Bankruptcy Prediction by Deep Learning

Giovanni Cialone

gcialone@stanford.edu

CS230 Winter 2020

Abstract

There is a lot of interest in bankruptcy predictive models. Academic research has mainly used traditional statistical analysis, but interest in 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, its major objective is an estimation of the probability of default over a certain time horizon, typically one year. The collected dataset consists of absolute values as well as financial ratios collected from the balance sheets of 14.965 Italian micro-small firms, 13,845 active and 1,120 bankrupted, with 88 observed variables. The volume of data processed places the research on the same scale of Moody's in the development of its rating model for public and private companies, RiskcalcTM. The study has been conducted using Deep Fully Connected and Convolutional Neural Networks. The results were compared for 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, with all the models showing relatively good performances.

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: it is well known that insolvency generates insolvency, producing a 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. The study was conducted first by gathering a wide dataset on small and medium-sized Italian companies. For each company, many balance sheet variables have been collected and then undergone pre-processing. Data were then analysed using some of the most advanced techniques in Machine Learning and Deep Learning, such as Fully Connected Neural Networks and Convolutional Neural Networks (CNNs). Subsequently the performance of these models have been compared on a real data test-set. The first part of the paper describes the state, analysing the proposals in the literature on this topic, followed by a brief description of the techniques of deep learning applied and an illustration of the dataset created for this application. Finally, the conclusions contain final considerations, takeaways 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. But it was only in 1966, Beaver [2] analysed a dataset of 13 balance sheet ratios for 38 companies (19 failed and 19 active). 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 focus, from this point on, has been 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] up to 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; after that, 70 publications have been produced by the end of the 90s. Between 2000 and 2004, a further 11 studies have been 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 logit and binary regression [8][9]. In the 1990s, as computing power increased, researchers had the opportunity to extend the number of techniques applied to the 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, nonstatistical 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 default probability 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 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.

Neural networks techniques are the most widely used but there are also other machine learning methods: Support Vector Machine SVM, 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 back to 2007, when a research group [13] developed two models using Feedforward 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 percent 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 and Viviani [17] pointed out the superiority of the Machine Learning tools over the traditional statistical approaches. The researchers analysed a sample of 3000 US banks (0f which 143 respect8 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. 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 IIAccords.

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. I selected 88 variables (inputs), 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 term. According to the literature, the chosen variables are closely related to gauge liquidity, profitability, financial solidity and operating performances, namely, the liquidity ratios, EBITDA, ROE, ROI and, among others, debt ratios. A company's liquidity is its ability to meetits short-term financial obligations. Liquidity ratios attempt to measure a company's ability to pay them. 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 how well certain core operations function within 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 t hat 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 most recent observations. Each record contains 88 variables, all of them being quantitative.

After pre-processing, 14.966 Italian micro-small firms have been selected: 13,846 active and 1,120 bankrupted. Regions with little data available have been aggregated. The final list has been cleaned for 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, RiskcalcTM[24]. Companies have been selected on the basis of market valuations, in fact they are of medium-small size and not listed on the stock exchange. In this sense, the project hasn't been able to rely on have these important assessment measures, which are included in almost all the studies on this topic. For each unit, the financial variables have been observed two times: two years and one year prior the observation of the target (Bankruptcy). Both ratios and values have been calculated or extracted from Balance Sheets.

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

Table 1. Extracted Samples

4. BANKRUPTCY ANALYSIS BY 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. This study's major concerns are related to the issue of comparing analytical models, so a careful pre-processing, conducted variable by variable, to look for the optimal transformations has not been performed.

To create the test set, the original dataset has been randomly split at a 30 to 70 ratio, obtaining one training data-set of 9,692 active and 785 failed companies respectively and one test data-set with 4,488 firms (335 bankrupted).

At the training time, due to the strongly imbalanced ratio between failed and not failed samples, 785 to 9692, data augmentation has been necessary. There are two popular methods to over-sample minority classes: (i) the Synthetic Minority Oversampling Technique (SMOTE) [26] and (ii) the Adaptive Synthetic (ADASYN) [25] sampling method. In this case study, SMOTE has been applied, oversampling training data to augment the minority class, the actual failed samples labelled at 1. After the procedure the two classes are balanced at a ratio of 9692 to 9691, i.e., the new training set totals 19.383 samples. The test-data were not over-sampled in order to obtain a correct evaluation of the model on new data.

To measure the different models applied, a comparison matrix has been created to compare the resulting binary classification on the test-set. Several well-known metrics for binary targets have been analysed: accuracy, sensitivity, specificity, max and average accuracy during test and 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 specifity.

A major issue in the use of the confusion matrix and its related measures can arise from its critical reliance on the chosen threshold probability value in order to select positive and negative class.

Tipically, the value is set at 0.5, 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 shows the performance of a binary classifier and its threshold's changes. It plots the sensitivity vs. 1-specificity (false positive rate). This rules out the problem of setting a threshold. The Area Under the Curve ROC (AUC) is a more reliable measure of a model performance as long as the target is binary and the objective is to select a relatively rare 'event'. That being said, in order to choose the "best" model to the purpose of the present study, AUC has to be considered as the optimal metrics.

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	

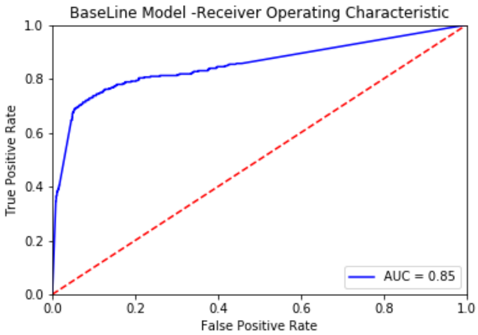
True Positives	242	275	243
False Positives	93	60	92
Sensibilty	0.7224	0.8205	0.7254
Specificity	0.9116	0.8148	0.9162
Accuracy	0.8975	0.8148	0.9018
Avg.Loss	5.5021	0.3063	1.9490
AUC	0.85	0.90	0.88
Max Accuracy	0.9496	0.9424	0.9510
Avg Accuracy	0.8708	0.8794	0.9010

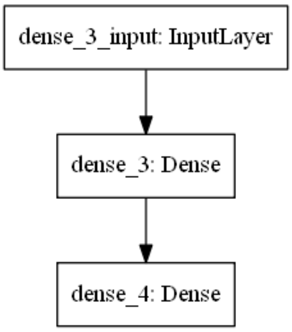
Table 3. Comparison Matrix

5.THE MODELS

The Baseline

A simple sequential fully connected model has been used as benchmark. The shape of the input layer is (88 x 19,383) Its architecture is the following:





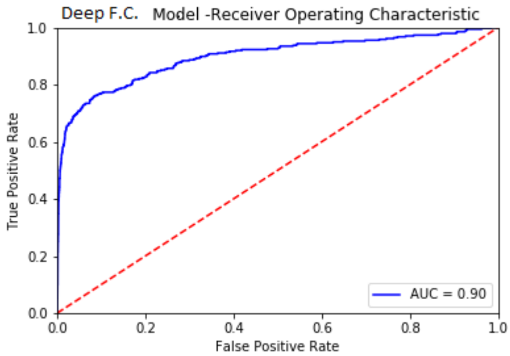

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_3 (Dense)              (None, 512)               45568
_________________________________________________________________
dense_4 (Dense)              (None, 1)                 513
=================================================================
Total params: 46,081
Trainable params: 46,081
Non-trainable params: 0
_________________________________________________________________
```

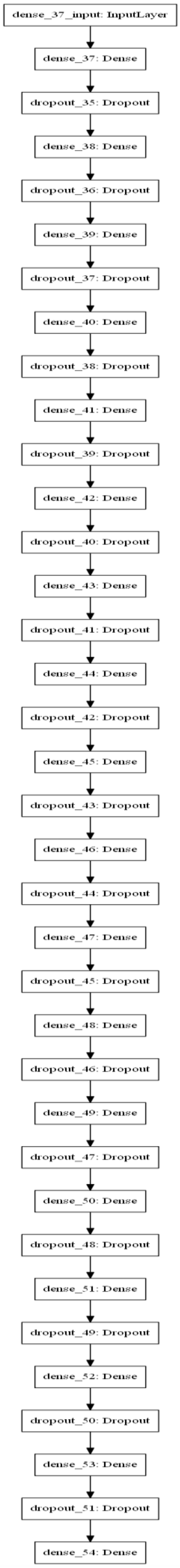
:

The Deep Sequential Model

The same input dataset has fed the deep sequential model. The architecture is much more complex than the Baseline Model: 01 input layer with shape (88, 19383), 17 inner layers, 512 neurons each and 1 output layer. This model totals 262,656 parameters of each of the layers from 2 to 17 and 45,568 parameters for the inner layer 1. Adding 513 parameters of the output layer, the final number of trainable parameters is 4,248,577. At the training time, a batchsize of 64 and a dropout of 0.1 have been set. All inner layer activation functions are "RELU". The output layer activation function for binary classification purpose is "Sigmoid",the optimizer "Adam".

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_55 (Dense)             (None, 512)               45568
_________________________________________________________________
dropout_52 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_56 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_53 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_57 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_54 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_58 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_55 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_59 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_56 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_60 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_57 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_61 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_58 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_62 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_59 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_63 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_60 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_64 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_61 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_65 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_62 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_66 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_63 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_67 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_64 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_68 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_65 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_69 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_66 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_70 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_67 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_71 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_68 (Dropout)         (None, 512)               0
_________________________________________________________________
dense_72 (Dense)             (None, 1)                 513
=================================================================
Total params: 4,248,577
Trainable params: 4,248,577
Non-trainable params: 0
_________________________________________________________________
```




dense_37_input: InputLayer
dense_37: Dense
dropout_35: Dropout
dense_38: Dense
dropout_36: Dropout
dense_39: Dense
dropout_37: Dropout
dense_40: Dense
dropout_38: Dropout
dense_41: Dense
dropout_39: Dropout
dense_42: Dense
dropout_40: Dropout
dense_43: Dense
dropout_41: Dropout
dense_44: Dense
dropout_42: Dropout
dense_45: Dense
dropout_43: Dropout
dense_46: Dense
dropout_44: Dropout
dense_47: Dense
dropout_45: Dropout
dense_48: Dense
dropout_46: Dropout
dense_49: Dense
dropout_47: Dropout
dense_50: Dense
dropout_48: Dropout
dense_51: Dense
dropout_49: Dropout
dense_52: Dense
dropout_50: Dropout
dense_53: Dense
dropout_51: Dropout
dense_54: Dense

The CNNs Model

To apply the CNN it is necessary to modify the data structure to obtain a 3D matrix. In fact, these expect that each unit corresponds to a data matrix. Actually, the data are obtained observing two years, so we have in total 88 variables, 44 for time -1 and 44 variables for time -2. So, each row of the original training set (88, 19385) has been reshaped in an array of dimensions ((7, 7) 2) for 19385 samples. To implement a 3D matrix, it's necessary to augment the number of the variables in order to increase them from 88 to 98. To do so, 10 additional zeros have been added to each layer. Zeros are neutral values and do not affect the processing. In this way, a 3D matrix with 2 square channels (Year -1 and Year-2) of dimension (7, 7) have built. 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. Concerning the model, the model summary gives an idea of the architecture.

Model: "CNNs"

Layer (type)	Output Shape	Param #
conv2d_2 (Conv2D)	(None, 2, 7, 32)	928
activation_2 (Activation)	(None, 2, 7, 32)	0
conv2d_3 (Conv2D)	(None, 1, 6, 32)	4128
activation_3 (Activation)	(None, 1, 6, 32)	0
max_pooling2d_1 (MaxPooling2	(None, 1, 3, 32)	0
dropout_1 (Dropout)	(None, 1, 3, 32)	0
conv2d_4 (Conv2D)	(None, 1, 3, 64)	18496
activation_4 (Activation)	(None, 1, 3, 64)	0
conv2d_5 (Conv2D)	(None, 1, 2, 64)	8256
activation_5 (Activation)	(None, 1, 2, 64)	0
max_pooling2d_2 (MaxPooling2	(None, 1, 1, 64)	0
dropout_2 (Dropout)	(None, 1, 1, 64)	0

flatten_2 (Flatten) (None, 64) 0

____

dense_2 (Dense) (None, 1) 65

____

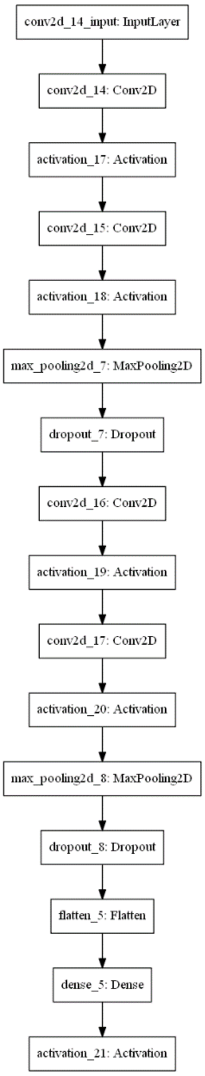
activation_6 (Activation) (None, 1) 0

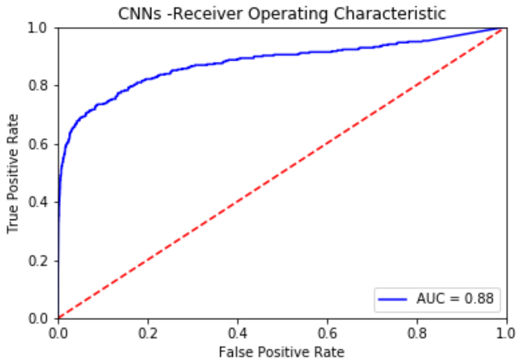
===

Total params: 31,873

Trainable params: 31,873

Non-trainable params: 0





6. CONCLUSION AND TAKEAWAYS

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 this paper, the large amount of data for small and medium-sized Italian companies collected from financial and income statements have been processed , applying two different Neural Networks architectures: (i) **a deep sequential model** and **(ii) a Convolutional architecture**, using a simple a very simple sequential one as a benchmark. The results obtained show that all models, including the baseline, achieve good results, probably due to the good quality of the data. The model with the best performances was the Sequential Architecture which reached the highest AUC value, 0.90 and the highest sensibility 0.8205. The CNN Architecture showed the best specificity (numbers of True Negatives caputered).

So, It's very likely that these architectures 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, this approach is completely independent of market values and can be applied to small and medium-sized enterprises. Ultimately, the models can find wider application, not only to the italian case but also to other countries where accounting standards are similar and the inputs variables have same metrics.

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., Boral A., Lea V.C. (2000). RISKCALC for private companies: MOODY's default model. Moody's.

[25] Haibo He et al. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning

[26] Chawla n. et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique