Information Systems Frontiers
https://doi.org/10.1007/s10796-020-10031-6

Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE

Salima Smiti[1] · Makram Soui[2]

Abstract
Imbalanced classification on bankruptcy prediction is considered as one of the most important topics in financial institutions. In this context, various statistical and artificial intelligence methods have been proposed. Recently, deep learning algorithms are experiencing a resurgence of interest, and are widely used to build a prediction and classification models. To this end, we propose a novel deep learning-based approach called BSM-SAES. This approach combines Borderline Synthetic Minority oversampling technique (BSM) and Stacked AutoEncoder (SAE) based on the Softmax classifier. The aim is to develop an accurate and reliable bankruptcy prediction model which includes the features extraction process. To assess the classification performance of our proposed model, k- nearest neighbor, decision tree, support vector machine, and artificial neural network, C5.0 that are machine learning methods, are applied. We evaluate our proposed approach on the Polish imbalanced datasets. The obtained results confirm the efficiency of our proposed model compared to other machine learning models regarding predicting and classifying the financial status of a firm.

Keywords Bankruptcy prediction · Deep learning · Stacked autoencoder · Borderline SMOTE · Imbalanced dataset

1 Introduction

In the current economic situation, bankruptcy prediction has become an important task that provides timely alerts to decision makers of the company. The ultimate aim of this prediction is to confirm economic stability and to enhance the efficiency of the commercial credit allocation. In addition, the number of company bankruptcies is crucial to a country's economy, and it can be considered as an indicator of economic development (Van Gestel et al. n.d.).

Furthermore, the economic, social costs and high number of companies that faced bankruptcies have let experts to better understand the bankruptcy risk (McKee and Lensberg 2002). Hence, the risk of bankruptcy can influence the country's economy. Thus, the bankruptcy prediction has gained an increasingly important role in the economy and society since it has an essential impact on the profitability of financial institutions. The aim is to evaluate a firm's financial situation and its future perspectives (Constand and Yazdipour 2011). Indeed, it has become more significant since the advent of the Basel II requirement. This consent highlights its importance for an accurate decision-making model. Besides, Basel II permits banks to assess the company's risk of bankruptcy based on its its own internal models, the probability of default, and the capital needed to face that loss (De Andrés et al. 2012).

Currently, majority of the existing works considered the bankruptcy prediction as binary classification problems. (Yeh et al. 2014a; Barboza et al. 2017; Min and Jeong 2009) differentiate two classes of the companies: bankrupt and non-bankrupt firms. In this context, many classification methods are proposed to assess the risk of bankruptcy. These methods can be classified into two main categories: (1) statistical methods and (2) artificial intelligence methods. The statistical methods have been widely used to built models for estimating bankruptcy. It was starting with Beaver (Beaver 1966) and Altman (Altman 1968) who applied Multivariate Discriminant Analysis (MDA), linear probit (Zmijewski 1984), and logit (Ohlson 1980). Nevertheless, these methods have certain boundaries inherent from the statistical techniques that reduce the application of this algorithm in the

✉ Salima Smiti
salima.smiti92@gmail.com

Makram Soui
m.soui@seu.edu.sa

[1] National School of Computer Science, University of Manouba, Manouba, Tunisia

[2] College of Computing and Informatics, Saudi Electronic University, Arabie Saoudite, Riyadh, Saudi Arabia

Published online: 03 August 2020

bankruptcy prediction. The main boundary lies in the assumption of linear separability, independence of the predictive variables, and multivariate normality (Ohlson 1980; Karels and Prakash 1987). But, these assumptions are infringed by multiple financial ratios. In addition, the traditional methods used a fixed function, and a set of assumptions of the statistical method make it hard to develop complex financial models (Yu et al. 2015). To overcome this drawback, different artificial intelligence methods have been proposed for designing bankruptcy prediction models. The most popular artificial intelligence algorithms used in bankruptcy prediction are: Support Vector Machine (SVM) (Barboza et al. 2017; Erdogan 2013; Xie et al. 2011), Artificial Neural Network (ANN) (Jardin 2010; Callejón et al. 2013), rough sets (McKee and Lensberg 2002; Sanchis et al. 2007), Decision Tree (DT) (Ocal et al. 2015; Gepp et al. 2010; Gepp and Kumar 2015) and Genetic Algorithm (GA) (Shin and Lee 2002; Smiti et al. 2018). The obtained results prove that these algorithms outperform the traditional statistical techniques (Ocal et al. 2015; Minar and Naher 2018; Chen 2011). But, these specified algorithms have some boundaries, such as the time-consuming process (it is very expensive) and local maximum or minimum problems (Yu et al. 2015).

The artificial intelligence methodincludes many categories like machine learning and Deep Learning (DL). Various artificial intelligence methods have been proposed for building bankruptcy prediction models. In the finance context, machine learning as well as an important branch of Artificial intelligence, have been successfully used in the area of finance and accounting (Barboza et al. 2017). More recently, researchers focused on deep learning using artificial neural networks. (Nguyen 2005; Lu et al. 2017). It has proved its efficiency to tackle many problematic in recent years (Dey 2017). Furthermore, it becomes popular and more efficient than traditional methods and machine learning methods in many fields such as language processing (Sutskever et al. 2014a; Collobert and Weston 2008; Lu and Li 2017), speech recognition (Dahl et al. 2012; Hinton et al. 2012; Bouallégue et al. 2016), computer vision (Srinivas et al. 2016; Krizhevsky and Sutskever 2012). Nevertheless, there were few existing works about bankruptcy evaluation using deep learning algorithms. In this work, we study the effectiveness of the stacked AutoEncoder (SAE) combined with a softmax classifier to predict the bankruptcy of financial companies.

The SAE has been applied in many fields such as image processing, speech recognition, machine translation, paraphrase detection (Jang et al. 2018). and human activity recognition (Mbarki et al. 2017; Smiti et al. 2019). But, there are few propositions concerning the use of the SAE in the economic or the financial area. The idea of our work is to study its application in the economic field by using it with a dataset in the company bankruptcy prediction.

Generally, the classification of imbalanced datasets has significant attention in different fields thanks to its wide real applications (Spelmen and Porkodi 2018; Johnson and Khoshgoftaar 2019; López et al. 2013). However, the imbalanced dataset problem has been neglected by most of the existing works about bankruptcy prediction. Many existing works used datasets with paired samples for training and testing bankruptcy prediction models, having the same bankrupt and non-bankrupt companies (Aliaj et al. 2018; Zhou et al. 2012; Ashoori and Mohammadi 2011). Nevertheless, the real-world bankruptcy dataset includes an imbalanced proportion of 10% bankrupt companies and 90% non-bankrupt companies (Shrivastava and Kumar 2018) (Zhou 2013). A predictive model created based on imbalanced dataset may ignore the bankrupt companies (the minority class) and classify the companies as non-bankrupt (majority class). This problem can has an impact on the efficiency of the created models. But, few studies focused on the imbalanced problem in bankruptcy prediction. The challenge of these studies is to derive high accuracy by correctly classifying the minority samples without any impact on majority instances. Hence, many methods have been suggested to address the imbalanced issue. These methods can be divided into two categories: undersampling and oversampling techniques (Zhou 2013). Undersampling aims to remove the majority class examples from the dataset while oversampling consists of generating new samples in the minority class instances until the majority class number become around equal to the minority class number. Over the last few years, oversampling techniques have become the most popular methods for handling imbalanced data (Yusof et al. 2017; Last et al. 2018). In fact, existing studies have prove that oversampling is more effective than undersampling methods for the bankruptcy prediction problem since undersampling can delete pertinent instances from the dataset (Nekooeimehr and Lai-Yuen 2016; Le et al. 2018) and hence the insufficient number of minority class samples may lead to a negative influence on the induction process. On the other hand, oversampling aims to maintain the original content of the dataset without losing any data (Ertekin 2013). For this reason, the oversampling method is more robust and performs better than the undersampling method (Garcı et al. 2013) (Zhou 2013). However, the oversampling is carried out by generating new synthetic instances from existing minority ones that can cause the variance of bankrupt financial ratios. Whereas this variance can improve the performance of the model. This leads to alleviating the classifier to be biased toward the most represented class. Many techniques have been applied to deal overgeneralization, the most widely known is the Synthetic Minority Oversampling Technique (SMOTE) (Skryjomski and Krawczyk 2017). The basic idea of this technique is to create new synthetic samples for minority class to achieve a balanced dataset distribution. Various variants of SMOTE exist, the most popular one is Borderline SMOTE (BSM) (Wang et al. 2015). Unlike SMOTE, BSM does not generate synthetic samples for noise instance, but only creates synthetic samples near the boundaries. Thus,

the BSM can enhance the classifier performances in the classification task.

There are several measurements of credit risk: the probability of default (PD), the exposure at default (EAD), the expected loss given default (LGD), and scoring model. In this work, we propose to build an accurate classification model based on stacked autoencoders + softmax as a deep learning technique in order to classify companies. To the best of our knowledge, this is the first attempt in financial institutions to adopt the borderline oversampling method in combination with stacked autoencoder and softmax classifiers to solve problems of the imbalanced bankruptcy dataset. Therefore, our study aims to simplify the bankruptcy decision-making process and reduce much as possible the time of classification. Our proposed model includes three steps: Borderline SMOTE oversampling, feature extraction and classification. The Borderline SMOTE oversampling technique is applied to obtain a balanced dataset. Then, the stacked auto-encoder is applied to reduce the dimensionality of the features to enhance the model accuracy. Finally, we used the softmax layer to classify the instances. In fact, companies are classified into two classes: bankrupt or non-bankrupt. The obtain results prove the efficiency of the proposed algorithm compared with machine learning algorithms and similar existing works using other classifiers based on the Area under the ROC curve (AUC) performance indicator. Our proposed model has a significant improvement in the classification accuracy compared to conventional methods which get more than 95% in terms of AUC. However, the experimental results confirm that the predictive model achieved worse time performance compared with other model in practice.

The rest of this paper is structured as follows: Section 2 reviews the related works 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 discusses the results of the classification performance of our approach. Finally, we conclude the paper with some pointers towards future work in section 5.

2 Related Work

Machine learning and deep learning algorithms have been effectively applied to address the bankruptcy prediction problem. These algorithms are considered artificial intelligence models. In the following part we will explain in depth how these algorithms have used in bankruptcy classification problem.

2.1 Machine Learning Algorithms

Several machine learning algorithms have been successfully applied to build accurate bankruptcy prediction models such as DT, NN, KNN, and SVM, etc.

Ocal et al. (Ocal et al. 2015) proposed predicting financial failure based on Chi-square Automatic Interaction Detector (CHAID) and C5.0 decision tree algorithms. The obtained results prove that the proposed models have acceptable prediction accuracy. The C5.0 algorithm is better than the CHIAD algorithm for evaluating unhealthy companies, but it has a lower accuracy than the CHAID decision tree algorithm. Furthermore, Fan et al. (2018) suggested and evaluated the one-class SVM, Isolation Forest, and multivariate Gaussian distribution algorithms to assess bankruptcy for a real-world dataset of Polish firms. For the AUC metric, the Isolation Forest gives the best results compared to one-class SVM, multivariate Gaussian distribution, and other classifiers. The result prove that the Isolation Forest can effectively overcome the impact of imbalanced learning problems.

Even though, the decision tree algorithms become popular for the bankruptcy prediction since its capability to derive interpretable rules and ability of handling continuous and categorical features (Daniel 2014; Nor et al. 2020). DTs are prone to errors in the bankruptcy prediction problem using on multiclass. Moreover, the process of rising and pruning a DT makes it computationally time-consuming to train (Sprengers 2005). In this context, ANN and SVM are used thanks due to its high capability to circumvent the complex problem (Jack and Nandi 2002).

For example, Jardin (2010) used the Neural Network (NN) to build a bankruptcy prediction model based on the French bankruptcy dataset. The obtained results show the efficiency of the proposed model for this problematic. In addition, the result confirms that the NN using a set of features selected based on suitable selection methods provides acceptable results than existing methods used already in the financial literature. In this way, Kasgari et al. (2012) proposed modeling bankruptcy prediction based on a multilayer perceptron (MLP). To this end, it used four financial ratios: quick assets to total assets, ratios of sales to current assets ratio, total liability to total assets operational income to sales, and ratios of sales to current assets ratio. The results confirm that the predictive MLP model reached the best results compared to other algorithms (linear regression, gene expression logit, and programming) using Iranian dataset.

Despite of the high accuracy of the NN compared to other classification algorithms, Ahn et al. (Ahn and Kim 2011) identified that there are a set of difficulties in their use that are coming for many parameters to be set by heuristics and hence the model may undergo an overfitting. This phenomenon may lead to the poor performance of the model. Therefore, many exiting works used SVM classifier to build an accurate bankruptcy prediction model, since it can improve the performance of a model by maximizing the margin to prevent overfitting (Erdogan 2013; Xie et al. 2011; Santoso and Wibowo 2018).

Santoso et al. (Santoso and Wibowo 2018) used the Linear Discriminant Analysis (LDA) and SVM with a variable selection technique to predict the bankruptcy risk of financial companies in Indonesia. The obtained results provedthat the hybrid stepwise-SVM model derives the high accuracy metric compared to other exiting classifier. They underlined that SVM algorithm is a powerful classifier to create robust decision support systems for firm loans. In a previous study, Zhang et al. (2018) suggested an improved Sequential Minimal Optimization (SMO) using on Four-Variable named FV-SMO. The main idea is to choose four attributes into the working set in each iteration. This model is proposed and evaluated using China, German, and Darden credit datasets. The experimental results prove that the FV-SMO algorithm attained the highest accuracy for predicting bankruptcy risk compared with the five popular classification algorithms in credit risk assessment such as Logistic Regression, Radial Basic Function, Baysenet, MLP, and DT.

Nevertheless, the major limitation of SVM and ANN is their difficulty to handle high dimensional data (Verleysen et al. 2003; Ghaddar and Naoum-sawaya 2017). This hurdle makes it less appropriate for estimating bankruptcy in large datasets. In the machine learning algorithms, the feature extraction is accomplished independently of the training phase (Ghaddar and Naoum-sawaya 2017; Nilsson et al. 2006).

Recently, many ensemble techniques have been used to study firm bankruptcy risk. Zieba et al. (2016) suggested a new technique called synthetic features which combines the econometric measures and the arithmetical operation for financial distress prediction based on eXtreme Gradient Boosting (XGB). The aim is to enhance the overall performance of the model which is built using a real-life dataset of Polish companies. It confirms that boosting has successfully used for predicting firm bankruptcy compared to other exiting techniques. Similarly, Wang et al. (2014) proposed an improved Boosting method known as FS-Boosting using feature selection to assess bankruptcy. This study emphasis the high performance of FS-Boosting for bankruptcy prediction for two datasets. This algorithm is effectively used to overcome the issue of financial distress based on balanced datasets.

The traditional classifiers such as logistic regression, SVM, ANN, and decision tree are appropriate for balanced training sets (Zhang et al. 2019). These methods often derive suboptimal classification results, i.e. a good coverage of the majority instances, while the minority instances are distorted (López et al. 2013). Moreover, SVM and NN can be sensitive to missing values and hard to train for large data (Verleysen et al. 2003). Besides, these methods have failed to obtain non-linear and complex patterns from a huge heterogeneous data (Najafabadi et al. 2015). However, DL allows using linear models for a big quantity of data analysis tasks by choosing such attributes, such as prediction and classification; which is an interesting step when designing models to transact with a large amount of data (Najafabadi et al. 2015). In this work, we study the use of DL algorithms like stacked autoencoders with softmax classifiers to enhance the accuracy of the bankruptcy prediction model.

2.2 Deep Learning Algorithms

Over the last few years, deep learning has attracted research attention, by outperforming machine learning algorithms such as kernel machines, in multiples fields (Addo et al. 2018) such as question answering (Bordes et al. 2014), natural language understanding (Collobert et al. 2011), particularly topic classification, sentiment analysis, language translation (Wang 2017; Jean et al. 2014; Sutskever et al. 2014b) and image classification (Ejbali and Zaied 2018; Said et al. 2016; Serikawa and Lu 2014). Deep learning is also called representation learning, it is a new variant of machine learning (Mai et al. 2019). But, it is rarely used in the domain of bankruptcy classification. Hence in this work, we study and argue existing works about the deep learning algorithms to assess bankruptcy.

Yeh et al. (2014b) proposed a Deep Belief Network (DBN) to tackle the imbalanced dataset problem for bankruptcy classification. It takes the stock returns of both bankrupt and non-bankrupt firms as input, and applied DBN with the Restricted Boltzmann Machine (RBM) to create the bankruptcy prediction model using the American dataset. The obtained results show the efficiency of the built model in predicting bankruptcy firm. Furthermore, it proves that DBN outperforms the SVM. Lee et al. (2017) tackled the bankruptcy prediction problem based on DBN which use RBM. The built model includes two phases: (1) unsupervised learning phase and (2) a fine-tuning phase. The first one is the pre-training, using an RBM with the full training dataset. The second one is the fine-tunes, which used the back-propagation algorithm and a recent dataset which reflects the recent relations between predictors and company performance. The classification performance is evaluated with existing algorithms such as Deep Belief Networks (DBN), Feed- Forward Neural Network (FNN) and Support Vector Regression (SVR), in term Root Mean Squared Error (RMSE). The experimental results confirm that the DBN algorithm is an effective for predicting bankruptcy and outperforms other exiting used algorithms. Ribeiro et al. (Ribeiro and Lopes 2011) also suggested the DBN for bankruptcy classification problem based on real French dataset. The obtained results prove that the proposed model is capable to derive an accurate model and outperforms the other applied classifiers.

More recently, Chaudhuri et al. (Chaudhuri and Ghosh 2018) proposed a Hierarchical Deep Architecture (HDA) for estimating bankruptcy. It is based on Hierarchical Rough Bayesian (HRB) and Fuzzy Rough Tensor Deep Stacking

Networks (FRTDSN) models. This algorithm is evaluated using Korean, American and European datasets and using the misclassification error. The obtained results prove that FRTDSN-HRB is an effective model for estimating bankruptcy and outperformed the soft computing and statistical models (Fuzzy- SVM, Modified fuzzy-SVM, hazard, rough Bayesian, Bayesian, and mixed logit). Lanbouri et al. (2015) suggested a hybrid method to assess bankruptcy which is includes two steps: pre-training step and classification step. First, the nodes for each layer of the deep network are chosen using the Local Receptive Field (LRF). Then, they used the stacked RBM to create a DBN as pre-training. After that, the SVM algorithm is applied to propose a bankruptcy classification model to classify companies as bankrupt or non-bankrupt firm.

This work studies the use of deep learning algorithms in the bankruptcy classification field to improve prediction accuracy. In the same time, we observe the more application of the machine learning algorithms such as DT, SVM, KNN, and ANN, etc. These algorithms are previously used in both fields of data mining and bankruptcy prediction. Nevertheless, deep learning is an advanced machine learning approach that can automatically generate useful features. In this context, we propose using stacked autoencoders with the softmax classifier for bankruptcy classification. First, we apply the oversampling methods to balance the dataset. Second, we used the SAE to reduce the dimensionality of the attributes. Then, we exploit the softmax classifier layer to classify companies as bankrupt or non-bankrupt.

3 Methods

3.1 Borderline-SMOTE

To overcome the imbalanced dataset problem for bankruptcy classification, there are two sampling methods: (1) undersampling reducesthe number of minority class examples to obtain a balanced class distribution. (2) oversampling increases the minority class examples by creating new synthetic samples (Zhou 2013). In the bankruptcy classification, the oversampling techniques ones performed better than the undersampling, there are a probability of removing interesting examples (Nekooeimehr and Lai-Yuen 2016). The SMOTE suggested by Chawla et al (2002) is one of the widely used oversampling method to address imbalanced dataset. This method creates new synthetic instances from the minority class by combining a certain instances with k similar minority class examples multiplied by Gaussian random distances until both classes become approximately equal. Han et al (2005) suggested Borderline-SMOTE (BSM) as an improved version of SMOTE algorithm which aims to produce synthetic samples by interpolating the k nearest neighbors of the minority instances closer to the border (Wang et al. 2015). SMOTE creates new instance for each minority sample while Borderline-SMOTE only extracts synthetic instances for the minority examples near the border of two classes because these instances are more important for classification (Toribio et al. 2012). Since deep learning algorithms require a large amount of training data, which is balanced by adding instances to the minority class (oversampling) rather than remove instances from the majority class (undersampling). However, oversampling can cause overfitting which occurs when adding the replicas of the existing minority to the main dataset. To address this problem, we suggest to use Borderline SMOTE in order to synthesize the minority class instances by a linear interpolation method.

In Borderline-SMOTE, minority class instances that are likely misclassified will get more training. First, it finds out the borderline minority instances and then uses these instances to create synthetic instances with their selected k nearest neighbors. The Borderline-SMOTE algorithm is adapted for bankruptcy prediction.

Once the new synthetic instances are generated by BSM, the size of the minority class is significantly increased. Suppose that the number of instances in minority class of the original data set is num and the number of instances in borderline instances set D is dnum, then M' new synthetic minority examples are generated attains (num + s × dnum) in the end. For the values of s and m are fixed by the size of instances that we need to synthesize. The value of k is set to 5. The following pseudo-code presented the adapted Borderline-SMOTE:

Algorithm : Borderline SMOTE algorithm

Input: T-Training set
M-Minority examples set
r,k-Number of nearest neighbors
s-Number of synthetic examples that account for the number of original examples in the given class

Output: Synthetic minority samples set: M'

1. D = ϕ //D is a set containing borderline samples
2. for all m_i in M do
3. $N_{mi} \leftarrow$ r nearest neighbors of m_i in T
4. n ← the number of samples in N_{mi} and not in M
5. if $r/2 \le n < r$ then //m_i is a borderline sample
6. add m_i to D
7. end if
8. end for
9. M' = ϕ //M' is a set containing synthetic samples
10. for all d_i in D do
11. $N_{di} \leftarrow$ k nearest neighbors of d_i in M
12. for i = 1 to s do
13. m ← choose a random sample from N_{di}
14. $d_i' \leftarrow d_i + p * (d_i - m)$ //p is a random number in(0, 1), d_i' is a synthetic sample
15. add d_i' to M'
16. end for
17. end for
18. M' = M' ∪ M //M' is the union of minority samples and synthetic samples
19. return M'

3.2 Classifiers

In this part, we detail the machine learning and deep learning classifiers applied in this work.

3.2.1 Support Vector Machine

It is proposed by Cortes and Vapnik (1995). It's widely used for bankruptcy classification. The aim is to search the optimal separating hyperplane which maximizes the margin of separation between the support-vector of the two classes. The support vector is defined by boundary cases selected from each class. The set of support vector machine be represented {<xi , yi>|i =1,2,..,l}, a support vector machine takes the following general form:

$$f(x) = sign\left(\sum_{i=1}^{l} y_i \alpha_i k(x, x_i) + \beta \right) \quad (1)$$

Where: k denotes a kernel function and is applied to performed non-linear transformations. The support vectors and the parameters αi, moreover β are established by solving the constrained quadratic optimization problem.

3.2.2 Artificial Neural Network

It is a flexible non-linear modeling algorithm (Akkoç 2012). It's based on input variables, also called explanatory variables, which communicates to one or more hidden layers, and result in response to the output variables. The basic idea behind the ANN is to simulate the structure of the human brain, in which a many neurons communicate by sending signals between them (Ripley 2008). The output of the hidden neuron is calculated by affected the weighted inputs and its intercept term bi as follows:

$$h_i = f^{(1)}\left(b_i^{(1)} + \sum_{j=1}^{n} W_{ij} x_i \right) \quad (2)$$

Where xi is the input data and ij denotes the weight connecting input neuron j to hidden neuron i. In the bankruptcy prediction, the input layer consists of n characteristics (financial ratios), xi, i=1,...,n; and a weight W affected to each characteristic. In a similar manner, the output of the output neurons is calculated:

$$z_i = f^{(2)}\left(b_i^{(2)} + \sum_{j=1}^{n_k} V_{ij} h_j \right) \quad (3)$$

With nk is the number of hidden neurons and Vij denotes the weight connecting hidden neuron i to the output neuron j. At the start, the networks are initialized with random weights. After that, iteratively the values of weights are adjusted to reduce the loss function. In addition, ANN has been criticized for its black box nature and for its long training process in developing the optimal model (Akkoç 2012).

3.2.3 Decision Tree

It is a decision support tool, which has a flowchart like a tree structure. In the tree structure, leaves indicate class labels while the branches represent conjunctions of input features that lead to those classifications (Sharma and Kumar 2016). In addition, there are various variations among the decision tree algorithm such as C4.5, J48, random forest and random tree, etc. The decision tree methods are based on maximizing the performance measure chosen as related to the target variable. It is a white-box model, which makes it easy to understand its output. In contrast to black-box patterns expressions whose cannot adequately reveal information hidden in the data and the bankruptcy prediction remains difficult. Moreover, the DT has the ability to generalizing the results and in handling a huge number of variables.

During the learning process, the algorithm builds a decision tree starting from a set of training datasets. To build a tree it divides the dataset recursively into several subsets. In order to find the best split point, it employs information gain as an attribute selection measure. Hence, the term attribute with the largest information gain is selected as the test attribute for the current node. To measure uncertainty entropy is used. Formally the entropy of a state S is defined as:

$$H(S) = -\sum_{i=1}^{k} p(i) log p(i) \quad (4)$$

where K is the number of branches and p the probability of each branch (Jaynes 1957).

3.2.4 K Nearest Neighbors (kNN)

The K Nearest Neighbors (KNN) is a non-parametric machine learning algorithm used for classification and regression. It's successfully applied to the large-scale dataset, among all the supervised machine learning algorithms. The KNN algorithm, classify data by a majority vote of its neighbors, with the current point being allocated to the most common class among its k closest neighbors (Dudani 1976). Chebyshev, Minkowski, Manhattan, and Euclidean parameter used in KNN to calculate the distance between new samples and existing samples in the dataset (Lee et al. 2012). In this work, the classification was performed using k=5 and euclidean metric for distance measure.

3.2.5 AutoEncoders

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 et al. 1986). In addition, an Autoencoder 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 autoencoder 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 autoencoder exist to force the hidden layer to learn better representations of the input features.

The main aim of the autoencoder is to reduce the dimensionality of the input data by removing the noise in the data. In addition, an autoencoder is the basic unit of stacked autoencoders. It contains 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 (5) and (6):

$$y = s(Wx + b) \tag{5}$$

$$z = s(W^t y + b') \tag{6}$$

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 this paper, we constructed a stacked autoencoder with two hidden layers and a softmax layer. SAE is as the name implies, a stack of single-level autoencoders. It is a deep learning model. The SAE consists of two layers of sparse autoencoders in which the outputs of each layer are fed into the input of the successive layer as shown in Figure 2. It stacking two or more autoencoders in succession and use the greedy layer-wise for training.

Autoencoders are stacked, so they transform the output of one hidden layer into an input of its successive autoencoder. However, given the input vectors, the objective of the autoencoder 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 equation (7), where xk and zk represent the kth element of x and z, respectively.

$$L(x,z) = -\sum[x_k \ln z_k + \overset{d}{(1-x_k)}_{k=1} ln(1-z_k)] \tag{7}$$

In addition, we can reduce the reconstruction error using the Gradient Descent method (Bottou 2010). The weights in equation (5) and equation (6) must be updated according to equation (8)-equation. (9) and equation (10), where represents the learning rate.

$$W = W - \alpha \frac{\partial L(x,z)}{\partial W}. \tag{8}$$

$$b = b - \alpha \frac{\partial L(x,z)}{\partial b} \tag{9}$$

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

3.2.6 Softmax Classifier

After extracting the relevant features for the training phase, a stacked autoencoder is applied. For the classification phase, we use the softmax classifier at the last layer for our deep learning model to classify the companies based on the softmax classifier as an activation function for classification. Each softmax unit represents one class and the output from softmax unit y_i is the probability that the current input data belongs to class K. The weight matrix between the top layer hidden units and the softmax layer, W_{ij} is trained by minimizing the cost function:

$$J_{softmax} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{k} y_i^{(n)} \log\left(\widetilde{y}_i^{(n)}\right) + \left(1-y_i^{(n)}\right)\log\left(1-\widetilde{y}_i^{(n)}\right) + \frac{\lambda}{2}\sum_i\sum_j W_{ij}^2 \tag{11}$$

Where N is the number of training examples in the mini-batch, y_i is the prediction class i of the input data, and if the input data belongs to category i, assign $\widetilde{y}_i$ to 1, otherwise assign it to 0. Then, we applied the fine-tuning with the backpropagation algorithm for all hidden layers of the network to enhance the performance of our SAE. The output from the softmax units can be defined as follow:

$$y_i = \frac{\exp\sum_j W_{ij} x_j}{\sum_{i=1}^{k} \exp\sum_j W_{ij} x_j} \tag{12}$$

where k is the number of classes. Similarly, the deep network can be used for a regression task if a regressor is placed on top layer hidden units.

In addition, the softmax layer is applied when there are two or more classes. This layer used two parameters L2WeightRegularization and SparsityRegularization received from the SAE to improve the classification performance and to reduce the overfitting. Thus, the use of these two parameters improve classification result and avoid the overfitting. The SparsityProportion parameter is used to

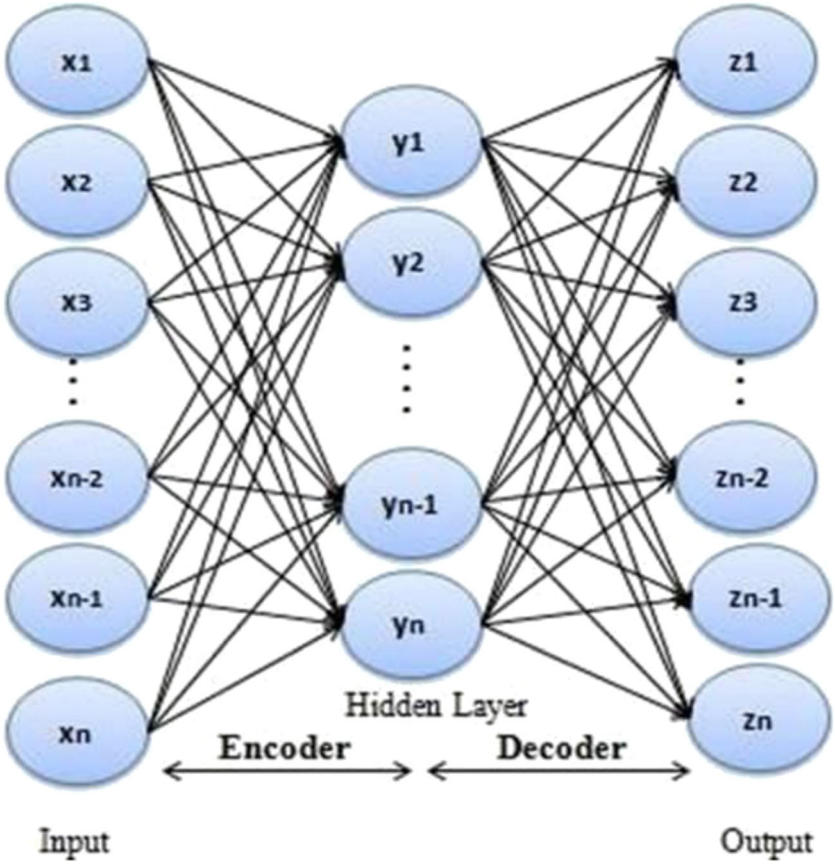

Fig. 1 The structure of an AutoEncoder

controls the average number of activations on a hidden layer. However, L2WeightRegularization is used to reduce overfitting.

As shown in Fig. 3 the proposed model includes three stages: oversampling stage, feature extraction stage and classification stage. First, we applied the oversampling method to balance the dataset. At this level, the Borderline SMOTE algorithm is used. Second, the stacked autoencoder is applied to extract the most important features. Finally, we applied the softmax layer for classification. This layer is stacked with autoencoders to obtain a deep neural network model.

The stacked autoencoder with softmax classifier offers good classification performance by determining the complex relations on the dataset. Each autoencoder, reduce the features of data by removing noisy attributes that facilitate classifying data into different classes. However, the major limitation of the stacked autoencoder is the long run time due to the time spent in the dimension reduction (Almotiri et al. 2017). In Table 1, we compared the computational complexity of the stacked autoencoder with a softmax classifier to the used machine learning methods using asymptotic measures (Cormen et al. 1990; Murphy 2012). As shown in Table 1, in the training phase the computational performance of our proposed method is lower compared to machine learning methods. Thanks to the dimension reduction of the stacked autoencoder, our proposed method provides high performance in the testing phase than the machine learning techniques.

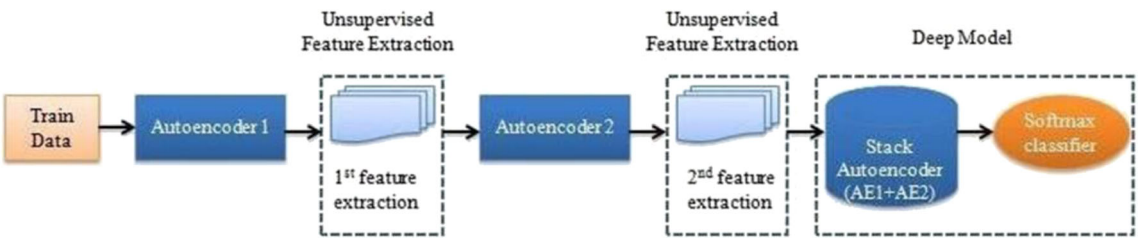

Fig. 2 The structure of a Stacked AutoEncoder

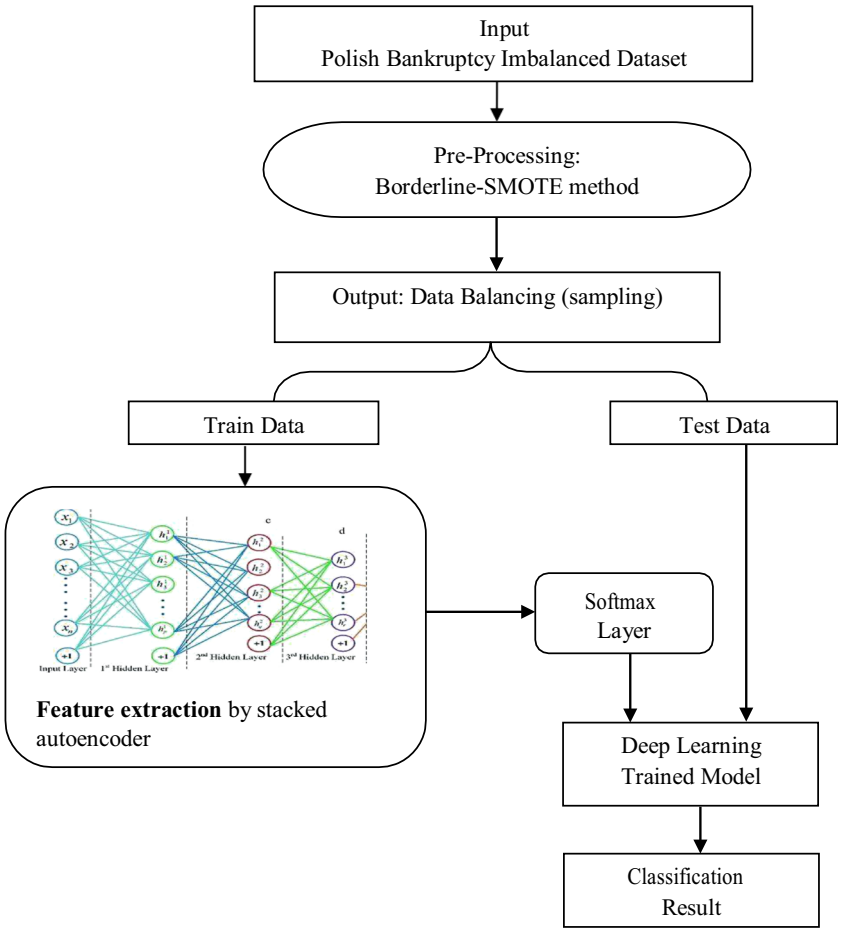

Fig. 3 The proposed bankruptcy prediction model

4 Experimental Setup

In this section, we assess the performance of our proposed deep learning method. We first describe the dataset and the evaluation measures are used to evaluate the performance. Then, we present the experimental results.

4.1 Description of the Experimental Database

The performance of the used algorithms is evaluated based on five-real world datasets of Polish companies. These datasets are obtained from the University of California, Irvine (UCI) Machine Learning Repository (https//archive.ics.uci.edu/ml/datasets.html). Although the companies are analyzed from 2002 to 2013, and still operating companies are evaluated from 2007 to 2013. Each observation presents 64 attributes (financial ratios). The sampling distributions according to the target variable are explained in Table 2. To enhance the reliability of the estimates and reduce the data dependency, Polish datasets are randomly separated into two parts (training and testing partitions). In this work, we use 70% of the dataset for training the classifier while the remaining 30% for testing the performance of the proposed model.

Table 1 Computational complexities of the used techniques

Method	Computational complexity	Parameter
Artificial neural network	O(*hNM*)	h: number of hidden neurons
Decision Tree	O(*H*)	N: number of observations
Random Forest	O(*MNlog(N)*)	M: number of features
K nearest neighbors	O(*MLog(k)NLog(N)*)	H: height of tree
Support vector machine using Linear	O(N^2)	k: number of neighbors
Support vector machine using Gaussian	O(N^2M)	K: Sum number hidden
Stacked autoencoder+softmax layer	O($NK + MhN$)	neurons of AE

Table 2 The distribution of the samples for 5 cases

Data set	Bankruptcy after	Bankrupt		Not bankrupt	
		Number	Rate	Number	Rate
1st year	5 years	271	3,86	6756	96,14
2nd year	4 years	400	3,94	9773	96,06
3rd year	3 years	495	4,71	10,008	95,29
4th year	2 years	515	5,26	9277	94,74
5th year	1 years	410	6,94	5500	93,06

4.2 Evaluation Criteria

Different evaluation criteria are applied to determine the feasibility and the effectiveness of the used techniques for predicting bankruptcy include 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 is the most commonly used measure to assess the performance of prediction models for an imbalanced dataset (Veganzones and Séverin 2018). The definition of this performance measure can be formulated using 2 2 confusion matrix as explained in Table 3.

In general, the accuracy rate is the most widespread performance criterion used in evaluating the predictive accuracy of classification models. Many bankruptcy prediction studies used accuracy as a measure for evaluating the performance of algorithms. However, this metric is inappropriate for the imbalanced dataset. To this aim, the area under the ROC curve (AUC) has been considered adequate criteria for evaluating and comparing the bankruptcy prediction model, because it is insensitive to misclassified costs and imbalanced distributions. Furthermore, AUC is commonly used for the binary classification problem; it illustrates the trade-off between the true positive rate and false-positive rate. AUC measure is given by the following equation:

$$AUC = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{13}$$

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

$$Sensitivity = \frac{TP}{TP + FN} \tag{14}$$

Table 3 A confusion matrix

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

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

$$Specificity = \frac{TN}{TN + FP} \tag{15}$$

Furthermore, the training time performance is an important measure used to evaluate the efficiency and the performance of the machine and deep learning classifiers. To this end, we will compare the training time of the used classifiers.

5 Results and Discussion

Experiments were described in detail in Fig. 3 combining the Borderline SMOTE as oversampling method with a deep learning model which consists of two autoencoders and softmax layer. The stacked autoencoder is preferred to the well known deep learning methods for many reasons: thanks to the inter- layer communication, learning in the stacked autoencoder is performed better than Restricted Boltzman Machine (RBM) (Wang et al. 2017). Moreover, we used the stacked autoencoder because it shows a better performance compared to deep belief network when trained on noisy datasets (Gaussian noise) (Wang et al. 2017; Kasun et al. 2013). Variational autoencoder (VAE) recently shows impressive performance on larger and more complex datasets such as bigdata, microarray (Sønderby et al. 2016).

First, we need to set out the deep learning parameter because it influences significantly the performance of search algorithms on a given search problem. The classification accuracy of the 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 are 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 autoencoder contains 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. The experimental results of the Polish datasets demonstrated that the SAE has high AUC when the number of hidden layers is two. 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 AUC for the five 5 datasets (shown in Table 4) when the number of neurons in the first hidden layer is 50 and 30 in the second hidden layer. Sparse Regularization is an important parameter that controls the regularization value of the objective function. This value must be

Table 4 AUC performance for Polish bankruptcy datasets

Oversampling method	Classifier	1st year	2nd year	3rd year	4th year	5th year
None	KNN	0.480	0.485	0.479	0.476	0.637
	DT	0.674	0.577	0.593	0.631	0.736
	SVM-Linear	0.480	0.485	0.479	0.492	0.500
	SVM-Gaussian	0.647	0.486	0.499	0.620	0.604
	ANN	0.520	0.480	0.481	0.499	0.510
	RF	0.770	0.700	0.666	0.710	0.774
	C5.0	0.845	0.890	0.881	0.90	0.882
	SAE + Softmax	0.864	0.894	0.890	0.902	0.905
BSM	KNN	0.859	0.723	0.862	0.826	0.856
	DT	0.840	0.863	0.834	0.858	0.908
	SVM-Linear	0.824	0.798	0.813	0.836	0.845
	SVM-Gaussian	0.806	0.786	0.757	0.751	0.833
	ANN	0.685	0.735	0.677	0.744	0.734
	RF	0.910	0.921	0.894	0.914	0.912
	C5.0	0.965	0.950	0.968	0.969	0.950
BSM	SAE + Softmax	0.961	0.962	0.950	0.980	0.962

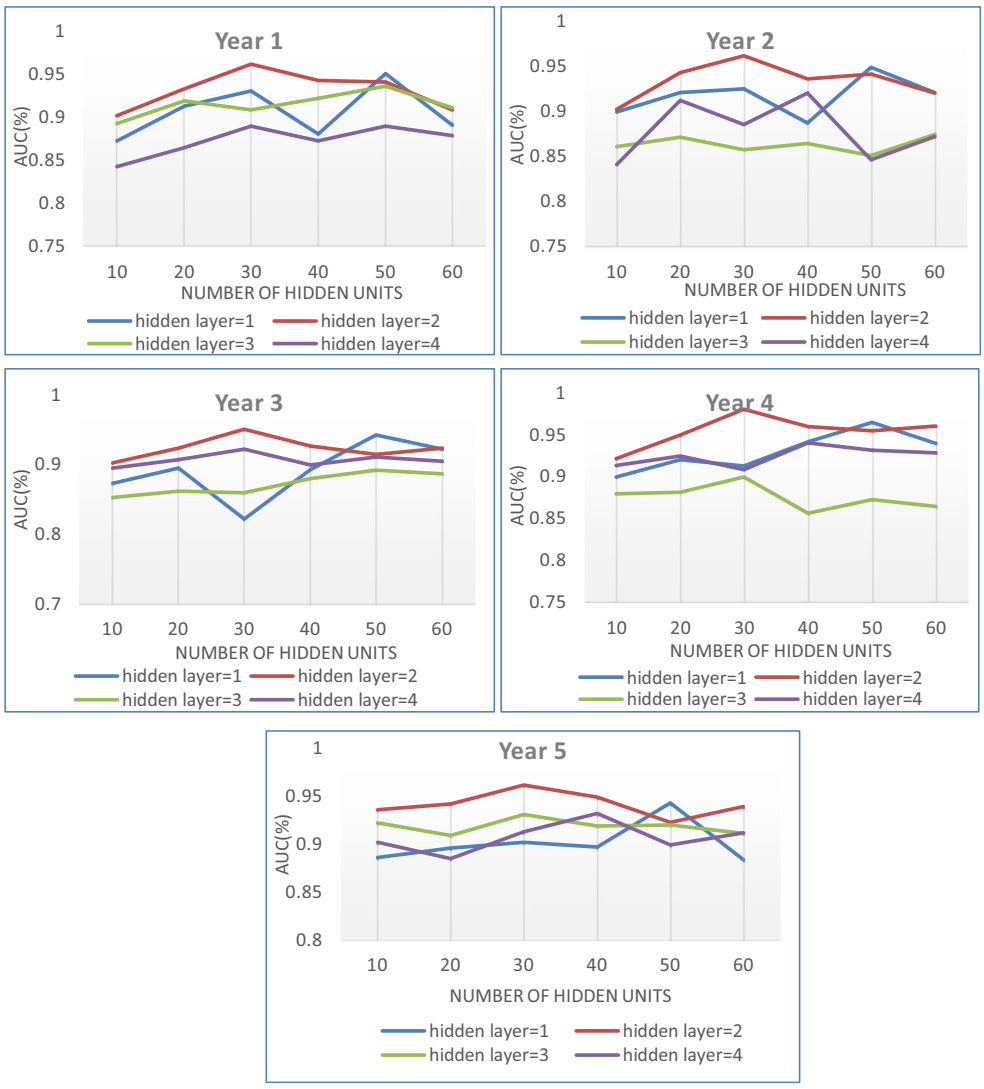

Fig. 4 The impact of the number of SAE hidden layers and the number of hidden units on classification accuracy for Polish datasets

Table 5 Significant test results of paired t-test ($\alpha = 0.05$)

Method A	Method B	AUC	
		t	*p* value
BSM-SAES	BSM-KNN	4.986	0.0076
	BSM-DT	8.009	0.0013
	BSM-SVM-Linear	18.527	0.0001
	BSM-SVM-Gaussian	10.406	0.0005
	BSM-ANN	22.668	<0.0001
	BSM-RF	12.905	0.0002
	BSM-C5.0	11.302	0.0003

between 0.5 and 5. L2 Regularization is used to reduce overfitting. Due to this parameter, the precise changes that lead to overfitting are stretched. In our study, L2 Regularization value is equal to 0.01 and Sparse Regularization value is set to 4. Figure 4 shows the effect of the number of hidden layers and the number of hidden units on the AUC of SAE. The datasets used in our experiment are characterized by an important number of features (64 features). The autoencoder is trained with a lesser number of neurons in the hidden layer than the input layer. For the second hidden layer, the number of the neuron is less than half (30 neurons) of the number of input features which is sufficient to extract the most relevant features. In this paper, A layer by layer greedy training method is used in the SAE model. This method provides a very promising results in terms of AUC.

At present, several machine learning techniques such as neural network, support vector machine, decision tree, and k-nearest neighbor have been successfully used to predict bankruptcy, and they usually have good prediction accuracy (Barboza et al. 2017; Wang 2017). Therefore, we compare our classifier performance with machine learning techniques. For KNN, the number of neighbors (K) is the most important parameter to improve classification accuracy. In addition, the optimal k which provides the lowest test error rate. For choosing K value is determined experimentally, starting with k=1. This process is repeated many time by incrementing K to found several neighbors. The k value which provides the maximum accuracy is selected. In general, the value of K increases with the size of the training dataset. In this study, KNN reaches the best performance in terms of AUC when k=5 using the euclidean metric for distance measure. For the ANN algorithm, we adjusted various parameters to optimize the model include the number of the hidden layers, the number of neurons in each hidden layer, and the initial weights. Methodology starts with ANN having one hidden layer and randomly selected neurons. The initial values of the weights are set to 1. The proposed methodology runs for a maximum 3 hidden layers with a different number of neurons or a different set of initial weights. Our method gives the best results when the number of hidden layers is set to 1, the number of neurons in each layer is set to 30 and the initial weights are equals to 1. For SVM, during the training, phase it searches the hyperplane with the largest margin that is the maximum marginal hyperplane. The experimental result demonstrated that the SVM has the best results in terms of AUC when the kernel scale is set to 4.5 for the Gaussian kernel function and 0.5 for the linear kernel function. The datasets used in this work were divided: 70% for training the classifiers and 30% for testing the performance of the results.

Table 4 reports the AUC result of the algorithms applied in experimental studies, to evaluate the efficiency of our approach for predicting the financial condition of the enterprises. From Table 4, we notice that for the Polish datasets, the deep neural networks method which is the stacked autoencoder combined with Borderline SMOTE, has the best AUC rate compared

Table 6 Time performance results for bankruptcy classification (milliseconds)

Oversampling method	Classifier	1st year	2nd year	3rd year	4th year	5th year
None	KNN	0.098	0.101	0.108	0.107	0.089
	DT	0.238	0.392	0.412	0.386	0.219
	SVM-Linear	0.051	0.041	0.058	0.042	0.014
	SVM-Gaussian	0.065	0.117	0.172	0.161	0.039
	ANN	0.640	0.488	1.234	0.730	0.797
	RF	5.360	8.299	8.496	8.125	6.074
	C5.0	0.528	0.548	0.452	0.989	0.479
BSM	KNN	2.108	2.765	1.289	1.390	0.435
	DT	1.444	1.646	1.858	1.771	1.264
	SVM-Linear	45.514	48.426	37.586	35.291	30.750
	SVM-Gaussian	14.167	21.132	17.559	16.442	11.023
	ANN	1.023	0.776	1.261	2.554	1.856
	RF	14.465	14.095	16.750	14.888	20.189
	C5.0	1.102	1.723	2.864	1.339	0.905
BSM	SAE + Softmax	60.419	96.994	100.030	85.177	52.633

Table 7 Recommended method and comparison with similar works for bankruptcy Polish dataset

Authors	Method	AUC				
		1st year	2nd year	3rd year	4th year	5th year
(Zieba et al. 2016)	EXGB	0.959	0.944	0.940	0.941	0.955
(Broelemann and Kasneci 2018)	XGB	0.944	0.903	0.902	0.925	0.944
(Fan et al. 2018)	Isolation Forest	0.930	0.950	0.940	0.950	0.960
Proposed method	BSM-SAES	**0.961**	**0. 962**	**0.950**	**0.980**	**0.962**

to the machine learning methods. Regarding the machine learning methods, it is shown that C5.0 outperforms the other methods. After applying the machine learning methods to the polish bankruptcy datasets, we compared the proposed model by applying the oversampling method to the proposed machine learning methods. The result shows that the Borderline SMOTE improves the performance of the machine learning methods. However, the proposed stacked autoencoder with Borderline SMOTE model achieves the highest AUC compared with the other proposed techniques.

First of all, we apply Borderline SMOTE as a data preprocessing method to balance the original data set by generating new instances from existing minority cases. Second, we split the data into the training set and testing set followed by extracting the important features using the stacked autoencoder, then we apply the softmax to classify the samples. We emphasis that after applying oversampling methods to my imbalanced datasets, the performance of the proposed classifier is enhanced. This result is more improved by applying the deep learning model, which is able to extract automatically useful features during the training step based on stacked autoencoders. This features extraction makes DL model highly accurate for bankruptcy prediction task. For this reason, our deep learning model outperforms the machine learning methods that performed the feature extraction step independently of the training phase.

For all datasets, the stacked autoencoder with softmax classifier based on the Borderline SMOTE oversampling method is significantly better than the other methods. In order to statistically confirm the hypothesis, the paired t-test is applied to test the significance of the above results. The null hypothesis is "the AUC of model A= AUC of model B ". The alternative hypothesis is "AUC of model A ≠ AUC of model B". The results are summarized in Table 5. As seen in this table, the proposed BSM-SAES is significantly better than the machine learning methods. Therefore, the null hypothesis is rejected at significance level equal 0.05. Concluding, Borderline SMOTE-Stacked autoencoder with softmax classifier performs better than the used methods.

In addition, Table 6 presents the training-time performances of the algorithms used in our experiments. As shown in Table 6, the proposed stacked autoencoder seems to be worse in terms of time performance than the machine learning models because in these models there is no dimension reduction. The comparative results of the proposed method with the existing similar works are shown in Table 7. The Polish bankruptcy prediction has been tackled by the ensemble methods, where, a large number of them are successfully used to deal with imbalanced data for bankruptcy prediction. As seen in Table 7, For the Polish benchmark datasets, BSM-SAES provides the highest AUC compared with other techniques that exist in the literature.

Based on the values in Tables 4, 5 and 7, we notice that BSM-SAES has the best performance on Polish bankruptcy dataset. It provides the highest AUC rate in comparison to machine learning techniques and to all reference classifiers of bankruptcy prediction. Therefore, we conclude that BSM-SAES is an efficient technique for predicting bankruptcy by correctly classifying firms either bankrupt or non-bankrupt. Thus, the proposed technique can be used as a feasible solution to enhance the accuracy in predicting the bankrupt firms.

6 Conclusion

Bankruptcy prediction is considered as the most critical and important issue in the financial institution. In existing studies, various methods have been proposed for supporting better decision-making. Most of these works used machine learning algorithms. However, these models are not perfect for bankruptcy prediction problem. They have failed to extract complex and non-linear patterns from big data. In this context, we proposed an accurate bankruptcy prediction model using deep learning algorithm. In the proposed model, a two-

layer autoencoder is appliedto 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 SAE performance. To solve the problem of the used imbalanced data, we combined the deep learning method with Borderline SMOTE oversampling method. The results of the conducted experiments show that the BSM- SAES outperforms the other's applied methods in terms of AUC. However, the proposed model achieved a worse training time performance due to the large amount of time spent in extracting the important features for the classification. However, we plan to continue our research for enhancing the AUC of the bankruptcy prediction model by using a comprehensible evaluation model based on IF-THEN rules.

References

P. Addo, D. Guegan, B. Hassani, P. Addo, D. Guegan, B. Hassani, C. Risk (2018) Credit Risk analysis using machine and deep learning models HAL id : Halshs-01719983 Centre d ' Economie de la Sorbonne documents de travail du

Ahn, A., & Kim, K. J. (2011). Corporate credit rating using multiclass classification models with order information. *International Journal of Economics and Management Engineering, 5*(12), 1800–1811.

Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid adaptive Neuro fuzzy inference system (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research, 222*(1), 168–178.

T. Aliaj, A. Anagnostopoulos, S. Piersanti (2018) Firms Default Prediction with Machine Learning.

J. Almotiri, K. Elleithy, A. Elleithy (2017) Comparison of autoencoder and principal component analysis followed by neural network for e-learning using hand-written recognition. In systems, applications and technology conference (LISAT). IEEE long island (pp. 1–5). IEEE.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*, 589–609. https://doi.org/10.2307/2329297.

Ashoori, S., & Mohammadi, S. (2011). Procedia computer compare failure prediction models based on feature selection technique : Empirical case from Iran. *Procedia Comput. Sci., 3*, 568–573. https://doi.org/10.1016/j.procs.2010.12.094.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications, 83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006.

W.H. Beaver. Financial Ratio as Predictors of Failure, Empirical Research in Accounting: Selected Studies 1966 (1966) Journal of Accounting Research Supplement to vol. 4 : https://doi.org/10.2307/2490171, 71-111.

A. Bordes, S. Chopra, and J. Weston (2014) Question answering with subgraph embeddings, Conference on Empirical Methods in Natural Language Processing (EMNLP): 615–620.

L. Bottou Large-scale machine learning with stochastic gradient descent, Proc. COMPSTAT 2010 - 19th Int. Conf. Comput. Stat. Keynote, Invit. Contrib. Pap. (2010) 177–186. https://doi.org/10.1007/978-3-7908-2604-3-16.

A. Bouallégue, S. Hassairi, R. Ejbali, M. Zaied (2016) Learning Deep Wavelet Networks for Recognition System of Arabic Words. International Joint Conference SOCO'16-CISIS'16- ICEUTE'16: 498–507.

K. Broelemann, G. Kasneci (2018) A Gradient-Based Split Criterion for Highly Accurate and Transparent Model Trees. http://arxiv.org/abs/1809.09703.

Callejón, A. M., Casado, A. M., Fernández, M. A., & Peláez, J. I. (2013). A system of insolvency prediction for industrial companies using a financial alternative model with neural networks. *International Journal Computational Intelligence Systems, 6*, 29–37. https://doi.org/10.1080/18756891.2013.754167.

Chaudhuri, A., & Ghosh, S. K. (2018). *Bankruptcy Prediction through Soft Computing based Deep Learning Technique*. https://doi.org/10.1007/978-981-10-6683-2.

Chawla, K. W. P., Bowyer, N. V., Hall, K. W., & L.O. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

Chen, M. Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications, 38*(9), 11261–11272. https://doi.org/10.1016/j.eswa.2011.02.173.

R. Collobert, J. Weston, A unified architecture for natural language processing, Proc, International Conference on Machine Learning.-ICML '08. (2008) 160–167. doi:https://doi.org/10.1145/1390156.1390177.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research, 12*, 2493–2537.

R. L. Constand, 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: https://doi.org/10.1007/978-1-4419-7527-0.

T. Cormen, C. Leiserson, R. Rivest, Introduction to algorithms (1990). MIT Press

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297.

Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Trans. Audio, Speech Lang. Process, 20*, 30–42. https://doi.org/10.1109/TASL.2011.2134090.

T. Daniel, Discovering knowledge in data: An introduction to data mining, second edition. IEE computer society (2014).

De Andrés, J., Landajo, M., & Lorca, P. (2012). Bankruptcy prediction models based on multinorm analysis: An alternative to accounting ratios. *Knowledge-Based Systems, 30*, 67–77. https://doi.org/10.1016/j.knosys.2011.11.005.

Dey, D. (2017). Growing Importance of Machine Learning in Compliance and Regulatory Reporting. *European Journal of Multidisciplinary studies, 6*, 255–258. https://doi.org/10.26417/ejms.v6i2.p255-258.

Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics, 6*(4), 325–327. https://doi.org/10.1109/TSMC.19765408784.

Ejbali, R., & Zaied, M. (2018). A dyadic multi-resolution deep convolutional neural wavelet network for image classification. *Multimedia Tools and Applications, 77*(5), 6149–6163. https://doi.org/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*, 1543–1555. https://doi.org/10.1080/00949655.2012.666550.

S. Ertekin (2013) Adaptive Oversampling for Imbalanced Data Classification. Information Sciences and Systems. doi:https://doi.org/10.1007/978-3-319-01604-7.

S. Fan, G. Liu, Z. Chen (2018) Anomaly detection methods for bankruptcy prediction, 2017 4th International Conference on Systems and Informatics, ICSAI 2017. 1456–1460. doi:https://doi.org/10.1109/ICSAI.2017.8248515.

V. Garcı, J.S. Sa, A.I. Marque (2013) On the suitability of resampling techniques for the class imbalance problem in credit scoring, 1060–1070. doi:https://doi.org/10.1057/jors.2012.120.

A. Gepp, K. Kumar (2015) Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques, Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015) 396–404. doi:https://doi.org/10.1016/j.procs.2015.06.046.

A. Gepp, K. Kumar, S. Bhattacharya (2010) Business failure prediction using decision trees, journal of forecasting, 536–555 doi: https://doi.org/10.1002/for.1153.

Ghaddar, B., & Naoum-sawaya, J. (2017). PT us CR. *Eur. J. Oper. Res.* https://doi.org/10.1016/j.ejor.2017.08.040.

Han, H., Wang, W. Y., & Mao, B. H. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. *ICIC*, 878–887.

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Vanhoucke, V., Nguyen, P., Sainath, T., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine, 29*, 82–97. https://doi.org/10.1109/MSP.2012.2205597.

Jack, L. B., & Nandi, A. K. (2002). Fault detection using support vector machines and artificial neural networks. *Augmented by Genetic Algorithms, 16*, 373–390. https://doi.org/10.1006/mssp.2001.1454.

Jang, M., Seo, S., & Kang, P. (2018) Recurrent neural network-based semantic variational autoencoder for sequence-to-sequence learning. http://arxiv.org/abs/1802.03238.

Jardin, P. D. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing, 73*, 2047–2060. https://doi.org/10.1016/j.neucom.2009.11.034.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review, 106/4*, 620–630.

S. Jean, K. Cho, R. Memisevic, Y. Bengio (2014) On using very large target vocabulary for neural machine translation, 000. http://arxiv.org/abs/1412.2007.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *J. Big Data, 6*. https://doi.org/10.1186/s40537-019-0192-5.

Karels, G. V., & Prakash, A. J. (1987). Multivariate normality and forecasting of business bankruptcy. *J. Bus. Financ. Account, 14*, 573–593. https://doi.org/10.1111/j.1468-5957.1987.tb00113.x.

Kasgari, A. A., Divsalar, M., Javid, M. R., & Ebrahimian, S. J. (2012). Prediction of bankruptcy Iranian corporations through artificial neural network and Probit-based analyses. *Neural Computing and Applications, 23*(3–4), 927–936. https://doi.org/10.1007/s00521-012-1017-z.

Kasun, L. L. C., Zhou, H., Huang, G. B., & Vong, C. M. (2013). Representational learning with extreme learning machine for big data. *IEEE Intelligent Systems, 28*(6), 31–34.

A. Krizhevsky, I. Sutskever, G.E. Hinton (2012) ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 1–9. https://doi.org/10.1016/j.protcy.2014.09.007.

Z. Lanbouri, S. Achchab (2015) A hybrid Deep belief network approach for Financial distress prediction, 2015 10th International Conference on Intelligent Systems Theories Application1–6. doi:https://doi.org/10.1109/SITA.2015.7358416.

F. Last, G. Douzas, F. Bacao (2018) Oversampling for imbalanced learning based on K-means and SMOTE information sciences 465.

T. Le, M.Y. Lee, J.R. Park, S.W. Baik (2018) Oversampling techniques for bankruptcy prediction: Novel features from a transaction dataset, Symmetry (Basel). 10. doi:https://doi.org/10.3390/sym10040079.

Lee, L. H., Wan, C. H., Rajkumar, R., & Isa, D. (2012). An enhanced support vector machine classification framework by using Euclidean distance function for text document categorization. *Applied Intelligence, 37*(1), 80–99. https://doi.org/10.1007/s10489-011-0314-z.

Lee, J., Jang, D., & Park, S. (2017). Deep learning-based corporate performance prediction model considering technical capability. *Sustain., 9*, 1–12. https://doi.org/10.3390/su9060899.

V. López, A. Fernández, S. García, V. Palade, F. Herrera (2013) An insight into classification with imbalanced data : Empirical results and current trends on using data intrinsic characteristics. doi:https://doi.org/10.1016/j.ins.2013.07.007

Lu, H., & Li, Y. (2017). *Brain Intelligence : Go beyond Artificial Intelligence*. https://doi.org/10.1007/s11036-017-0932-8.

H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, S. Serikawa (2017) Motor anomaly detection for unmanned aerial vehicles using reinforcement learning, 4662. doi:https://doi.org/10.1109/JIOT.2017.2737479.

Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research, 274*, 743–758. https://doi.org/10.1016/j.ejor.2018.10.024.

Mbarki, N. E. H., Ejbali, R., & Zaied, M. (2017). Recognition of human activities in smart homes using stacked autoencoders. *Tenth Int. Conf. Adv. Comput. Interact. ACHI, 2017*, 176–180.

McKee, T. E., & Lensberg, T. (2002). Genetic programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research, 138*, 436–451. https://doi.org/10.1016/S0377-2217(01)00130-8.

Min, J. H., & Jeong, C. (2009). A binary classification method for bankruptcy prediction. *Expert Systems with Applications, 36*, 5256–5263. https://doi.org/10.1016/j.eswa.2008.06.073.

M.R. Minar, J. Naher (2018) Recent Advances in Deep Learning: An Overview, 2006 1–31. doi:https://doi.org/10.13140/RG.2.2.24831.10403.

K. Murphy, Machine learning: A probabilistic perspective (2012). MIT Press.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data., 2*, 1–21. https://doi.org/10.1186/s40537-014-0007-7.

Nekooeimehr, I., & Lai-Yuen, S. K. (2016). Adaptive semi-unsupervised weighted oversampling (A- SUWO) for imbalanced datasets. *Expert Systems with Applications, 46*, 405–416. https://doi.org/10.1016/j.eswa.2015.10.031.

Nguyen, H. G. (2005). Using Neutral Network in Predicting Corporate Failure. *Journal of Social Sciences (15493652), 1*(4), 199–202.

R. Nilsson, M. Pe, J. Bjorkegren, J. Tegnér (2006) Evaluating feature selection for SVMs in high dimensions. European Conference on Machine Learning.

S. H. Syed Nor, S. Ismail, B.W. Yap, Personal bankruptcy prediction using decision tree model, (2020). doi:https://doi.org/10.1108/JEFAS-08-2018-0076.

Ocal, N., Ercan, M. K., & Kadıoğlu, E. (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). https://doi.org/10.5539/ijef.v7n7p189.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*, 109. https://doi.org/10.2307/2490395.

B. B. Ribeiro, and N. Lopes (2011) Deep belief networks for financial prediction, International Conference on Neural Information Processing; Springer: Berlin/Heidelberg, Germany : 766–773.

Ripley, B. D. (2008). Pattern recognition and neural networks. *Technometrics, 39*(2), 403.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation (no. ICS-8506). *California Univ San Diego La Jolla Inst For Cognitive Science, 1*, 318–362.

S. Said, O. Jemai, S. Hassairi, R. Ejbali, M. Zaied, C. Ben Amar (2016) Deep wavelet network for image classification, IEEE International Conference on Systems, Man, and Cybernetics, SMC - Conference Proceedings: 922–927.

Sanchis, A., Segovia, M. J., Gil, J. A., Heras, A., & Vilar, J. L. (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*, 1554–1573. https://doi.org/10.1016/j.ejor.2006.01.045.

Santoso, N., & Wibowo, W. (2018). Financial Distress Prediction using Linear Discriminant Analysis and Support Vector Machine. *International Conference on Science (ICOS)*. https://doi.org/10.1088/1742-6596/979/1/012089.

Serikawa, S., & Lu, H. (2014). Underwater image dehazing using joint trilateral filter q. *Computers and Electrical Engineering, 40*, 41–50. https://doi.org/10.1016/j.compeleceng.2013.10.016.

Sharma, H., & Kumar, S. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. *Int. J. Sci. Res., 5*, 2094–2097. https://doi.org/10.21275/v5i4.nov162954.

Shin, K. S., & Lee, Y.-J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications, 23*, 321–328. https://doi.org/10.1016/S0957-4174(02)00051-9.

A. Shrivastava, K. Kumar (2018) Business Distress Prediction Using Bayesian Logistic Model for Indian Firms. doi:https://doi.org/10.3390/risks6040113.

Skryjomski, P., & Krawczyk, B. (2017). Influence of minority class instance types on SMOTE imbalanced data oversampling. *Proc. Mach. Learn. Res., 74*, 7–21 http://proceedings.mlr.press/v74/skryjomski17a/skryjomski17a.pdf.

S. Smiti, M. Soui, I. Gasmi (2018) A comparative study of rule based classification algorithms for credit Risk assessment. The 31st IBIMA, International Business Information Management Conference.

S. Smiti, M. Soui, R. Ejbali, G Khaled (2019) Comparing Approaches for Combining Data sampling and Stacked Autoencoder to address Bankruptcy Prediction. the 34th IBIMA, International Business Information Management Conference.

Sønderby, C. K., Raiko, T., Maaløe, L., Sønderby, S. K., & Winther, O. (2016). Ladder variational autoencoders. *Advances in Neural Information Processing Systems*, 3738–3746.

V.S. Spelmen, R. Porkodi (2018) A review on handling imbalanced data, 2018 Int Conf. Curr. Trends Towar. Converging Technol. 1–11.

M. A. Sprengers (2005) Bankruptcy prediction using classification and regression trees.

S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. S Kruthiventi, R.V. Babu, (2016) A Taxonomy of Deep Convolutional Neural Nets for Computer Vision, Frontiers in Robotics and AI. doi: https://doi.org/10.3389/frobt.2015.00036.

I. Sutskever, O. Vinyals, Q. V Le (2014a) Sequence to sequence learning with neural networks, Advances in Neural Information Processing Systems. 3104–3112. doi:https://doi.org/10.1007/s10107-014-0839-0.

I. Sutskever, O. Vinyals, Q. V Le (2014b), Sequence to sequence learning with neural networks, Adv. Neural Information Processing Systems 3104–3112. doi:https://doi.org/10.1007/s10107-014-0839-0.

Toribio, P., Alejo, R., Valdovinos, R. M., & Pacheco Sanchez, J. H. (2012). Using Gabriel graphs in borderline-SMOTE to deal with severe two-class imbalance problems on neural networks. *Artificial Intelligence Research and Development*. https://doi.org/10.3233/978-1-61499-139-7-29.

T. Van Gestel, B. Baesens, J. Suykens, M. Espinoza, D.-E. Baestaens, J. Vanthienen, De Moor B., Bankruptcy prediction with least squares support vector machine classifiers, 2003 IEEE international conference computer intelligence financial engineering 2003. Proceedings (n.d.) 1–8. doi:https://doi.org/10.1109/CIFER.2003.1196234.

Veganzones, D., & Séverin, E. (2018). An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems, 112*, 111–124. https://doi.org/10.1016/j.dss.2018.06.011.].

M. Verleysen, D. François, G. Simon, V. Wertz, On the effects of dimensionality on data analysis with, (2003) 105–112.

Wang, N. (2017). Bankruptcy prediction using machine learning. *J. Math. Financ., 07*, 908–918. https://doi.org/10.4236/jmf.2017.74049.

Wang, G., Ma, J., & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications, 41*, 2353–2361. https://doi.org/10.1016/j.eswa.2013.09.033.

Wang, K. J., Adrian, A. M., Chen, K. H., & Wang, K. M. (2015). A hybrid classifier combining borderline- SMOTE with AIRS algorithm for estimating brain metastasis from lung cancer: A case study in Taiwan. *Computer Methods and Programs in Biomedicine, 119*, 63–76. https://doi.org/10.1016/j.cmpb.2015.03.003.

H. Wang, B. Raj, E. P Xing. On the Origin of Deep Learning (2017), abs/1702.07800, Access Date: 29.08.2018 Web.

Xie, C., Luo, C., & Yu, X. (2011). Financial distress prediction based on SVM and MDA methods: The case of Chinese listed companies. *Quality and Quantity, 45*, 671–686. https://doi.org/10.1007/s11135-010-9376-y.

Yeh, C. C., Chi, D. J., & Lin, Y. R. (2014a). Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences, 254*, 98–110. https://doi.org/10.1016/j.ins.2013.07.011.

S. Yeh, C. Wang, M.-F. Tsai (2014b) Corporate default prediction via deep learning, International Inst. Forecast. http://forecasters.org/wp/wp-content/uploads/gravity_forms/7-2a51b93047891f1ec3608bdbd77ca58d/2014/07/Yeh_Shu-Hao_ISF2014.pdf.

Yu, L., Yang, Z., & Tang, L. (2015). A novel multistage deep belief network based extreme learning machine ensemble learning paradigm for credit risk assessment. *Flexible Services and Manufacturing, 28*, 576–592. https://doi.org/10.1007/s10696-015-9226-2.

R. Yusof, K.A. Kasmiran, A. Mustapha, U. Tun, H. Onn (2017) Techniques for handling imbalanced datasets when producing classifier models. Journal of Theoretical and Applied Information Technology.

Zhang, Q., Wang, J., Lu, A., Wang, S., & Ma, J. (2018). An improved SMO algorithm for financial credit Risk assessment–evidence from China's banking. *Neurocomputing, 272*, 314–325. https://doi.org/10.1016/j.neucom.2017.07.002.

L. Zhang, B. Tan, T. Liu, X. Sun (2019) Classification study for the imbalanced data based on Biased-SVM and the modified over-sampling algorithm Classification study for the imbalanced data based on Biased-SVM and the modified over-sampling algorithm, . doi:https://doi.org/10.1088/1742-6596/1237/2/022052.

Zhou, L. (2013). Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Syst., 41*, 16–25. https://doi.org/10.1016/j.knosys.2012.12.007.

Zhou, L., Lai, K. K., & Yen, J. (2012). Empirical models based on features ranking techniques for corporate financial distress prediction.

Comput. Math. with Appl., 64, 2484–2496. https://doi.org/10.1016/j.camwa.2012.06.003.

Zieba, M., Tomczak, S. K., & Tomczak, J. K. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications, 58*, 93–101. https://doi.org/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.

Salima Smiti obtained her baccalaureate degree in 2011 from Mohamed Ali School Elhamma. She obtained her diploma and her master degree in Computer Sciences from the Higher Institute of Management of Gabes in 2014 and in 2017. Since 2018, she is pursuing her doctoral studies at the National School of Computer Science of Manouba, Tunisia (ENSI), and she is a member of Research Groups on COSMOS research laboratory. She is currently teaching at the Higher Institute of Management of Gabes. Her current research interests are focused on the application of machine learning and deep learning techniques on the field of the financial institution.

Makram Soui received a PhD in Computer Science from the University of Valenciennes, France, in 2010. He is currently an Assistant Professor in College of Computing and Informatics Saudi Electronic University, Riyadh. He teaches: Artificial intelligence, Data Mining, Information retrieval, Ontology and semantic web, Human-Computer Interaction, Fuzzy logic. He is a member of (Automated Reasoning and Human-Computer Interaction) research group in LAMIH France. He works in collaboration with Institute of Management of Gabes in Tunisia.