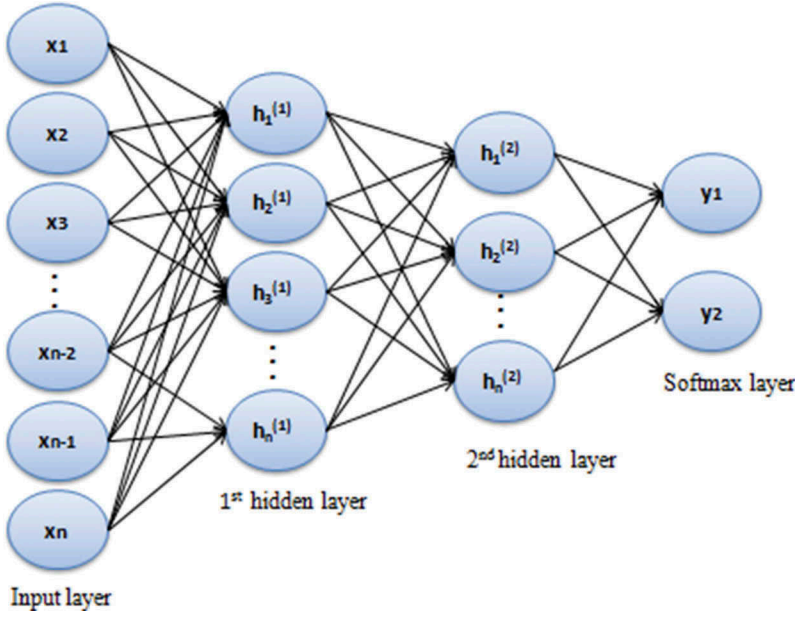


Figure 2. The structure of the used stacked auto-encoders with softmax layer.

where x is the input vector, Y is the predicted class of x ; W denotes the weight matrix and b denotes the bias vector of this layer, W_i and W_j are the i^{th} and j^{th} row of matrix, respectively, b_i and b_j are respectively the i^{th} and j^{th} element of vector b . However, softmax is the non-linearity function used in this work. The predicted label y_{pred} of an input vector x corresponds to the class with the highest probability, as defined in Eq. (8). The calculation of the prediction error sample data set D (Loss (D)) depends on the true labels, as shown in Eq. (9), where y_i is the true label of x_i . In addition, the Gradient Descent method is used to minimize the Loss (D), which is similar to the process of minimizing the reconstruction error as explained above.

$$P(Y = i|x, W, b) = \text{softmax}(Wx + b) = \frac{e^{w_i x + b_i}}{\sum_j e^{w_j x + b_j}} \quad (7)$$

$$y_{pred} = \text{argmax}(P(Y = i|x, W, b)) \quad (8)$$

$$\text{Loss}(D) = - \sum_{i=0}^D \ln(P(Y = y_i|x_i, W, b)) \quad (9)$$

Experimental Setup

In this section, we assess the performance of our used algorithm (SAE+softmax classifier). We first describe the dataset and the evaluation measures used to evaluate the performance. Then, we show the experimental results.

Description of the Experimental Database

The performance of the used algorithms is evaluated using two real-world datasets. The first dataset is selected from the CD-ROM database. It regroups 132 observations divided into binary classes: 66 represents bankrupted firms and 66 firms that did not bankrupt. Each observation presents 24 attributes (financial ratios) using data from the Compustat tapes and Moody's Industrial Manual for the year that was 2 years to the year of bankruptcy. The main characteristics of this dataset have been defined in Table 1.

The second dataset of Polish companies is obtained from the University of California, Irvine (UCI) Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets.html>). Although the companies were analyzed from 2002 to 2013, and still operating companies were evaluated from 2007 to 2013, our prediction model is primarily based around the analysis of the financial rates from the first year of forecasting period and the corresponding class label that indicates bankruptcy status after 5 years. The dataset regroup 7027 observation divided into binary classes: 271 represents bankrupted firms and 6756 firms that did not bankrupt. Each observation presents 64 attributes (financial ratios). The features of the Polish dataset are explained in Table 2.

To enhance the reliability of the estimates and reduce the data dependency, Polish and Darden datasets are randomly separated into two parts (training and testing partitions). In this work, we use 80% of the dataset for training the classifier while the remaining 20% for testing the performance of the model. For polish dataset, we applied the random undersampling to balance class distribution.

Evaluation Criteria

Different evaluation criteria are applied to determine the feasibility and the effectiveness of the used techniques for predicting bankruptcy include

Table 1. Financial ratios of darden dataset.

ID	Description	ID	Description
R1	Cash/Current debt	R13	Income/Sales
R2	Cash/Sales	R14	Income/Total assets
R3	Cash/Total assets	R15	Income/Total debt
R4	Cash/Total debt	R16	Income plus depreciation/Sales
R5	Cash flow from operations/Sales	R17	Income plus depreciation/Total assets
R6	Cash flow from operations/Total	R18	Income plus depreciation/Total assets debt
R7	Cash flow from operations/Total	R19	Sales/Receivables debt
R8	Cost of goods sold/Inventory	R20	Sales/Total assets
R9	Current assets/Current debt	R21	Total assets/Total debt
R10	Current assets/Sales	R22	Working capital from operations/Sales
R11	Current assets/Total assets	R23	Working capital from operations/Total assets
R12	Current Debt/Total debt	R24	Working capital from operations/Total debt

Table 2. Financial ratios of polish dataset.

ID	Description	ID	Description
R1	net profit/total assets	R33	operating expenses/short- term
R2	total liabilities/total assets	R34	operating expenses/total liabilities
R3	working capital/total assets	R35	profit on sales/total assets
R4	current assets/short-term liabilities	R36	total sales/total assets
R5	[(cash+short-term securities +	R37	(current assets- inventories)/long-receivables – short-term liabilities)/(term liabilities (operating expenses-depreciations)] *365
R6	retained earnings/total assets	R38	constant capital/total assets
R7	EBIT/total assets	R39	profit on sales/sales
R8	book value of equity/total liabilities	R40	(current assets – inventory receivables)/short-term liabilities
R9	sales/total assets	R41	total liabilities/((profit on operating activities+depreciation)*(12/365))
R10	equity/total assets	R42	profit on operating activities/sales
R11	(gross profit + extraordinary items +	R43	rotation receivables + inventory financial expenses)/total assets turnover in days
R12	gross profit/short-term liabilities	R44	(receivables*365)/sales
R13	(gross profit + depreciation)/sales	R45	net profit/inventory
R14	(gross profit + interest)/total assets	R46	(current assets – inventory)/short- term liabilities
R15	(total liabilities * 365)/(gross profit +	R47	(inventory * 365)/cost of depreciation) products sold
R16	(gross profit + depreciation)/total	R48	EBITDA (profit on operating activities – depreciation)/total liabilities assets
R17	total assets/total liabilities	R49	EBITDA (profit on operatingactivities – depreciation)/sales
R18	gross profit/total assets	R50	current assets/total liabilities
R19	gross profit/sales	R51	short-term liabilities/total assets
R20	(inventory * 365)/sales	R52	(short-term liabilities * 365)/cost of products sold)
R21	sales (n)/sales (n-1)	R53	equity/fixed assets
R22	profit on operating activities/total assets	R54	constant capital/fixed assets
R23	net profit/sales	R55	working capital
R24	gross profit (in 3 years)/total assets	R56	(sales – cost of products sold)/sales
R25	(equity – share capital)/total assets	R57	(current assets – inventory – short-term liabilities)/(sales – gross profit – depreciation)

(Continued)

Table 2. (Continued).

ID	Description	ID	Description
R26	(net profit + depreciation)/total liabilities	R58	total costs/total sales
R27	profit on operating activities/financial expenses	R59	long-term liabilities/equity
R28	working capital/fixed assets	R60	sales/inventory
R29	logarithm of total assets	R61	sales/receivables
R30	(total liabilities – cash)/sales	R62	(short-term liabilities *365)/sales
R31	(gross profit + interest)/sales	R63	sales/short-term liabilities
R32	(current liabilities * 365)/cost of products sold	R64	sales/fixed assets

Kolmogorov–Smirnov statistic, Gini coefficient, accuracy, area under the ROC curve (AUC), mean squared error and error rate. Among all these standard evaluation criteria, AUC and accuracy are the most commonly used measures to assess the performance of prediction models (Marqués, Garcia, and Sanchez 2012). The definition of these performance measures can be formulated using 2×2 confusion matrix as that explained in Table 3.

In general, the accuracy rate is the most widespread performance criterion used in evaluating the predictive accuracy of classification models. That's way, most bankruptcy prediction studies used the accuracy as a measure for evaluating the performance of algorithms. In addition, it defined as the fraction of correctly classified instances (both bankrupt and non-bankrupt firms) on a particular dataset. The accuracy function is calculated using the following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (10)$$

The most common evaluation criterion in bankruptcy prediction is used to evaluate the performance of the classifier on balanced datasets such as accuracy metric. Thus, it was inappropriate for the imbalanced dataset. To this aim, the area under the ROC curve (AUC) has been considered an adequate criteria for evaluating and comparing bankruptcy prediction model, because it is insensitive to misclassified costs and imbalanced distributions. Furthermore, AUC is widely used for the binary problem; it determines the trade-off between the true positive rate and false positive rate.

$$AUC = \frac{Sensitivity + Specificity}{2} \quad (11)$$

where sensitivity corresponds to the percentage of non-bankruptcy firms that have been predicted correctly.

$$Sensitivity = \frac{TP}{TP + FN} \quad (12)$$

Whereas specificity measures the percentage of bankruptcy firms predicted as bankruptcy.

$$Specificity = \frac{TN}{TN + FP} \quad (13)$$

Table 3. A confusion matrix.

Predicted class (%)	Actual class (%)	
	Non-bankrupt	Bankrupt
Non-Bankrupt	TP	FN
Bankrupt	FP	TN

Empirical Analysis

Our work addresses two main experiments, which are described below. In this section, we explain the design of these two experiments. The purpose of this work is to evaluate the performance of our proposed method regarding its accuracy for predicting bankruptcy of firms.

Experiment 1: SAE with Softmax Classifier Performance

The classification accuracy of SAE technique in either supervised pre-training and fine-tuning stages is affected by various parameters. The main parameters of the SAE are max epoch, the number of hidden layers, learning rate batch size, the number of neurons in the hidden layer, etc. In addition, the experiments were developed and implemented using Matlab 2017b platform. Our final classification result of all used model computes training and testing set on average. The proposed stacked auto-encoders contain two hidden layers and a final softmax classifier layer which capable to classify firms into bankrupt or non-bankrupt.

In this paper, we adjusted two critical parameters to improve the model: the number of hidden layers and the number of units in each layer. In this experiment, we choose 1 to 4 hidden layers in the network. As shown in Figures 3 and 4, the effect of the number of hidden layers and the number of units for each of them on the accuracy of SAE. The experimental result of the Darden dataset demonstrated that the SAE has high accuracy when the number of hidden layers is two (shown in Figure 3). At this point, we use the SAE with two hidden layers then we change the number of neurons (units) per layer to experiment again. The result confirms that we achieve the highest accuracy of 87.9% when the number of neurons in each hidden layer is 16.

For the Polish dataset, the result shows that the SAE has the highest accuracy rate 98% when the number of hidden layers is two and the number of neuron in each hidden layer is 48. Figure 4 shows the effect of the number of hidden layers and the number of hidden units on the overall accuracy of SAE.

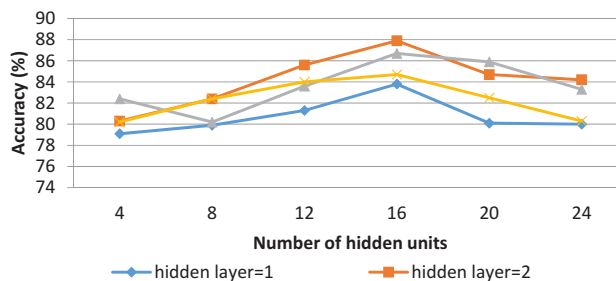


Figure 3. The impact of the number of SAE hidden layers and the number of hidden units on classification accuracy for Darden dataset.

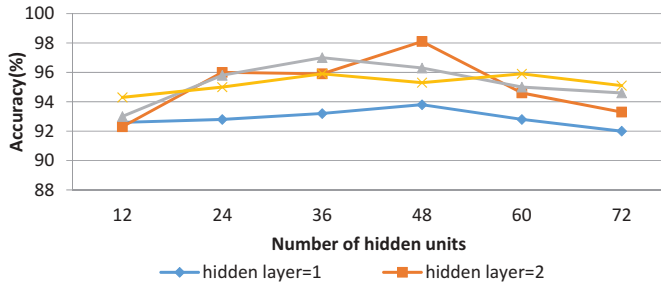


Figure 4. The impact of the number of SAE hidden layers and the number of hidden units on classification accuracy for Polish dataset.

Experiment 2: Comparative Study

At present, several data mining techniques such as neural network, support vector machine, decision tree have been successfully used to predict bankruptcy, and they usually have good prediction accuracy. Therefore, we compare our classifier performance with similar classifiers applied to two the used dataset. These two following tables report accuracy and AUC of some used algorithms, to evaluate the efficiently of our approach for predicting the financial condition of the enterprises. Tables 4 and 5 support the claim that our classifier outperforms these similar works.

Table 4 presents the comparison of the results obtained by SAE+softmax classifier and some existing work taking into account the two evaluation criteria: accuracy and AUC based on small balanced dataset. As shown in this table, the SAE with softmax classifier has the highest accuracy rate 87.9% followed by the method proposed by (Wang, Ma, and Yang 2014) which achieves 86.79% accuracy for Darden testing dataset. However, this work does not take into account the AUC criteria. Therefore, it’s AUC measure comparison is not possible. When comparing

Table 4. Comparison of our approach with similar works for Darden dataset.

	Methods	Accuracy	AUC
Our model (Wang, Ma, and Yang 2014)	SAE+softmax	87.9	87.7
	LRA	73.90	–
	NB*	77.78	–
	DT	75.99	–
	ANN	75.69	–
	SVM	79.99	–
	Bagging	81.22	–
	Boosting	79.40	–
	FS-Boosting	86.79	–
	(Zhang et al. 2017)	RBF	74.2
MLP		68.9	79.1
Bayesnet		72	76.6
J48		62.9	63
Logistic		77.8	80.3
FV-SMO		74.3	82.6

NB: Naive Bayes

Table 5. Comparison of our approach with similar works for Polish dataset.

	Methods	Accuracy	AUC
Our model (Zieba, Tomczak, and Tomczak 2016)	SAE+softmax	0.98	0.961
	LDA	–	0.639
	MLP	–	0.543
	LR	–	0.620
	AB*	–	0.916
	SVM	–	0.502
	RF	–	0.851
	XGB*	–	0.945
	XGBE*	–	0.953
	EXGB*	–	0.959
(Zhang et al. 2017)	Isolation Forest	–	0.93
	One-Class SVM	–	0.92
	Multivariate Gaussian	–	0.89
	NN	–	0.84
	GBDT*	–	0.85

AB: AdaBoost.

XGB: Boosted trees trained with Extreme Gradient Boosting.

XGBE: only the last tree of the ensemble of boosted trees, i.e., f_k , trained with the ensemble of boosted trees with synthetic features.

EXGB: Ensemble of boosted trees trained with the Ensemble of boosted trees with synthetic features.

GBDT: Gradient boosted decision tree.

our classifier with the classifier proposed by (Zhang et al. 2017), considering the AUC (area under the ROC curve space) measure, we can observe that SAE with the softmax classifier generate the best result performance. Similarly, Table 5 presents the same performance indicator comparison for Polish dataset. This dataset is a high dimensional and imbalanced dataset. The existing works about bankruptcy prediction which used the same dataset (Polish dataset) have been focused on feature selection and data reduction considering the AUC performance indicator. So, it's accuracy measure comparison is not possible. For this benchmark dataset, SAE with the softmax classifier provides a good accuracy rate (98%) and achieves the highest AUC compared with other techniques that exist in the literature.

Based on the values in Tables 4 and 5, we notice that SAE with the softmax classifier has the best performance on two bankruptcy datasets in any condition. It provides the highest accuracy and AUC rate in comparison to all reference classifiers that were applied to solve the problem of bankruptcy prediction. Therefore, we conclude that the SAE+softmax classifier is an efficient technique for predicting bankruptcy by correctly classify firm either bankrupt or non-bankrupt. Thus, the proposed SAE technique can be used as a feasible solution to enhance the accuracy in predicting the bankrupt firms.

Conclusion

Bankruptcy prediction is considered as the most important issue in the field of financial research. Therefore, many studied have been conducted to

predict this type of risk. Still, most of this work used machine learning algorithms. However, machine learning models are not perfect when applied in bankruptcy prediction problem. They have failed to extract complex and non-linear patterns from big data. In this context, this paper proposes a new perspective on the bankruptcy prediction problem using deep learning algorithm. We applied the stacked auto-encoders with the softmax classifier to solve the problem of bankruptcy prediction. In the proposed model, a two-layer auto-encoder is used to learn the attributes followed by a softmax classifier layer which provides the probability of each class label. Finally, fine-tuning using back propagation is applied to all the hidden layers to improve the SAE performance. The results clearly indicate that the SAE with the softmax classifier outperforms the reference methods.

Unfortunately, this paper has some limitations which future work can aim at solving. First, we evaluate the prediction accuracy and classification ability of the SAE+softmax classifier with the machine learning algorithms. It would also be meaningful to compare this work with some of the other deep learning algorithms such as the deep belief networks (DBN) and the convolutional neural networks (CNN). Second, in this work, we focused only on increasing the accuracy of bankruptcy prediction model. However, on the field of the financial institutions, transparency and interpretability are highly recommended which guarantee the comprehensibility of the decision by both experts and applicants. Therefore, we plan to investigate a rule-based classification technique capable of classifying and to evaluate easily new firms, in term of future research.

References

- Addo, P., D. Guegan, B. Hassani, P. Addo, D. Guegan, B. Hassani, and C. Risk. 2018. Credit risk analysis using machine and deep learning models. *Documents De Travail Du Centre D'economie De La Sorbonne* 106–12. doi:10.3390/risks6020038.
- Ahn, H., and K. J. Kim. 2011. Corporate credit rating using multiclass classification models with order information. *International Journal of Economics and Management Engineering* 60 (12):95–100.
- Altman, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23 (4):589–609. doi:10.1111/j.1540-6261.1968.tb00843.x.
- Andrés, J. D., M. Landajo, and P. Lorca. 2012. Bankruptcy prediction models based on multi norm analysis: An alternative to accounting ratios. *Knowledge-Based Systems* 30:67–77. doi:10.1016/j.knosys.2011.11.005.
- Barboza, F., H. Kimura, and E. Altman. 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications* 83:405–17. doi:10.1016/j.eswa.2017.04.006.
- Beaver, W. H. 1966. Financial ratio as predictors of failure, empirical research in accounting: selected studies 1966. *Journal of Accounting Research* 4:71–111. doi:10.2307/2490171.
- Bordes, A., S. Chopra, and J. Weston. 2014. Question answering with subgraph embeddings. *Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha*: 615–20.

- Bottou, L. 2010. Large-Scale Machine Learning with Stochastic Gradient Descent. Paper presented at the International Conference on Computational Statistics, Paris, August 22–27.
- Bouallégué, A., S. Hassairi, R. Ejbali, and M. Zaied. 2016. Learning deep wavelet networks for recognition system of arabic words. *International Joint Conference SOCO'16-CISIS'16-ICEUTE'16*: 498–507.
- Callegón, A. M., A. M. Casado, M. A. Fernández, and J. I. Peláez. 2013. A System of Insolvency Prediction for industrial companies using a financial alternative model with neural networks. *International Journal of Computational Intelligence Systems* 6 (1):29–37. <http://www.tandfonline.com/loi/tcis20>.
- Chaudhuri, A., and S. K. Ghosh, 2018. Bankruptcy Prediction through Soft Computing based Deep Learning Technique.
- Chen, M. Y. 2011. Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications* 38 (9):11261–72. doi:10.1016/j.eswa.2011.02.173.
- Collobert, R., and J. Weston. 2008. A unified architecture for natural language processing. *Proceedings of the 25th International Conference on Machine Learning - ICML'08*. 160–67.
- Collobert, R., J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. 2011. Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research* 12:2493–537.
- Constand, R. L., and R. Yazdipour. 2011. Firm failure prediction models: A critique and a review of recent developments, in: *Advances in Entrepreneurial Finance*. Springer-Verlag New York 185–204. doi:10.1007/978-1-4419-7527-0.
- Dahl, G. E., D. Yu, L. Deng, and A. Acero. 2012. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech and Language Processing* 20 (1):30–42. doi:10.1109/TASL.2011.2134090.
- Dey, D. 2017. Growing importance of machine learning in compliance and regulatory reporting. *European Journal of Multidisciplinary Studies* 6:255–58. doi:10.26417/ejms.v6i2.
- Ejbali, R., and M. Zaied. 2018. A dyadic multi-resolution deep convolutional neural wavelet network for image classification. *Multimedia Tools and Applications* 77 (5):6149–63. doi:10.1007/s11042-017-4523-2.
- Erdogan, B. E. 2013. Prediction of bankruptcy using support vector machines: An application to bank bankruptcy. *Journal of Statistical Computation and Simulation* 83 (8):1543–55. doi:10.1080/00949655.2012.666550.
- Fan, S., G. Liu, and Z. Chen. 2017. Anomaly detection methods for bankruptcy prediction. International Conference on Systems and Informatics.
- Gepp, A., and K. Kumar. 2015. predicting financial distress: A comparison of survival analysis and decision tree techniques. *Procedia Computer Science* 54:396–404. doi:10.1016/j.procs.2015.06.046.
- Gepp, A., K. Kumar, and S. Bhattacharya. 2009. Business failure prediction using decision trees. *Journal of Forecasting* 29:536–55. doi:10.1002/for.1153.
- Hinton, G., L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, and B. Kingsbury. 2012. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine* 29 (6):82–97. doi:10.1109/MSP.2012.2205597.
- Jang, M., S. Seo, and P. Kang. 2018. Recurrent Neural Network-Based Semantic Variational Autoencoder for Sequence-to-Sequence Learning.
- Jardin, P. D. 2010. Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing* 73(10-12):2047–60. doi:10.1016/j.neucom.2009.11.034.

- Jean, S., K. Cho, and R. Memisevic. 2014. On using very large target vocabulary for neural machine translation. In *Computer Science*, 1–10.
- Karels, G. V., and A. J. Prakash. 1987. Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance & Accounting* 14 (4):573–93. doi:10.1111/jbfa.1987.14.issue-4.
- Kasgari, A. A., M. Divsalar, M. R. Javid, and S. J. Ebrahimian. 2012. Prediction of bankruptcy Iranian corporations through artificial neural network and Probit-based analyses. *Neural Computing and Applications* 23 (3–4):927–36. doi:10.1007/s00521-012-1017-z.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. Image Net Classification with Deep Convolutional Neural Networks. In *Advances In Neural Information Processing Systems*, 1–9.
- Labouri, Z., and S. Achchab. 2015. A hybrid Deep belief network approach for Financial distress prediction. *International Conference on Intelligent Systems: Theories and Applications (SITA)*: 1–6.
- Lee, J., D. Jang, and S. Park. 2017. Deep learning-based corporate performance prediction model considering technical capability. *Sustainability* 9 (6):899. doi:10.3390/su9060899.
- Marqués, A. I., V. Garcia, and J. S. Sanchez. 2012. Two-level classifier ensembles for credit risk assessment. *Expert Systems with Applications* 39:10916–22. doi:10.1016/j.eswa.2012.03.033.
- Mbarki, N. E. H., R. Ejbali, and M. Zaied. 2017. Recognition of human activities in smart homes using stacked autoencoders. *The Tenth International Conference on Advances in Computer-Human Interactions, ACHI*: 176–80.
- McKee, T. E., and T. Lensberg. 2002. Genetic programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research* 138 (2):436–51. doi:10.1016/S0377-2217(01)00130-8.
- Min, J. H., and C. Jeong. 2009. A binary classification method for bankruptcy prediction. *Expert Systems with Applications* 36:5256–63. doi:10.1016/j.eswa.2008.06.073.
- Minar, M. R., and J. Naher. 2018. Recent Advances in Deep Learning: An Overview, (February): 0–31.
- Najafabadi, M. M., F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic. 2015. Deep learning applications and challenges in big data analytics. *Journal of Big Data* 2 (1):1–21. doi:10.1186/s40537-014-0007-7.
- Nguyen, H. G. 2005. Using Neural Network in Predicting Corporate Failure. *Journal of Social Sciences* (15493652). 1 (4):199–202. doi:10.3844/jssp.2005.199.202.
- Ocal, N., M. K. Ercan, and E. Kadioğlu. 2015. Predicting financial failure using decision tree algorithms: An empirical test on the manufacturing industry at borsa istanbul. *International Journal of Economics and Finance* 7:7. doi:10.5539/ijef.v7n7p189.
- Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18 (1):109–31. <http://www.jstor.org/stable/2490395>.
- Ribeiro, B., and N. Lopes. 2011. Deep belief networks for financial prediction. *International Conference on Neural Information Processing*; Springer: Berlin/Heidelberg, Germany, 766–73.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. Learning internal representations by error propagation (No. ICS-8506). *California Univ San Diego La Jolla Inst For Cognitive Science* 1:318–62.
- Said, S., O. Jemai, S. Hassairi, R. Ejbali, M. Zaied, and C. Ben Amar. 2016. Deep wavelet network for image classification. *IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*: 922–27.
- Sanchis, A., M. J. Segovia, J. A. Gil, A. Heras, and J. L. Vilar. 2007. Rough Sets and the role of the monetary policy in financial stability (macroeconomic problem) and the prediction of

- insolvency in insurance sector (microeconomic problem). *European Journal of Operational Research* 181 (3):1554–73. doi:10.1016/j.ejor.2006.01.045.
- Santoso, N., and W. Wibowo. 2018. Financial distress prediction using linear discriminant analysis and support vector machine. *International Conference on Science (ICOS)*. doi:10.1088/1742-6596/979/1/012089
- Shin, K. S., and Y. J. Lee. 2002. A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications* 23 (3):321–28.
- Srinivas, S., R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. S. Kruthiventi, and R. V. Babu. 2016. A taxonomy of deep convolutional neural nets for computer vision. *Frontiers in Robotics and AI*. doi:10.3389/frobt.2015.00036.
- Sutskever, I., O. Vinyals, and Q. V. Le. 2014. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems (NIPS)* 3104–12.
- Van Gestel, T., B. Baesens, J. Suykens, M. Espinoza, D. E. Baestaens, J. Vanthienen, and B. D. Moor. 2003. Bankruptcy prediction with least squares support vector machine classifiers. *International Conference on Computational Intelligence for Financial Engineering*. Proceedings. 1–8.
- Wang, G., J. Ma, and S. Yang. 2014. An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems With Applications* 41:2353–61. doi:10.1016/j.eswa.2013.09.033.
- Wang, N. 2017. Bankruptcy Prediction Using Machine Learning. *Journal of Mathematical Finance* 07 (04):908–18. doi:10.4236/jmf.2017.74049.
- Xie, C., C. Luo, and X. Yu. 2011. Financial distress prediction based on SVM and MDA methods: The case of Chinese listed companies. *Quality & Quantity* 45 (3):671–86. doi:10.1007/s11135-010-9376-y.
- Yeh, C. C., D. J. Chi, and Y. R. Lin. 2014. Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences* 254:98–110. doi:10.1016/j.ins.2013.07.011.
- Yeh, S., and C. Wang. 2014. Corporate Default Prediction via Deep Learning. In *International Institute of Forecasting*.
- Yu, L., Z. Yang, and L. Tang. 2015. A novel multistage deep belief network based extreme learning machine ensemble learning paradigm for credit risk assessment. *Flexible Services and Manufacturing*. doi:10.1007/s10696-015-9226-2.
- Zhang, Q., J. Wang, A. Lu, S. Wang, and J. Ma. 2017. An improved SMO algorithm for financial credit risk assessment—evidence from China’s banking. *Neurocomputing*. doi:10.1016/j.neucom.2017.07.002.
- Zieba, M., S. K. Tomczak, and J. K. Tomczak. 2016. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems With Applications* 58:93–101. doi:10.1016/j.eswa.2016.04.001.
- Zmijewski, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* 22:59–82. <http://www.jstor.org/stable/2490859>.