INSOLVENCY PREDICTION ANALYSIS OF ITALIAN SMALL FIRMS BY DEEP LEARNING

Agostino Di Ciaccio[1] and Giovanni Cialone[2]

[1]Dept. of Statistics, university of Rome, La Sapienza, Italy
[2]Senior partner of Kairos Advisory srl., Italy

ABSTRACT

To improve credit risk management, there is a lot of interest in bankruptcy predictive models. Academic research has mainly used traditional statistical techniques, but interest in the capability of machine learning methods is growing. This Italian case study pursues the goal of developing a commercial firms insolvency prediction model. In compliance with the Basel II Accords, the major objective of the model is an estimation of the probability of default over a given time horizon, typically one year.

The collected dataset consists of absolute values as well as financial ratios collected from the balance sheets of 14.966 Italian micro-small firms, 13,846 ongoing and 1,120 bankrupted, with 82 observed variables. The volume of data processed places the research on a scale like that used by Moody's in the development of its rating model for public and private companies, Riskcalc^{TM}. The study has been conducted using Gradient Boosting, Random Forests, Logistic Regression and some deep learning techniques: Convolutional Neural Networks and Recurrent Neural Networks. The results were compared with respect to the predictive performance on a test set, considering accuracy, sensitivity and AUC. The results obtained show that the choice of the variables was very effective, since all the models show good performances, better than those obtained in previous works. Gradient Boosting was the preferred model, although an increase in observation times would probably favour Recurrent Neural Networks.

KEYWORDS

Credit risk, Bankruptcy prediction, Deep learning

1. INTRODUCTION

Corporate insolvency and credit risk assessment have been the subject of much academic and professional research over the past half century. There are several reasons why this research is important: we know that insolvency generates insolvency, producing the domino effect, but it is also a general economic and social question. The current economic situation, with various turbulences in the financial markets, requires important developments in the forecast of bankruptcies.

Our study was conducted first by gathering an important database on small and medium-sized Italian companies. For every company we have collected many variables taken from the budgets and we have carried out a pre-processing. The data were then analysed using some advanced techniques in Machine Learning and Deep Learning, such as Random Forests, Gradient Boosting, Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN). We then compared the performance of these models on a test-set of real data.

In the first part of the article we describe the state of the art, analysing the proposals in the literature on this topic, then we briefly introduce the techniques of machine learning and deep learning that will be applied. Subsequently we illustrate the database we have created for this application and the results obtained. Finally, the conclusions contain some considerations and future research ideas.

2. LITERATURE REVIEW

The first methodological approaches to predicting insolvency date back to 1932. Paul Joseph FitzPatrick is the author of one of the first articles on bankruptcy prediction [1]. He presented data for 19 pairs of companies and compared the financial ratios as possible indicators of failure. Only in 1966, Beaver [2] analysed a dataset of 13 balance sheet ratios for 38 companies (19 failed and 19 ongoing). His univariate model is the first research based on statistics: "Financial ratios as predictors of failure". Beaver compared the average of the values of 30 financial ratios of 79 failed and 79 not failed companies across 38 economic sectors.

Starting in 1968, the attention, from this point on, has been placed on the first multivariate research by Altman on "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy" [3].

Successively, Altman [4] emphasized that the use of the first formulation of the discriminant function, conceived in 1968, must be considered outdated, since the coefficients associated with it must be redefined in relation to the context to which they belong, such as industry, size, etc..

In 2007, an article, "A Review of Bankruptcy Prediction Studies: 1930 to Present"[5] highlights how, starting from Altman's research, the number and complexity of default prediction models rose significantly. In fact, 165 studies have been published on prediction models of failures, from the pioneer works of Beaver [2] and Altman [1] to the year 2004.

Only one research [6], in addition to Beaver and Altman works, had been published by the end of the 1960s. The numbers rose to 28 papers in the 1970s and to 53 studies in the 1980s; then we have 70 publications produced before the end of the 90s. Between 2000 and 2004, a further 11 studies were published. All the papers published in the 1970s were characterized by the prevalence of the application of Multiple Discriminant Analysis (MDA), which monopolized about a quarter of the literature of failure prediction techniques [7].

Since the end of the 1970s, some criticisms concerning the violation of the statistical hypotheses underlying the MDA approach have led researchers to focus their efforts on the development of conditional probability models with particular emphasis on log it and binary regression[8][9]. In the 1990s, as computing power increased, researchers had the opportunity to extend the number of techniques applied to insolvency prediction. Much of the work of the 1990s focuses on systems of Artificial Intelligence, Neural Networks (NN), Genetic Algorithms (GA), Case based reasoning (CBR), and Decision Tree (DT). The first application of the Neural Network system to the prediction of insolvency, ever published, is the one by Bell, Ribar, and Verchio [10].

At the same time, by the beginning of the 1970s, a different approach of a theoretical, non-statistical nature, evolved. It is based on a conceptual framework derived from the option pricing model developed by Black and Scholes as the Contingent Claim Analysis (CCA) theory [11], further developed by Robert Cox Merton [12]. The most famous version of this model is the one developed by Kealhofer, McQuown and Vasicek (KMV), belonging to a US company specialized in providing estimates of the probability of default of listed companies and, more recently, of private companies. The company was acquired by Moody's Corp. in April 2002.

In recent years a new methodological approach has been developed oriented towards the analysis of big data. This approach comes from the combination of the availability of a large amount of data, a large processing capacity and innovative analysis techniques.

The technique of neural networks is the most evident but there are also other machine learning methods: Support Vector Machine, Random Forests, Gradient Boosting. The first conceptual framework of Artificial Neural Networks can be traced back to the 1950s, but only in the last few years they have achieved great analytical skills thanks to the increase in computational power and the development of new specific software. In Italy, one of the first attempts to build a model for default prediction of companies, using neural networks, dates to 2007, when a research group [13] developed two models using feed forward neural networks.

Another interesting case of the application of neural networks is the one carried out by Jackson and Wood [14]. This research is of particular interest since a comparison was made between traditional statistical models and neural networks.

A different approach was presented by Zhang et al. [15].During a research conducted on a sample of 1000 companies of which 500 were in default, they selected 25 financial ratios for each of them, using Genetic Algorithm (GA) combined with the Ant Colony Algorithm (ACA). In their research, the fitting on the validation data exposed a classification error of the various configurations, ranging from a maximum of 8.9% of the GA to a minimum of 7.9 of GACA (a modified Genetic Algorithm in combination with Colony Ant).

In 2017, Barboza, Kimura and Altman [16] showed the results of a comparison between machine learning models (Support Vector Machines, Neural Networks, Bagging, Boosting and Random Forests) and traditional statistical ones, such as Discriminating Analysis and Logistic Regression, in order to predict bankruptcy one year prior to the event.

Recently, Le & Viviani [17] pointed out the superiority of the Machine Learning tools with respect to the traditional statistical approaches. They analysed a sample of 3000 US banks (in which 1438 were failed and 1562 were active banks) by two traditional statistical approaches (Discriminant Analysis and Logistic Regression) and three machine learning approaches (Artificial Neural Network, Support Vector Machines and K-Nearest Neighbours). Deep learning techniques, as the Convolutional Neural Networks (CNN)have only been reported in a small number of studies on the prediction of stock price movements to financial analyses, but in [18] a transposition of financial ratio values into images was processed through the deep learning architecture of GoogleNet, a CNN networks with 22 layers and about 4 million of parameters.

3.MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

The term machine learning is very general and includes many techniques used in various fields. Our problem is to do a supervised binary classification, that is we must estimate a model that can predict companies that will fail within a year. Among the machine learning techniques available, we consider two powerful ensemble methods: Random Forests, a method proposed by Breiman [19] such as the natural development of bagging and classification trees, and Gradient Boosting, proposed by Friedman [20] as a Boosting development. In addition to these methods we will also consider the well-known classical Logistic Regression model, that often proves to be an excellent classifier in the binary case.

Furthermore, we have considered Deep Neural Networks (DNN) [21], very popular machine learning algorithms in recent years, characterized by a large number of neurons organized in different hidden layers. DNN has been shown to overcome other techniques, in terms of accuracy and speed, for many applications. Today there are many variants of DNN, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) that allow to obtain results that were once unthinkable. These techniques are also the basis for the development of artificial intelligence that is becoming the driving force behind the greatest technological development of this century.

Roughly, a NN layer is nothing but a collection of neurons that accept an input and provide an output, where the inputs for each of these neurons are processed through the activation functions assigned to the neurons. The number and the structure of the hidden layers differ between different networks depending on the complexity of the problem to be solved.

Convolutional neural networks are a category of neural networks that has proven to be very effective in areas such as image recognition and classification. CNNs have been successful in identifying faces, objects and road signs as well as enhancing their vision in robots and self-driving cars.Successful CNNs were designed to classify objects within images, with many layers and millions of parameters without overfitting.

Recurrent neural networks are a recent variant of the neural networks strongly used in all areas in which current observations are strongly dependent on previous observations. Typical applications are, for example, in the analysis of time series and in the natural language processing, where it is essential to consider the sequence of the observed values.
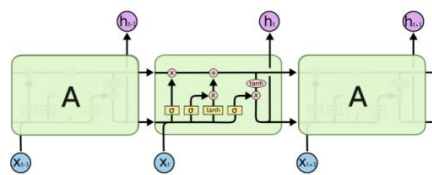

Figure 1. Scheme of a LSTM layer

Like RNN, LSTM (Long Short-Term Memory) layers kept a context of memory within their pipeline to allow for tackling sequential and temporal problems. LSTMs have been successfully used in a multitude of real-world problems, from text analysis to self-driving car technologies [21].

4. THE DATA-BASE: AN ITALIAN CASE-STUDY

To improve credit risk management, there is a lot of interest in bankruptcy predictive models. Academic research has mainly used traditional statistical techniques, but interest in the capability of machine learning methods is growing [22][23]. This Italian case study pursues the goal of developing a commercial firms insolvency prediction model, in compliance with the Basel II Accords.

The present research utilized the company's AIDA database - Bureau Van Dijk, a Moody's Group company. Data can be extracted in an indexed format through search keys and complex queries, then processed, evaluated and exported in multiple formats. From the database, information on the financial statements and financial ratios of the companies have been retrieved.

We selected 82 variables, as the most meaningful in terms of capacity of pointing out the critical issues related to a firm financial and economic equilibrium in the long run.

According to the literature, the chosen variables are closely related to gauge the liquidity, the profitability, the financial solidity and the operating performances, namely, the liquidity ratios, EBITDA, ROE, ROI and, among others, debt ratios. A company's liquidity is its ability to meet its short-term financial obligations. Liquidity ratios attempt to measure a company's ability to pay off its short-term debt obligations. This is done by comparing a company's most liquid assets, those that can be easily converted to cash, with its short-term liabilities.

Profitability ratios are a class of financial metrics that are used to assess a business's ability to generate earnings relative to its associated expenses.

Operating performance ratios are tools which measure the function of certain core operations for an organization or business. Particularly, these ratios reveal information about how efficiently that organization is using resources to generate sales and cash.

The debt ratio is a financial ratio that measures the extent of a company's leverage. The debt ratio is defined as the ratio of total debt to total assets, expressed as a decimal or percentage. It can be interpreted as the proportion of a company's assets that are financed by debt.

Overall, the group of variables selected was considered suitable to estimate the probability of insolvency of a bankrupt company, within one year of the available observations. Each record contains 82 variables, only one variable (Region) have categorical level, all the other variables are quantitative.

Table 1. Territorial distribution of the extracted original sample

ABRUZZO	2.796	SICILIA	6.456
CALABRIA	2.148	TOSCANA	11.125
CAMPANIA	10.400	TRENTINO ALTO ADIGE, SARDEGNA	5.236
EMILIA ROMAGNA	13.753	UMBRIA	2.351
FRIULI V.G.	3.460	VAL D'AOSTA, LIGURIA MOLISE, BASILICATA	5.061
LAZIO	12.621	VENETO	15.654
LOMBARDIA	26.570	PUGLIA	6.628
MARCHE	4.885	PIEMONTE	10.502
Total not failed companies	139.646	Total failed companies	1.579

After the pre-processing procedure, 14.966 Italian micro-small firms have been selected: 13,846 ongoing and 1,120 bankrupted. Some regions for which little data was available have been aggregated. The final list has been sorted out removing all records exposing missing data or negligible values. The volume of data processed places the research on a scale similar to that used by Moody's in the development of its rating model for public and private companies, Riskcalc^{TM}[24].

We have not used market valuations, in fact the companies considered are of medium-small size and not listed on the stock exchange. In this sense, we do not have these important assessment measures available, while are included in almost all the studies on this topic.

For each unit, the financial variables have been observed 2 times: 2 years and 1 year prior the observation of the target (Bankruptcy). Both ratios and values have been calculated or extracted from Balance Sheets.The extracted financial variables are shown in table 2.

5. BANKRUPTCY ANALYSIS BY MACHINE LEARNING AND DEEP LEARNING

Data pre-processing can be a very complex and time-consuming phase and the results of the analysis are very dependent on the effectiveness of this procedure. In this study we are mainly interested in comparing analytical models, so we did not perform a careful pre-processing, conducted variable by variable, to look for the optimal transformations. Anyway, it was performed a logarithmic transformation followed by a normalization of all the quantitative variables.

To create the test set, we split randomly the database obtaining one training data-set with 11,000 firms (832 bankruptcy) and one test data-set with 3,966 firms (288 bankruptcy).

For Random Forest and Gradient Boosting, to obtain better results, we applied also an oversampling on the training data to balance the classes of the target. The test-data were not oversampled in order to obtain a correct evaluation of the model on new data. Conversely, the NNs were not sensitive to oversampling.

Table 2. Extracted financial variables

Current Assets EUR_Year - 2	Return on Sales Year -2
Current Assets EUR_Year - 1	Return on Sales Year -1
Date of Last Available Balance Sheet	Return on Investiment Year -2
Bank Debt to Sales Year -2	Return on Investiment Year -1
Bank Debt to Sales Year -1	Gross Sales Year -2
Debt/EBITDA ratio Year -2	Gross Sales Year -1
Debt/EBITDA ratio Year -1	Gross working capital Turnover (times) Year -2
Total Debt to Equity Year -2 ratio	Gross working capital Turnover (times) Year -1
Total Debt to Equity Year -1 ratio	Invested Capital Turnover (times) Year -2
Receivables Average Collection Period -Year -2	Invested Capital Turnover (times) Year -1
Receivables Average Collection Period -Year -1	Operation Headquarters - Region
Payables Average Settlement Period -Year -2	Total Fixed Assets Year -2
Payables Average Settlement Period -Year -1	Total Fixed Assets Year -1
Ebidta to Interest Expenses Year -2	Total Liabilities and Equity Year -2
Ebidta to Interest Expenses Year -1	Total Liabilities and Equity Year -1
Working Capital to Revenues Year -2	Total Current Liabilities Year -2
Working Capital to Revenues Year -1	Total Current Liabilities Year -1
Current Ratio Year -2	Net Income after Taxes Year -2
Current Ratio Year -1	Net Income after Taxes Year -1
Total Fixed Tangible Assets to Equity Year -2	Profits/Losses Year -2
Total Fixed Tangible Assets to Equity Year -1	Profits/Losses Year -1
(Equity+Long Term Debts)/Fixed Assets Year -2	Equity to Total Assets Year -2
(Equity+Long Term Debts)/Fixed Assets Year -1	Equity to Total Assets Year -1
Current Debts to Total Debts Year -2	(Assets – Inventories)/Debts Year -2
Current Debts to Total Debts Year -1	(Assets – Inventories)/Debts Year -1
Long Term Debts to Total Debts Year -2	Long Term debt Year- 2
Long Term Debts to Total Debts Year -1	Long Term debt Year- 1
Interest Expenses to Gross Sales Year -2	Total Assets Year- 2
Interest Expenses to Gross Sales Year -1	Total Assets Year -1
Total Equity Year -2	EBITDA Year-2
Total Equity Year -1	EBITDA Year-1
Net Financial Position Year -2	EBITDA /Gross Sales Year -2
Net Financial Position Year -1	EBITDA /Gross Sales Year -1
Company Name	Total credits Year-2
Total Assets to Equity Year -2	Total credits Year-1
Total Assets to Equity Year -1	Total debt Year-2
Return on Equity Year -2	Total debt Year-1
Return on Equity Year -1	Number of employees Year-2
Return on Assets Year -2	Number of employees Year-1
Return on Assets Year -1	

To compare the different models applied, we compared the resulting binary classification on the test-set. Several well-known measures for binary targets were analysed: accuracy, sensitivity, precision, area under the curve ROC (AUC). In particular, the confusion matrix can give an idea of the classification result, given a threshold, and allows to calculate accuracy, sensitivity and precision.

The main defect of the confusion matrix, and of the connected measures, is that it depends strictly on the probability threshold value chosen to distinguish between the positive and negative class. Generally, the value 0.5 is chosen, but for a target with unbalanced classes this threshold value is not the best choice. Moreover, if the interest is to identify failure events (1), the threshold value should be chosen to maximize this objective.

The ROC curve illustrates the performance of a binary classifier as its threshold is varied. It plots the sensitivity vs. 1-specificity (false positive rate). This eliminate the problem of fixing the threshold. The Area Under the Curve ROC (AUC) is a more reliable measure of the performance of a model if we have a binary target and we are interested to individuate the 'event'.
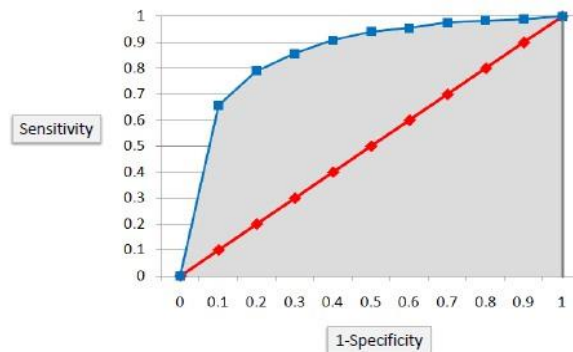

Figure 2. A typical ROC curve (blue)

To apply the CNN and the RNN it is necessary to modify the data structure to obtain three-dimensional matrices. In fact, these techniques expect that each unit corresponds to a data matrix. Actually, our data are obtained observing two years, so we have 40 variables for time -1 and 40 variables for time -2, plus two common variables. So, we can reshape each row of data as an array of dimensions (42, 2). The use of Convolutional Neural Networks and recurrent Neural Networks allows to consider, even if in a different way, the temporal relation existing between the observed data. Obviously, 2 times are insufficient to catch the temporal evolution and this approach would be greatly enriched by having more observation times available. These methods have proven to be very effective in various fields, such as the analysis of images, textual data, and repeated observations over time.

Table 3 – Characteristics of the models applied to our data

Random forests	500 estimators, min samples in a leaf=3, min samples split=5
Gradient Tree Boosting	500 estimators, max depth=3, subsample=0.7
Logistic Regression Model	weight of classes: inverse of frequency.
NN	4 hidden dense layers with 1024 neurons each, plus the normalization layers. A total of 3,248,130 parameters.
CNN	3 convolutional layers with 256 neurons, 3 dense layers with 1024 neurons,plus pooling and normalization layers. A total of 1,523,842 parameters.
CRNN	2 convolutional layers, 1 LSTM layer, 2 dense layers, plus pooling-regularization layers. A total of 288,962 parameters.

Table 4 – Architecture of our CRNN

Layer	Output Shape	Params
Conv1D	(2, 64)	5440
MaxPooling	(1,64)	0
Dropout	(1,64)	0
Batchnormalization	(1,64)	256
Conv1D	(1,128)	16512
MaxPooling	(1,128)	0
LSTM	(128)	131584
Batchnormalization	(128)	512
Dropout	(128)	0
Batchnormalization	(128)	512
Dense	(1024)	132096
Dropout	(1024)	0
Dense	(2)	2050

Total number of parameters: *288,962*

Table 5 summarizes all the results of the applied techniques and at the bottom also provides the elements fora comparison with the recent work of Barboza et al. [16]. The value of the AUC obtained by all the models,in our application, can be considered very good, even if Gradient Boosting, CRNN and Random Forest stand out. One reason for the good predictive ability of models is certainly due to the richness and quality of the data used.

Table 5 – Summary of the results on test data

Models	Precision threshold value 0.5	Sensitivity threshold value 0.5	AUC
Deep N.N.	0.83	0.88	0.975
Convolutional N.N.	0.87	0.85	0.972
Recurrent Conv. N.N.	0.86	0.85	0.980
Random Forest	0.85	0.84	0.979
Gradient Boosting	0.84	0.86	0.981
Logistic regression	0.79	0.82	0.973
Results of Barboza-Kimura-Altman (2017):			
Random forest	0.06	0.83	0.929
Neural Networks	0.03	0.93	0.900
MDA	0.01	0.64	0.636

Considering the comparison with the work of Barboza, obviously the data-sets of the two works are different, but in both cases table 5 provides the evaluation of the models on "new data"(the test-set). The differences are notable, especially in the value of the precision, despite the fact that the authors state that they used a balanced training data-set. However, the poor performance of Random Forests and Neural Networks obtained by [16] could be explained by the excessive simplicity of the models applied.

Barboza et al.[16] in their analysis considered a balanced training set with 898 firms, and the application of NN on the test-set (13,300 firms) gave the confusion matrix (pag. 413) of table 6.

Table 6. Confusion matrix of Barboza et al. [16].

Predicted by Barboza et al.

Observed		0	1
	0	9582	3585
	1	9	124

In this confusion matrix, we can note that only 3.3% of predicted bankruptcy are true! Similar results are obtained with the other models. It is evident that, considering a probability threshold 0.5, as in the previous confusion matrix, these models are useless. On the other hand, analysing the AUC, they obtained values between 0.90 and 0.92 for NN, Logistic Regression, Random Forests, which are much more interesting results and indicate that the previous threshold value 0.5 is not adequate. This is confirmed also by inspecting the ROC curve.

We reported the confusion matrices for 3 methods applied to our test-set (3,966 firms), considering a 0.5 probability threshold in table 7.

Table 7. Confusion matrices for three models on test data. Threshold 0.5.

Pred. by Gradient Boosting			Pred. by Random Forests			Pred. by CRNN		
	0	1		0	1		0	1
0	3631	47	0	3636	42	0	3638	40
1	40	248	1	47	241	1	43	245

From the table 7 you can see that all the models are very effective, with only minor differences. The ROC curves of these 3 models substantially coincide and in fig.3 we have reported the ROC curve of the CRNN model, which shows the excellent predictive capacity obtained. The AUC indices shown in table 5 show values around 0.98 for these models, which is an excellent result.
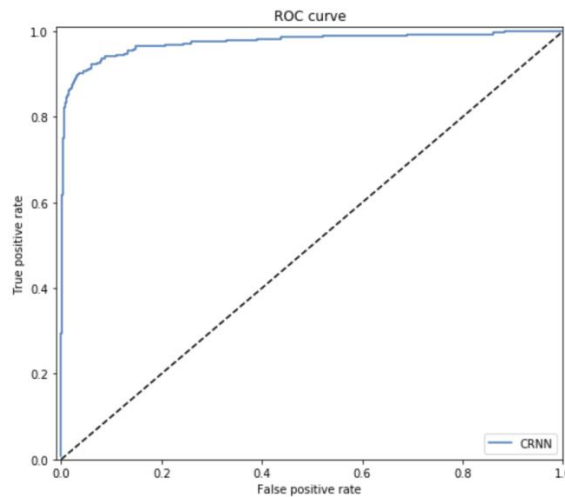

Figure 3. ROC curve for the model CRNN on the test data.

6. CONCLUSION

Academic literature contains a wide range of techniques that have been proposed over the past five decades to predict business insolvency. Among these models there are the most classic ones such as Multiple Discriminant Analysis and Logistic Regression, while among the most recent approaches we have machine learning techniques, such as Random Forests, Boosting and NN. Moreover, there are approaches that are less statistical, like the Contingent Claim analysis. In the literature, in recent years, advanced machine learning techniques, in particular the Deep Neural Networks, have been studied extensively. In our work, we have collected a large amount of data for small and medium-sized Italian companies taking information from the budgets, applying several machine learning and deep learning techniques, and using the logistic regression model as a benchmark.

The results obtained show that all models, including logistic regression, achieve good results, probably due to the wealth and good quality of the data. The model with the best performances was Gradient Boosting, which with a modest computational effort reached the highest AUC value. The neural networks showed a good performance, especially in the recurrent version, which took into account the two times of observation, but required a much higher computational effort. In general, the use of neural networks requires a lot of data, and the use of recurrent

networks also requires several successive moments of observation, characteristics that our data did not have. We are therefore confident that these new techniques will provide, in a future wider investigation, more interesting results.

It is worth noting that the results obtained in this paper show a predictive capacity of the applied methods higher than that of similar works in the literature, that generally use only listed companies. On the contrary, our approach is completely independent of market values and can be applied to small and medium-sized enterprises, as in the Italian productive structure.

REFERENCES

[1] FitzPatrick, P. J. (1932). A comparison of ratios of successful industrial enterprises with those of failed companies. The Certified Public Accountant, 598-605 (Oct), 656662 (Nov) , 727-731 (Dec).

[2] Beaver W. (1966). Financial ratios as predictors of failure. Journal of Accounting 5(5), 71-111.

[3] Altman E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate. The Journal of Finance 23(4), 589-609.

[4] Altman E. I. (2000). Predicting Financial Distress of Companies: Revisiting The Z-score and ZETA models. Stern School of Business, New York university, pp. 1-54.

[5] Jodi L., Gissel D.G. (2007). A Review of Bankruptcy Prediction Studies: 1930 to present. Journal of Financial Education, Vol. 33 (Winter 2007), 1-42.

[6] Daniel T. (1968). Discriminant: analysis for the prediction of business failures. PH.D. dissertation, University of Alabama.

[7] Richard H.G., Jackson, A.W. (2013). The performance of insolvency prediction and credit risk. The British Accounting Review, 45, 183-202.

[8] Ohlson J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research 18(1), 109-131.

[9] Zavgren C.V. (1985). Assessing the vulnerability to failure of American industrial firms: A logistic Analysis. Journal of Business Finance & Accounting 12(1), 19-45.

[10] Bell T.B., Ribar G.S., Verchio J.R. (1990), Neural nets vs. logistic regression: A comparison of each model's ability to predict commercial bank failures, in Auditing Symposium X: Proceedings of the 1990 Deloitte & Touche/University of Kansas Symposium on Auditing Problems, pp. 29-53.

[11] Black F., Scholes M. (1973). The pricing of options and corporate liabilities. Journal of Political Economy, 81.

[12] Merton R. C. (1974). On the pricing of corporate debt: the risk structure of interest rates. Journal of Finance, 29(2), 449–470.

[13] Angelini E., di Tollo G., Roli A. (2007). A neural network approach for credit risk evaluation. The Quarterly Review of Economics and Finance. Elsevier, v. 48(4), pp. 733-755.

[14] Jackson, R.H.G., Wood, A. (2013) The performance of insolvency prediction and credit risk models in the UK: A comparative study. The British Accounting Review 45, 183-202 (2013).

[15] Zhang Y., Wang S., Ji G. (2013). A Rule-Based Model for Bankruptcy Prediction Based on an Improved Genetic Ant Colony Algorithm, Vol. 2013, A. ID 753251. Hindawi Publishing Corp. - Mathematical Problems in Engineering.

[16] Barboza F., Kimura H., Altman E. (2017) Machine learning models and bankruptcy prediction. Expert System with Applications, 83, 405-417.

[17] Le H.H., Viviani J.L. (2018). An improvement by implementing a machine-learning approach to classical financial ratios. Research in International Business and Finance, 44 (2018) pp. 16-25

[18] Hosaka T. (2018). Bankruptcy prediction using image financial ratios and convolutional neural network. Expert Systems with Applications, Volume 117, 1 March 2019, pp. 287-299 .

[19] Breiman L. (2001). Random Forests, Machine Learning. 45 (1): 5–32.

[20] Friedman J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, Vol. 29, No. 5, pp. 1189-1232

[21] Goodfellow I., Bengio Y., Courville A. (2016). Deep Learning, series Adaptive Computation and Machine Learning, MIT press, ISBN-13: 978-0262035613

[22] Mai F., Tian S., Lee C., Ma L. (2018). Deep learning models for bankruptcy prediction using textual disclosures. European Journal of Operational Research (2018)

[23] Alaka H. (2017). Big data analytics for construction firms insolvency prediction models. Bristol: University of West of England.

[24] Falkenstein E.G., BoralA., LeaV.C. (2000). RISKCALC for private companies: MOODY's default model. Moody's.