UNIVERSITÀ
CATTOLICA
del Sacro Cuore

DIPARTIMENTO DI DISCIPLINE MATEMATICHE,
FINANZA MATEMATICA ED ECONOMETRIA

WORKING PAPER N. 19/3

Machine Learning models for bankruptcy prediction in Italy: do industrial variables count?

Daniela Bragoli, Camilla Ferretti,
Piero Ganugi, Giovanni Marseguerra,
Davide Mezzogori, Francesco Zammori

VP VITA E PENSIERO

Università Cattolica del Sacro Cuore

DIPARTIMENTO DI DISCIPLINE MATEMATICHE,
FINANZA MATEMATICA ED ECONOMETRIA

WORKING PAPER N. 19/3

Machine Learning models for bankruptcy prediction in Italy: do industrial variables count?

Daniela Bragoli, Camilla Ferretti,
Piero Ganugi, Giovanni Marseguerra,
Davide Mezzogori, Francesco Zammori

VP VITA E PENSIERO

Daniela Bragoli, Department of Mathematics for Economic, Financial and Actuarial Sciences, Università Cattolica del Sacro Cuore, Milano (Italy).

Camilla Ferretti, Department of Economic and Social Sciences, Università Cattolica del Sacro Cuore, Piacenza (Italy).

Piero Ganugi, Department of Engineering and Architecture, Università di Parma, Parma (Italy).

Giovanni Marseguerra, Department of Mathematics for Economic, Financial and Actuarial Sciences, Università Cattolica del Sacro Cuore, Milano (Italy).

Davide Mezzogori, Department of Engineering and Architecture, Università di Parma, Parma (Italy).

Francesco Zammori, Department of Engineering and Architecture, Università di Parma, Parma (Italy).

✉ daniela.bragoli@unicatt.it

✉ camilla.ferretti@unicatt.it

✉ piero.ganugi@unipr.it

✉ giovanni.marseguerra@unicatt.it

✉ davide.mezzogori@unipr.it

✉ francesco.zammori@unipr.it

www.vitaepensiero.it

ISBN 978-88-343-4111-7

Abstract

We aim to provide a predictive model, specifically designed for the Italian economy, which classifies solvent and insolvent firms one year in advance, using AIDA Bureau van Dijk dataset from 2007 to 2015. We apply a full battery of bankruptcy forecasting models, including both traditional and more sophisticated machine learning techniques, and add to the financial ratios used in the literature a set of industrial/regional variables. We find that XGBoost is the best performer and that industrial/regional variables are important. Moreover, belonging to a district, having a high mark up and a greater market share diminish bankruptcy probability.

Keywords: Firm distress analysis, machine learning, logistic regression, industrial variables.

JEL: G33, C45, C52, R11, L23.

1 Introduction

Research on financial distress prediction is relevant not only to lending institutions, both in deciding whether to grant a loan and in devising policies to monitor existing ones, but also to investors, regulatory authorities, managers and so on. Studying the determinants of firm bankruptcy then becomes of vital importance, not only from an economic point of view - the failure of firms represents a cost for employees, entrepreneurs, creditors and for the whole society - but also from a policy perspective.
A first stream of contributions, led by the seminal papers by Altman [1968] and Foster [1986], focuses on critical financial ratios that can help entrepreneurs and funders predict insolvency.
While the financial nature of default events clearly suggests to primarily look for financial causes, the probability to stay in the market, as well as the financial stability of a firm, is deeply interconnected with the ability to perform well along the economic or industrial aspects of its operation. Thus, it is likely that looking exclusively at financial indicators cannot offer but a partial account of the main determinants of default. Related to this point a second stream of the literature aims to determine the causes of firm bankruptcy by looking at variables beyond those that come from accounting books, e.g. productivity, profitability and growth [Bottazzi et al., 2011], size and age [Mueller and Stegmaier, 2015], corporate governance indicators [Liang et al., 2016] and the institutional framework in which the firm operates [Eklund et al., 2018].
Other studies propose methodologies and tools to improve firm bankruptcy prediction models. Balcaen and Ooghe [2006] have highlighted the problems related to the classic statistical methodologies for bankruptcy prediction, whereas more recently Barboza et al. [2017] have compared statistical models (e.g. logistic regression) with machine learning techniques, whereas Zhao

et al. [2017] have analysed the discriminatory power of the features (predictive variables) related to bankruptcy prediction.[1]
The aim of this paper is to bring together these two streams of the literature providing a new bankruptcy model for the Italian economy, which considers jointly financial ratios and more structural/industrial variables with a special focus on regional aspects. For this purpose, we apply a full battery of bankruptcy forecasting models, which combine more traditional models, such as logistic regression, with more sophisticated techniques based on machine learning, focusing on Aida Bureau van Dijk balance sheet information on manufacturing Italian firms from 2007 to 2015.
Our aim is not only to select the most accurate forecasting model, but to shed some light on what variables are important predictors and whether the industrial/regional indicators should be considered by credit institutions to assess the financial vulnerability of firms.
Our results show that indeed industrial variables and regional indicators have a significant impact on the probability of bankruptcy. In particular, belonging to an industrial district, having a high mark-up and a high market share diminish the probability of bankruptcy. Moreover, the XGBoost technique is the best performer in terms of predictive ability and outperforms the Logistic Regression and also the other Machine Learning models. The XGBoost has also the advantage of being able to exploit, better than all the other models considered, the information on the industrial/regional indicators given that the bankruptcy prediction error reduces from 12.31% to 10.72% when augmenting the model with these additional variables.
The rest of the paper is structured as follows. In Section 2 we present the literature on predictive variables focusing on indus-

[1]See [Kumar and Ravi, 2007] for a survey on bankruptcy prediction in banks and firms via statistical and intelligent techniques.

trial/regional indicators. Section 3 describes the data, Section 4 the methods and Section 5 the evaluation exercise. Section 6 highlights the main results and Section 7 concludes.

2 The literature on predictive variables

The list of variables with which to feed the model is crucial in a firm bankruptcy forecasting exercise. The literature has started to focus primarily on financial ratios. The seminal work by Altman [1968] identified a set of financial ratios that were the first under consideration by many researchers and subsequently used in later studies which eventually proposed a very large number of ratios. Courtis [1978], for example, has identified 79 financial ratios that were grouped in three main categories: 1) profitability, 2) managerial performance, 3) solvency ratios.

The performance and survival of firms though might be influenced by several factors external to the firm, i.e. the environment, national and international economic conditions. Mensah [1984] noted that different economic environments as well as different sectors lead to different models for the prediction of failures.

Other studies explore the possibility that firms' performance might be influenced not only by financial ratios, but also by qualitative variables, i.e. quality of management, research and development, market trend [Zopounidis, 1987], the social importance of the firm, and the strength of its bank relationship [Suzuki and Wright, 1985].

Judging from the dates of these contributions, the idea of expanding the initial set of financial ratios is not new to the literature. However, there have been far more recent contributions with the aim of augmenting the financial ratios with other groups of variables and showing the importance of these new variables in increasing the forecasting performance of the model.

For example, Bottazzi et al. [2011] focus on productivity, profitability and growth as additional variables, Mueller and Stegmaier [2015] select size and age, Liang et al. [2016] favour corporate governance indicators, and finally Eklund et al. [2018] introduce the institutional framework.
To our knowledge, we are the first to add industrial/regional indicators to the financial ratios *à* la Altman. We start considering the following financial ratios as in Barboza et al. [2017]: Net Working Capital/Total Assets, Earnings before interest and taxes/Total Assets, Net Worth/Total Debt, Total Sales/Total assets, Earnings before interest and taxes/Total Sales, growth rates of Total Assets, growth rates of Total Sales, Return on Equity$_t$-Return on Equity$_{t-1}$.[2]
To the financial ratios we add the following industrial/regional variables: sectors (food products, textiles, leather and related, wood products and cork, glass, metal and machinery and equipment), regional dummies (North East, North West, Centre and South), whether the firm belongs to an industrial district (dummy 0/1 variable), a non-parametric measure of market power,[3] defined as (Total Sales/(Labour Cost + Nominal Materials)) - 1 and a measure of the firm market share (firm value added over sector value added).
These variables could be potentially relevant for any country, but are even more important in the Italian context. Italy is characterized by a prevalence of non-listed manufacturing SMEs, Industrial Districts (ID) represent around one fourth of the Italian productive system, in particular 24.4% of firms belong to ID and

[2]Compared to Barboza et al. [2017] we only consider indicators for which we have available and reliable information, we do not consider some variables typical of listed firms, which are few in number in the Italian economy.

[3]See the European Central Bank Competitiveness Research Network (CompNet) study on different ways to calculate mark-up measures. See https://www.ecb.europa.eu/home/pdf/research/compnet/CompNet-database-user_guide-round4.pdf

24.5% of employees are employed in ID.[4] Italy is also characterized by a regional divergence between North and South.
In the next Section we will argue why ID membership and firms' mark-up should be important predictors for firms' bankruptcy.

2.1 District membership

In a world of dramatically improved communications systems and corporations that are increasingly mobile internationally, it is puzzling why certain places are able to sustain their attractiveness to both capital and labor. ID are a successful example of such phenomena thanks to the role of small, innovative firms, embedded within a regionally cooperative system of industrial governance, which enables them to adapt and flourish despite globalizing tendencies.
The first contribution on the concept of ID dates back to Marshall [1890], who defines the localization of industry as a 'concentration of many small businesses of a similar character in particular localities'. The disadvantage of the small scale is compensated by the localization externalities that firms belonging to a district enjoy. The key idea is that firms located close to other firms operating in the same industry, benefit from reduced transportation costs, availability of specialized workers and suppliers, and diffusion of intra-industry knowledge and technological spillovers. According to the literature on ID [Marshall, 1890, Hart, 2009, Bellandi, 2009] these factors enable small firms localized in the same industrial area to benefit of the same economies (external-scale economies) present inside large firms (internal-scale economies).
The Italian revisiting of the Marshallian ID concept introduced by Becattini [1990], Brusco [1982], Sforzi [1989] highlights more the role of cooperation and the link between social and economic

[4]See ISTAT website https://www.istat.it/it/archivio/150320.

forces that interact within the same geographical area. Trust among district members is central to their ability to cooperate and act collectively.
Alongside this new theoretical definition of ID, a new body of empirical literature emerged. These works attempt to establish the presence of a 'district effect', i.e. they try to identify empirically the agglomerative benefits that firms derive from membership. Signorini [1994] and other research in this field show unanimously that firms in ID do indeed benefit from agglomeration advantages.
Another very vast and more recent stream of the literature focuses on the impact on economic growth (in terms of employment and productivity) of three different types of local externalities: localization economies, Jacob's externalities and urbanization economies.
These studies are spanned over time starting from the '90 [Glaeser et al., 1992, Henderson et al., 1995] cover different countries [De Lucio et al., 2002, Cingano and Schivardi, 2004, Martin et al., 2011], but are rather not unanimous in their conclusion. More recent contributions have focused on the role of agglomeration in fostering innovation productivity and export [Boschma and Iammarino, 2009, Antonietti and Cainelli, 2011].
While these papers all refer to the long-run effects of agglomeration on growth and productivity, short-run effects are less studied. However, an interesting stream of the literature emphasizes the benefits of agglomeration economies over the business cycle with a particular attention to recessions [Guiso and Schivardi, 2007, Brunello and Langella, 2016].
Why should firms in ID behave differently from other companies during recessions? According to this literature the social interactions among entrepreneurs, which characterize ID of the marshallian type, and social capital, which also affects localization economies through mutual trust and cooperation, are the

drivers for making ID firms different during recessions.
According to Guiso and Schivardi [2007] the intense social interactions within the ID are likely to amplify the responses to negative shocks acting as a social multiplier. A similar result is found by Brunello and Langella [2016] who investigate the impact of agglomeration economies on firm entry during recessions and show that firm entry in ID has declined more during recession than in comparable areas. On the other hand, social capital[5], which is found to be highly present in ID [Trigilia, 2001, Soubeyran and Weber, 2002], might increase the trust among firms and between firms and other institutions in the territory.
This could for example translate into a better access to credit through relationship lending. The level of trust between district firms and local banks, that share the same territory, might be crucial in the process of credit supply, given that banks sharing the same territory evaluate firms solvability not only implementing a credit scoring approach, but also accounting for the entire background of 'soft' and not codified information, which is crucial to fully characterize firms belonging to ID [Alessandrini and Zazzaro, 2009]. A higher trust towards district firms could translate into a higher availability of credit that will in turn promote investments and innovation. Though the empirical literature on the role of ID membership on bankruptcy is missing, we have tried to hint at some possible theoretical explanations, related to localization externalities and social capital, that could be potential drivers for reducing the probability of ID firms of exiting the markets.

[5]i.e. the set of norms and values that creates the fabric of society glues individuals and institutions together and constitutes a necessary link for its governance.

2.2 Mark up

As a measure of mark-up we use the Price Cost Margin (PCM) defined in Section 2. This indicator is related to the notion of firm profitability, which has been widely considered in the past literature on bankruptcy models.
The reason we introduce this variable in the augmented specification is twofold. The first is related to the fact that the PCM, differently from more traditional indicators of profitability such as Return on Equity (RoE) and EbIT (Earnings before Interests and Taxes), measures the profits related to the core business of the firm, whereas the other two variables comprise both the core business, but also the financial and accessory activities.
The second is related to the fact that the PCM, quantifying the mark-up that firms are able to extract from customers, identifies the market power of a firm. An important theoretical feature of this measure is that the higher the market competition, the smaller should be the PCM. In fact, in absence of barriers to entry, prices should be equal to the marginal costs. A positive and persistent PCM typically suggests that firms have at least a certain degree of market power. Having a high mark-up thus implies greater profits generated by the core business and higher market power.
The role of market power is not new to the literature. From a social welfare perspective, most of the literature has been in favour of the benefits of marginal cost pricing. Structural reforms and deregulation as a mean of lowering entry barriers, are perennial topics for macroeconomic policy around the world.
However, as Dixit and Stiglitz [1977] point out there is a trade-off between quantity and variety. With scale economies, resources can be saved by producing fewer goods and larger quantities of each, however this leaves less variety, which entails some welfare loss. Bilbiie et al. [2008], show that the welfare impact of deregulation or 'more competition' fluctuates over the business cycle

along with the consumer's taste for variety and firms' profit incentive for entry.
From the firm perspective, a high mark-up, on one hand, might be related to higher profits and thus more financial resources to increase investments and innovative activities that could reduce production costs [Cassiman and Vanormelingen, 2013].[6]
On the other hand a high mark-up might also mean more product diversification (variety) and higher barriers of entry for external firms. These two factors could be potentially important drivers to reduce the firm's probability of going bankrupt.

3 Data

The analysis is based on balance sheet information on manufacturing Italian firms extracted from AIDA Bureau van Dijk, from 2007 to 2015, which allows to compute the response variable and all the selected covariates with the exception of the ID variable. The latter is obtained merging, through the ZIP code of the firm's operative branch, AIDA with the Industrial District Database provided by the Italian National Statistical Institute (ISTAT).
We construct our response variable based on the AIDA field 'status', i.e. we create a dummy variable which takes the value of 1 if the status is 'bankruptcy', and 0 otherwise. For brevity, from here on bankrupt firms will be identified as B (Bankrupt), while sound firms will be referred to as NB (Not Bankrupt).
We clean the data to exclude missing observations, inconsistencies or extreme values. Regarding B companies, as in Barboza et al. [2017], we only consider the balance sheet in the year before

[6] A part of the literature, differently from this view highlights the inefficiencies stemming from high market power, i.e. when industries are able to charge relatively high prices and benefit from large rents, they might have fewer incentives to improve their efficiency [Cette et al., 2016].

the bankruptcy event, whereas for NB companies, we check all available years. It is worth noting that for NB, since we do not use panel methodologies, balance sheets of the same company, corresponding to different years, are considered as different statistical units in the final dataset.

Given the high imbalance ratio (15%) of B over NB, aggravated by the fact that NB are counted as nine observations in our dataset, we follow a mixed strategy, both downsizing NB and using class weighted loss functions.

Specifically, we downsize the number of NB observations, keeping only three different balance sheets for each NB firm, equally spaced in time. A third of the firms are associated to years 2007, 2010, 2013, a third to years 2008, 2011, 2014, and the last third to years 2009, 2012, 2015.

The final dataset consists of 4,774 B and of 22,359 NB considered in three equispaced years (67,077 NB observations), for a total of n =71,851 balance sheets, with an imbalance ratio of 7.12%.

As already stated, to further adjust such ratio we use a 'class weighted loss function' to perform the classification. Following King and Zeng [2001], we assign different weights to B and NB observations, defined as $w_i = \frac{n}{2n_i}$ with $i = B, NB$ and n_i =number of observations in the corresponding class. In our analysis we obtain $w_B = 7.52$ and $w_{NB} = 0.53$.[7]

In the Appendix (Table A1), we report the summary statistics of the variables considered for B and NB.

[7]Differently from the previous literature, which considers datasets equally balanced between B and NB, we decide to keep the imbalance in the data. The 'class weighted loss function' has the advantage of avoiding the loss in information due to the downsizing of NB firms, which are typically more numerous than B. We feel that this methodological refinement represents an important novelty in the field of bankruptcy prediction. For further details, the reader is referred to the following section.

4 Models

4.1 Logistic Regression (LR)

Earlier studies in credit risk modelling employed Univariate and then Multivariate Discriminant Analysis with the purpose of developing bankruptcy prediction models [Beaver, 1966, Altman, 1968]. Starting from the 80's, Logistic regression (LR) has been considered a popular alternative to multivariate analysis for credit risk modelling [Ohlson, 1980].
Here we resume LR to have a benchmark for comparing the more sophisticated techniques we will present in the next sections. In addition, LR permits to evaluate the significance of the explanatory variables and the sign of their coefficients, allowing us to give an economic intuition of some important determinants in bankruptcy prediction.
As it is well known, given a binary variable Y distributed as a Bernoulli with parameter π, through LR we suppose that π depends on personal covariates, and we estimate $P(Y_i = 1|X_i = x_i)$, where $x_i = (x_{i1}, \ldots, x_{ip})$ is the vector of explanatory variables observed for the i-th firm, $i = 1, \ldots, n$. As before, we aim to predict the dummy variable $Y = 1$ if bankruptcy occurs, 0 otherwise and we use the logit model, where the bankruptcy probability π_i depends upon the covariates through the following link function:

$$\pi_i = P(Y_i = 1|X_i = x_i) = \frac{exp(x_i \cdot \beta)}{[1 + exp(x_i \cdot \beta)]}$$

in which $x_i \cdot \beta = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$, and $\beta_0, \ldots, \beta_p$ are $p+1$ parameters to be estimated.
The LR model is estimated using Maximum Likelihood, and the resulting coefficients are associated with a test of significance. It is worth noting that the log-likelihood function is a sum of n terms, each one corresponding to a statistical unit, and consequently it can be split into two parts, corresponding respectively

to firms observed to have $y_i = 1$ and $y_i = 0$ as follows:

$$L = \sum [y_i \cdot log(\pi_i) + (1 - y_i) \cdot log(1 - \pi_i)] = \sum_{y_i=1} log(\pi_i) + \sum_{y_i=0} log(1 - \pi_i) = L_1 + L_0$$

.

If positive events (number of observed $y_i = 1$) are rare in the sample under study, as in our exercise, the estimated probabilities π_i tend to be too small and biased, together with the related standard errors which depend on $\pi_i \cdot (1 - \pi_i)$. To account for this bias, we exploit the aforementioned method proposed in King and Zeng [2001], i.e. in order to consider the imbalance ratio, we estimate the parameters maximizing the modified log-likelihood function $L_w = w_1 \cdot L_1 + w_0 \cdot L_0$, where $w_1 = w_B = 7.52$ and $w_0 = w_{NB} = 0.53$. In this light we will refer to this methodology as a Weighted Logistic Regression (WLR).

4.2 Machine Learning (ML)

Aiming to compare the WLR with some state-of-the art machine learning (ML) techniques, we also perform the classification tasks using Neural Networks, Random Forest and XGBoost methods. All these methods are implemented in Python, with Keras and Scikit-learn packages, and their hyperparameters are fine-tuned using cross-validation. The implementation details of each technique are described next; a more comprehensive explanation can be found in Appendix B.

Neural Networks (NN) are one of the most widespread artificial intelligence methods, widely used for regression, patter recognition and data analysis [LeCun et al., 2015].
For every $i = 1, \ldots, n$ the vector of observed covariates x_i is fed as input in the NN algorithm, and elaborated through a sequence of steps ('layers') formed by many 'neurons'. Every neuron j in a layer firstly computes the weighted sum s_j of the inputs furnished by all the neurons in the preceding layer, and then produces its

own output calculating the 'activating function' $f(s_j)$. Such outputs are in turn fed as inputs for the neurons in the following layer, and so on. Weights for the weighted sums are the parameters to be trained (see Appendix B).
In this exercise we use a fully connected feedforward NN made of 3 hidden layers, with 16 neurons each, based on the 'relu' activation function $f(s_j) = max(0, s_j)$. See Glorot et al. [2011].
As it is customary in classification problems, the last layer has a single neuron that generates the response value $\hat{y}_i$ (in our case: the probability for the i-th firm to be bankrupted) using the standard logistic function as activating function.
Generally, weights are estimated minimizing a given loss function, based on the difference between observed and estimated classification for the units in the training set. To consider the imbalance ratio we use the weighted binary cross-entropy loss function as follows:

$$-\frac{w_B}{n_B}\sum_{y_i=1} L(y_i, \hat{y}_i) - \frac{w_{NB}}{n_{NB}}\sum_{y_i=0} L(y_i, \hat{y}_i),$$

where w_i and n_i, $i = B, NB$ have been previously defined, and $L(y_i, \hat{y}_i) = y_i \cdot log(\hat{y}_i) + (1 - y_i) \cdot log(1 - \hat{y}_i)$.

Random Forests (RF), introduced by Breiman [1996], are an 'ensemble method' based on decision tree models, successfully used for firm bankruptcy prediction [Bou-Hamad et al., 2011, Barboza et al., 2017].
Ensemble method means that many machine learning algorithms are combined together so that the resulting model is more powerful than any single component in the ensemble. In the case of RF, many classification trees are used (see Appendix B for more details). The advantage of assembling trees is to obtain a more robust classification and thus to increase forecasting performance [Yeh et al., 2014].

It is worth to remember that a decision tree is a flow-chart structure (i.e. directed graph) able to split the covariates' space in many non-overlapping regions, starting from a unique initial node and following a path made of many partitioning nodes. Every node splits observations according to a given covariate, and every possible path defines a region and leads to a final node ('leaf'), which contains the predicted classification (B or NB). In our study, the RF is implemented with 500 trees built on bootstrapped samples and each tree is characterized by a max depth of 15 internal nodes and by a max number of leaf nodes of 20. The final classification is obtained computing the majority vote among the 500 outputs provided by the trees.
In our RF, at every node we choose as a splitting criterion the heterogeneity Gini index (see Appendix B). In the case of imbalanced classes, as in our case, the splitting criterion is to maximize the following quantity:

$$WID = \frac{n_{node}}{n}\left[G_{node} - \frac{n_{right}}{n_{node}}G_{right} - \frac{n_{left}}{n_{node}}G_{left}\right],$$

where n_{node} is the number of firms in the considered node, and n_{right}/n_{left} are the numbers of firms split in the right/left branch. All these quantities are weighted sums: for example $n_{node} = w_B \cdot n_{B,node} + w_{NB} \cdot n_{NB,node}$, where $n_{B,node}$ is the number of B training firms observed in the node, and so on.

XGboost (XGB) (eXtreme Gradient Boosting method), firstly introduced by Chen and Guestrin [2016], is an extremely performing algorithm to implement gradient-boosted decision trees and it has been used for bankruptcy prediction [Zieba et al., 2016] and risk modelling [Wang and Ni, 2019]. XGB is an ensemble method in which each tree is built sequentially, as opposed to the RFs (see Appendix B for more details).
Roughly speaking, GB acts iteratively as follows: in the first step a (small) tree is built, which provides the (raw) classification $\hat{y}_i^1$

minimizing the cost function $\sum_i L(y_i, \hat{y}_i^1)$. In the second step, GB tries to improve $\hat{y}_i^1$ by minimizing $\sum_i L(y_i, \hat{y}_i^1 + f_1(x_i))$, in which $f_1(x_i)$ ideally is the best fit among all the possible decision trees based on the x_i's as covariates and the residuals $y_i - \hat{y}_i^1$ as responses. Successive steps are similar. Obviously, it is not possible to check all the possible trees, then some approximations are needed (see Appendix A for more details).
In our analysis the generated number of trees is equal to 5,000 with a max depth of 100. We also implement a sampling strategy of the covariates, with a threshold equal to 50%, so that no more than half of the covariates can be considered at each split.
It is worth to note that XGB does not allow to specify class weights for the loss function. However, it has a specific parameter, 'the scale positive weights', which can be implemented to account for the imbalance ratio in the dataset. Specifically, it can be used to adjust the weights associated to the classification errors of the minority class. In the analysis, we use a scaled weight for the B class equal to 1.0E+10. Given such a high value, we also have to use a low learning rate equal to 9.0E-04.

5 Evaluation

In order to measure the predictive performance of our models we conduct an out-of-sample exercise randomly splitting the whole dataset into a training set and a test set (respectively 75% and 25% of firms in the dataset). We correctly implement a stratified split so to reproduce the proportion of B and NB observations both in the training and in the tests set. In the training set we estimate the models' parameters in the case of WLR and NN and we create model instances in the case of RF and XGB. In the test set we verify the predictive performance of each model.
To compare the predictive power of the different models we used,

we calculate a set of accuracy indices. Given that we have a classification objective, prediction models are traditionally measured against a confusion matrix, which reports True Negatives (TN) = NB correctly classified, False Positives (FP)= NB misclassified as B, False negatives (FN)= B misclassified as NB, and True Positives (TP)= B correctly classified. On this basis we calculate the following quantities:
Type 1 error measures the percentage of misclassified B over all observations classified as B and is calculated as FN/(FN+TP);
Type 2 error measures the percentage of misclassified NB over all observations classified as NB and is calculated as FP/(FP+TN);
Recall measures the percentage of correctly classified B over the number of actual B and is calculated as TP/(FN+TP);
Precision measures the percentage of correctly classified B over all observations classified as B and is calculated as TP/(TP+FP);
F1 score is the harmonic mean between precision and recall, it ranges from 0 to 1;
F2 score is equivalent to the F1 score but weights the recall double with respect to the precision, it ranges from 0 to 1.
Type 1 and Type 2 are the usual measures used in the literature, the remaining indicators are considered because they are able to solve possible biases related to our unbalanced dataset. For example the F2 score, which places more emphasis on the Recall, is able to counteract the fact that in our case the Precision measure is highly influenced by the imbalance ratio between B and NB. F1 and F2 scores are alternative measures of reporting accuracy, more suitable in unbalanced scenarios. For the same reason, the *accuracy score* (TN+TP)/(TN+TP+FN+FP) is in this case not informative and we choose to not report it.
Given that the training and test sets are randomly selected, to reduce variability many random partitions are usually performed, and the aforementioned indices are averaged over such repetitions. We use here a *repeated random sub-sampling validation*,

i.e. we randomly split the whole dataset into training and test for 200 times, and for every split we estimate the described models. Results are averaged on these 200 repetitions. In addition, in WLR we build a confidence interval around the averaged regression coefficients in order to test significance.

6 Results

We start by showing the results on the predictive ability of different models, logistic regression and machine learning techniques, comparing the two specifications (only financial ratios and financial ratios + industrial variables).
As already explained in the previous Section the evaluation exercise is out-of-sample, i.e. the models are estimated in the training sets and tested in the tests sets. The rationale of this procedure is to mimic the activity of a credit institution which has some information on its client firms, divided into B and NB, and needs to classify a new client as B or NB in order to decide whether or not to grant a new loan. If the credit institution grants a loan to a firm, which was erroneously classified as NB, it will have a loss in its balance sheet, else if the credit institution does not grant a loan to a firm, which was erroneously classified as B, it loses a profit opportunity. The first type of error is what we have previously defined as Type 1 and the second is what we have previously defined as Type 2. Table 1 shows also the other metrics used in the literature.
There is usually a trade-off between Type 1 and Type 2 errors, i.e. we cannot expect to minimize both of them at the same time. From a credit institution perspective though, minimizing the error in classifying as sound a firm that will eventually become insolvent is of crucial relevance, given that the bank has the aim of reducing the number of NPL (Non Performing Loans) in its balance sheet.

Table 1: Predictive performance across models

	financial ratios					
	T1	T2	F1	F2	Recall	Precision
Logistic Regression	15.26	19.12	37.39	56.24	84.74	23.99
Random Forest	16.22	10.36	50.91	66.57	83.78	36.57
XGBoost	12.31	17.42	40.57	59.87	87.69	26.39
Neural Network	17.02	11.02	49.53	65.15	82.98	35.54
	financial ratios + industrial variables					
	T1	T2	F1	F2	Recall	Precision
Logistic Regression	**15.06**	19.71	36.79	55.75	**84.94**	23.48
Random Forest	**15.55**	11.26	49.32	65.71	**84.45**	34.84
XGBoost	**10.72**	19.22	38.89	58.80	**89.28**	24.86
Neural Network	17.26	**10.89**	49.72	65.19	82.74	**35.79**

Notes. T1=Type1 error; T2=Type2 error; Recall= 1-Type1 error, i.e. percentage of firms correctly classified as B. In bold we report the improvement of the financial ratios model augmented with industrial variables. F1 and F2 range from 0 to 1.

From the Type 1 error perspective the best model in terms of predictive performance is XGB followed by WLR and RF, whereas the NN seems not to work as well as the other two computational techniques. On the other hand, RF and NN work much better in reducing the Type 2 error.

We can thus divide models into two groups. WLR and XGB provide forecasts that are in line with a more risk averse policy in terms of granting loans, whereas RF and NN are in line with a less risk averse policy, which weighs more the cost related to the opportunity of losing profits rather than the direct cost of having NPLs.

XGB is also the model with the lowest Type 1 error which translates in the ability to classify correctly around 90% of bankrupt firms, which is a great achievement in line with other results in the literature [Barboza et al., 2017, Bottazzi et al., 2011].

To summarize, the logistic regression - the benchmark regression model used in the literature - produces comparable results with

other techniques only in Type 1 errors, whereas Type 2 errors are consistently higher. The ML techniques have different outcomes in terms of forecasting performance. RF and NN work well in reducing Type 2 errors only, XGB seems to be the best performer in both errors, but strikingly good with Type 1 errors. In relation to the importance of the industrial and regional variables in adding predictive power, Table 1 also shows that with the exception of NN all the other types of models see a reduction in Type 1 error when adding industrial and regional variables. Once again XGB seems to be able to use more efficiently the information coming from firms' industrial structure to reduce the prediction Type 1 error. For this model the advantage of using industrial variables is substantially increasing the capacity of correctly classifying B firms from 87.69% to 89.28%. This result is particularly striking given that the model was already performing very well only with financial ratios.

Given that industrial and regional variables seem to be important for firms' bankruptcy forecasting we expect that these variables have a significant role in determining firms' probability of becoming insolvent. For this purpose we report the average result of the logistic regression over the training sets (in-sample results), in order to check the sign and significance of the different variables. Table 2 reports the mean of the coefficients and their significance. Results show that indeed industrial variables have a significant impact on the probability of bankruptcy. In particular, belonging to an industrial district, having a high mark up and a high market share diminish the probability of bankruptcy. Regarding sectors and regional dummies results (which we do not report in the table) we find that food and machinery have a lower probability of bankruptcy compared to other sectors and that Southern regions have a higher bankruptcy probability with respect to other regions.

The relevance of regional disparity is an expected result given the

Table 2: Logit coefficients - results over 200 training samples

	M1	M2
Financial Ratios		
Net working capital/total assets	-3.975***	-4.022***
Net Worth/Total Debt	-0.827***	-0.824***
Total Sales/Total Assets	-1.264***	-1.270***
EBIT/Total Assets	-10.428***	-10.151***
TA growth	0.404***	0.385***
TS growth	0.160***	0.164***
ROE variation	0.004	0.004
Industrial variables		
Mark up		-0.113**
Market share		-6.398***
District dummy		-0.142***
Sector dummies		yes
Regional dummies		yes

Notes. M1= only financial, M2= financial + industrial. Coefficient are averaged across the 200 random samples and significance is based on empirical confidence intervals around the averaged coefficients. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

recent increasing dualism of the Italian economy. Also the result on sectors is not surprising. The Machinery and Food industries have a stronger capacity, in comparison to the other sectors, to propose a differentiated product and thus increase their competitive advantage crowding out foreign competitors. The first novel result of this paper concerns the positive relation between district membership and firm's solvency. The vast empirical literature on ID is silent on this issue and has focused primarily on the benefits that agglomeration economies have on economic growth through local externalities. The result of the paper highlights a different advantage linked to ID membership, i.e. bankruptcy reduction. A possible explanation could be related to the presence of social

capital, which increases the level of trust among firms and institutions sharing the same territory. A higher level of trust might in turn translate, for example, into easier access to credit which could be decisive in curbing the probability of going bankrupt. Also the result concerning the positive relation between mark up and solvency is worth to note. It seems to suggest that a high mark-up is associated with an efficient use of the firms' large rent and/or to a greater market power. Finally, results show that the size of the single firm relative to its sector (market share) is also relevant to reduce the probability of going bankrupt.

7 Concluding Remarks

We provide a predictive model, specifically assigned for the Italian economy with the aim of correctly classifying solvent and insolvent firms one year in advance.
Our results seem to suggest two different possible takeaways for economists and practitioners. The first is methodological. The WLR, which is the benchmark regression model used in the literature, produces comparable results with other techniques only in Type 1 errors, whereas Type 2 errors are consistently higher. The ML techniques have different outcomes in terms of forecasting performance. RF and NN work well in reducing Type 2 errors only, XGB seems to be the best performer in both errors, but strikingly good with Type 1 errors.
The second takeaway is related to pinning down the set of variables with which to feed our bankruptcy forecasting models. For the Italian economy industrial and regional variables seem to be relevant not only in determining the probability of bankruptcy, but also in incrementing the forecasting performance of the models. It is important to account for sectoral and regional disparities, and to consider the industrial structure of the firms.
This result is also relevant for the literature on ID and mark-up

given that belonging to a district and having a high mark-up increase the ability of firms to be solvent.

A Summary Statistics

Table 3: Summary statistics for B and NB

	NWC/TA	EBIT/TA	NW/TD	TS/TA	EBIT/TS	grTA	grTS	varROE	PCM	MS
					NB					
OBS	67077	67077	67077	67077	67077	67077	67077	67077	67077	67077
mean	0.244	0.057	0.714	1.441	0.040	0.094	0.165	-0.012	0.681	0.004
std	0.203	0.099	0.957	0.600	0.073	0.354	0.657	0.497	0.571	0.023
min	-0.085	-0.241	0.000	0.737	-0.279	-0.999	-0.999	-4.987	0.000	-0.490
25%	0.084	0.001	0.155	1.048	0.000	-0.060	-0.068	-0.094	0.334	0.000
50%	0.213	0.032	0.375	1.311	0.023	0.038	0.050	-0.003	0.535	0.001
75%	0.381	0.092	0.873	1.670	0.065	0.169	0.206	0.071	0.834	0.002
max	0.998	0.982	9.974	21.253	0.935	9.757	9.968	4.999	4.999	0.982
					B					
OBS	4774	4774	4774	4774	4774	4774	4774	4774	4774	4774
mean	-0.115	-0.131	0.086	1.020	-0.171	0.173	0.967	156.692	0.550	0.001
std	0.602	0.431	0.685	0.938	0.547	1.290	26.827	10697.286	0.520	0.028
min	-12.905	-8.620	-8.575	0.010	-14.497	-0.955	-0.975	-268.748	0.000	-1.488
25%	-0.187	-0.097	0.010	0.597	-0.136	-0.111	-0.211	-0.357	0.241	0.000
50%	-0.017	-0.025	0.055	0.868	-0.031	0.022	-0.032	-0.012	0.416	0.000
75%	0.110	-0.004	0.144	1.245	-0.004	0.207	0.200	0.276	0.690	0.001
max	0.991	0.875	30.695	41.700	0.759	38.287	1754.601	739099.870	4.935	0.377

Notes. NWC/TA= Net Working Capital/ Total Assets; EBIT/TA= Earning before interest and taxes/Total Assets; NW/TD=Net Worth/Total Debt; Total Sales/Total Assets; EBIT/TS=Earning before interest and taxes/Total Sales; grTA=growth rates of Total Assets; grTS=growth rates of Total Sales; varROE= ROE_t-ROE_{t-1}; PCM=Price Cost Margin; MS= Market Share. NB=Non Bankrupt; B=Bankrupt; OBS= number of observations.

B Methodology

Feedforward Neural Networks

The *feedforward neural network* is the simplest type of artificial neural network. This network is based on computational units called *neurons*, that are grouped into *layers* and that are interconnected in a feed-forward way: neuron in one layer has directed connections to the neurons of the subsequent layer, but not with the neurons of the same layer. Also, the signal moves only forward, from the input neurons, through the hidden ones (if any) and to the output neuron(s).
According to the 'universal approximation theorem',[8] a network architecture with at least three layers is needed, as represented in Figure 1a.
As we can see, the first layer, the input one, associates one neuron to each covariate which must be processed. These values are then processed by the second layer, known as the hidden layer. Each neuron j in the hidden layer is connected with weighted links w_{lj} to all preceding neurons l. Weights are used by neuron j to compute a weighted sum s_j of the input covariates.
This weighted sum is then passed through a non-linear activation function $f(\cdot)$, to produce the final output of the neuron $f(s_j)$, which will be passed forward to the output layer. The choice of the activation function is up to the analyst, but common choices are the hyperbolic tangent or the rectified linear unit (*relu*). The neuron computation here described, which uses the relu function, is depicted in Figure 1b.
In this last layer there are as many neurons as the number of response variables (a single neuron in the case of binary classifi-

[8]The universal approximation theorem states that every continuous function that maps intervals of real numbers to some output interval of real numbers can be approximated arbitrarily closely by a multi-layer NN with just one hidden layer to learn an appropriate representation of the input data

cation), fully-connected with all preceding neurons with weighted links. As before, a weighted sum is made by each neuron. In case of regression, this is the final output of the neuron, while in case of binary classification a logistic function is used to constrain the output in the range [0,1].
All the weights are learned during the training phase, characterized by a forward and a backward propagation phase. In the forward phase each observation of the training set is fed into the neural network and the resulting output is collected. In the backward phase the error between the collected output and the observed value of the response is computed and the weights are adjusted to minimize a global loss function, using a gradient descent optimization method.
The most common one is the backpropagation algorithm [Rumelhart et al., 1995]. The backpropagation algorithm essentially computes, using the chain-rule of derivatives, the partial derivative of the loss function with respect to all weights. This value, coupled with a learning rate of choice, is used to iteratively adjust the weights in a gradient descent fashion.
Weights are updated after a 'batch' of k observations has been processed: if the batch size k equals the number of observations n the procedure is the classic gradient descent, while if $k < n$ the procedure is called stochastic gradient descent. To reach convergence, the training is repeated a certain number of times, known as 'epochs'.

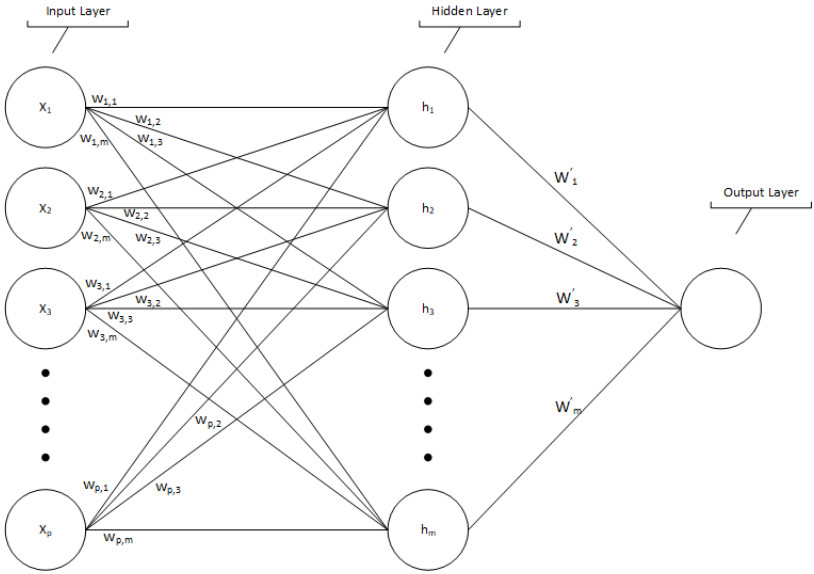

Figure 1: **Figure 1a**

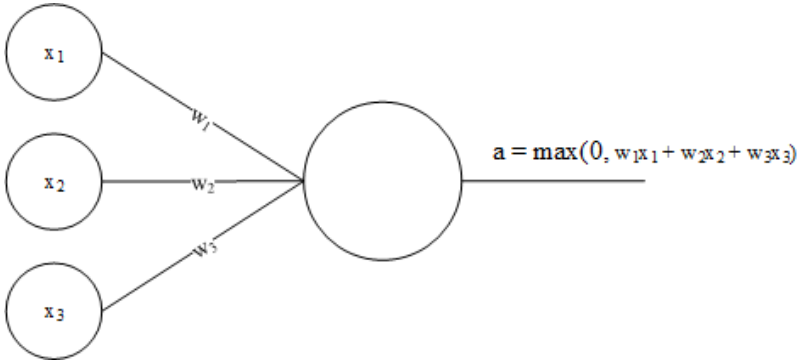

Figure 2: **Figure 1b**

Random Forest

Random Forest is an ensemble learning method based on Decision Trees. Decision trees can be applied to both regression and classification problems and involve segmenting the covariate space into a number of non-overlapping regions (partitions) using simple rules. The set of splitting rules used to partition the input space can be summarized in a flow-chart structure which can be represented using binary trees.
The building process of a decision tree follows a top-down greedy approach known as recursive binary splitting, meaning that starting from a common trunk, at every node of the tree a covariate X_i and a cut point s are chosen so that observations having $X_i < s$ (respectively $X_i \geq s$) will follow the left (respectively right) branch arising from the node.
At every iteration the best split is found relatively to the partitions already made. With this aim, the best-splitting predictor is selected, and the related best cut-point s is found such that a given metric that measures the progress of the learning of the tree results to be optimal.
For regression tasks the metric is usually the residual sum of squares (RSS), while for classification problems either the *Gini index* or the *Entropy index* can be used.
In particular, the Gini index is evaluated as follows: suppose to have n units in the training set and consider a specific node which contains n_{node} observations. If $n_{B,node}$ of them are observed to be bankrupt, and the remaining $n_{NB,node}$ are active, the Gini index in the selected node is $G_{node} = \sum_{i=B,NB} p_i(1-p_i)$, where $p_i = \frac{n_{i,node}}{n_{node}}$ is the percentage of B and NB firms (the entropy index is analogously evaluated substituting $p_i(1-p_i)$ with $-p_i log(p_i)$).
In the successive step, the n_{node} units have to be split so that n_{left} of them follow the left branch and the remaining n_{right} follow the right branch. The Gini index for the left branch is given

by $G_{left} = \sum_{i=B,NB} p_{i,left}(1 - p_{i,left})$, where $p_{i,left} = \frac{n_{i,left}}{n_{left}}$ and $n_{i,left}$ is the number of observed B/NB firms following the left branch. G_{right} is evaluated in a similar manner and the global Gini index for this particular split is given by

$$G = \frac{n_{left}}{n_{node}} G_{left} + \frac{n_{right}}{n_{node}} G_{right}.$$

Note that if firms in the node are perfectly split (for example, all the B firms on the right and all the NB firms on the left) we obtain $G = 0$. Then, in the given node, every possible combination of the covariate X_i and the cut-point s is checked, and that couple is chosen which minimize G (or equivalently, maximize the difference $G_{node} - G$.

Once the best predictor and its corresponding best cut-point is found, the process carries on until a stopping criterion is reached, for example until no region contains more than a given number of observation, or until a maximum number of regions is reached, or until the optimization metrics does not improve more than a given threshold. Additional techniques such as tree pruning can be used to reduce the tree complexity, reducing possible overfitting.

While decision trees can be simple and useful for interpretation, they are typically not competitive with other supervised techniques. Most common issues are over-fitting and high model variance. Random forest prevents these shortcomings constructing a multitude of decision trees and combining each tree output in a final response.

The forest is composed by a given number of trees trained as follows: 1) we draw B bootstrap samples from the original data, 2) we produce a tree for each bootstrap sample. At each node, select at random m out of p covariates where $m \in \{1, \ldots, p\}$ is a parameter chosen by the analyst at the start (usually $m = \sqrt{p}$). The splitting can be stopped when a minimum node size is reached or

using different stopping criteria such as maximum depth [James et al., 2013].
The methodology in training has two main characteristics. The first is that the trees are fed with different versions of the dataset obtained through a bootstrap sampling. The second is that, at each node, in order to identify the best predictor to use for splitting, the tree can choose from a reduced random subset of $m < p$ predictors. The first characteristic is shared with the bagging algorithm, while the second is introduced with the aim of decorrelating the trees' outputs. The RF method can be seen as a refinement of the bagging algorithm introduced by [Breiman, 1996].
During the prediction phase, for a given test observation, the output of each decision tree in the ensemble is recorded. The final output is obtained by averaging the different trees outputs, when the predictive variable is continuous, or computing the majority vote, in case of discrete predictive variable (such as our bankruptcy prediction task).

XGBoost

XGBoost is used for supervised learning problems, both for regression and classification tasks, and it is a particularly efficient implementation [Chen and Guestrin, 2016] of Gradient Tree Boosting [Friedman, 2001].
We remind that in general it is always possible to express the output $\hat{y}_i$ of a given machine learning algorithm (tree included) as a function of the explanatory variables: $\hat{y}_i = \phi(x_i) = \phi(x_{i1}, \ldots, x_{ip})$. In case of a tree with T leaves, each one producing the classification c_l, $l = 1, \ldots, T$, for example we may write $\hat{y}_i = \sum_{l=1}^{T} c_l \cdot 1(i \in l)$, in which $i \in l$ indicates that the $x_{i1}, \ldots, x_{ip}$ are such as to make the i-th unit to fall in the l-th leaf. In any case, the leading aim is to minimize the cost function $\sum_{i=1}^{n} L(y_i, \hat{y}_i)$.

The Gradient Tree Boosting algorithm is an ensemble model in which decision trees are sequentially constructed. Informally, at each stage of the sequential procedure, the gradient boosting algorithm tries to improve over the preceding imperfect model, by constructing a new estimator to add to the ensemble. In this way a better overall model is obtained.
The final output after K steps is

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \ f_k \in \mathcal{F},$$

where $\mathcal{F}$ is the functional space containing all the possible decision trees. The corresponding cost function to be minimized in the XGBoost algorithm is set to be:

$$\mathcal{L} = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

in which $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda ||w||^2$ is a regularization term which penalizes too much complex models, to avoid over-fitting, with γ, λ parameters to be tuned. T is the number of leaves of the t-th tree, and $w = (w_1, \ldots, w_T)$ are weights for the leaves which should be optimal in the following sense: the k-th tree in the sequence provides a response w_{ik} for the i-th unit (in our case, the probability to be bankrupted), and the final response is the sum $w_{i1} + \cdots + w_{iK}$ which minimizes $\mathcal{L}$.
More in details, let $\hat{y}_i^{(t-1)}$ be the prediction obtained at the $t-1$-th step: in the next step it is adjusted looking for the new tree f_t which minimize the quantity

$$\mathcal{L}^{(t)} = \sum_{i} L(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t).$$

where f_t uses as response variables no longer the original y_i's, but the residuals obtained in the previous step. This formula

can be simplified through the second-order approximation:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t), \tag{1}$$

(see Chen and Guestrin, 2016 for more details) in which g_i and h_i are the first- and second-order partial derivatives of $L(y_i, \hat{y}_i^{(t-1)})$ with respect to $\hat{y}_i^{(t-1)}$, $i = 1, \ldots, n$. For example, if $L(y_i, \hat{y}_i^{(t-1)}) = (y_i - \hat{y}_i)^2$ (*square loss*), we have $g_i = 2(\hat{y}_i^{(t-1)} - y_i)$ and $h_i = 2$. If the structure of a given tree q is known, the optimal weight for the l-th leaf is:

$$w_l^* = -\frac{\sum_{i \in l} g_i}{\sum_{i \in l} h_i + \lambda}$$

and re-elaborating Eq. 1 one more time, we obtain:

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2} \sum_{l=1}^{T} \frac{\left(\sum_{i \in l} g_i \right)^2}{\sum_{i \in l} h_i + \lambda} + \gamma T \tag{2}$$

However, the best tree structure q cannot be known apriori, being infeasible to enumerate all possible trees structures. So, the tree is built using another greedy approach. Specifically, the tree is built traditionally, and at each branch the best split is chosen using Eq. 2 rather than the Gini index (or entropy index). Indeed, Eq. 2 is a generalization for a wider range of objective functions of the impurity score of decision trees.

References

Pietro Alessandrini and Alberto Zazzaro. Bank's Localism and Industrial Districts. In *A Handbook of Industrial Districts*, pages 471–482. Edward Elgar Publishing, 2009.

Edward I Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609, 1968.

Roberto Antonietti and Giulio Cainelli. The role of spatial agglomeration in a structural model of innovation, productivity and export: a firm-level analysis. *The Annals of Regional Science*, 46(3):577–600, 2011.

Sofie Balcaen and Hubert Ooghe. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38(1): 63–93, 2006.

Flavio Barboza, Herbert Kimura, and Edward I Altman. Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83:405–417, 2017.

William H Beaver. Financial ratios as predictors of failure. *Journal of Accounting Research*, pages 71–111, 1966.

Giacomo Becattini. The Marshallian industrial district as a socio-economic notion. In *Industrial Districts and Inter-Firm Co-operation in Italy*, pages 37–51. International Institute for Labour Studies, 1990.

M Bellandi. External Economies, specific public goods and policies. In *A Handbook of Industrial Districts*, pages 712–725. Edward Elgar Publishing, 2009.

Florin O Bilbiie, Fabio Ghironi, and Marc J Melitz. Monopoly power and endogenous product variety: Distortions and remedies. Technical report, National Bureau of Economic Research, 2008.

Ron Boschma and Simona Iammarino. Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, 85(3):289–311, 2009.

Giulio Bottazzi, Marco Grazzi, Angelo Secchi, and Federico Tamagni. Financial and economic determinants of firm default. *Journal of Evolutionary Economics*, 21(3):373–406, 2011.

Imad Bou-Hamad, Denis Larocque, and Hatem Ben-Ameur. Discrete-time survival trees and forests with time-varying covariates: application to bankruptcy data. *Statistical Modelling*, 11(5):429–446, 2011.

Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

Giorgio Brunello and Monica Langella. Local agglomeration, entrepreneurship and the 2008 recession: Evidence from Italian industrial districts. *Regional Science and Urban Economics*, 58:104–114, 2016.

Sebastiano Brusco. The Emilian model: productive decentralisation and social integration. *Cambridge Journal of Economics*, 6(2):167–184, 1982.

Bruno Cassiman and Stijn Vanormelingen. Profiting from innovation: Firm level evidence on markups. *Available at SSRN 2381996*, 2013.

Gilbert Cette, Jimmy Lopez, and Jacques Mairesse. Market regulations, prices, and productivity. *American Economic Review*, 106(5):104–08, 2016.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.

Federico Cingano and Fabiano Schivardi. Identifying the sources of local productivity growth. *Journal of the European Economic Association*, 2(4):720–742, 2004.

John K Courtis. Modelling a financial ratios categoric framework. *Journal of Business Finance & Accounting*, 5(4):371–386, 1978.

Juan J De Lucio, Jose A Herce, and Ana Goicolea. The effects of externalities on productivity growth in Spanish industry. *Regional Science and Urban Economics*, 32(2):241–258, 2002.

Avinash K Dixit and Joseph E Stiglitz. Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(3):297–308, 1977.

Johan Eklund, Nadine Levratto, and Giovanni B Ramello. Entrepreneurship and failure: two sides of the same coin? *Small Business Economics*, pages 1–10, 2018.

G Foster. Distress analysis and financial information. In Editor Charles T. Horngren, editor, *Financial Statement Analysis second edition*, chapter 15, pages 534–572. Prentice Hall International Editions, Englewood Cliffs, NJ, 1986.

Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.

Edward L Glaeser, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer. Growth in cities. *Journal of Political Economy*, 100(6):1126–1152, 1992.

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.

Luigi Guiso and Fabiano Schivardi. Spillovers in industrial districts. *The Economic Journal*, 117(516):68–93, 2007.

N Hart. External and internal economies. In *A Handbook of Industrial Districts*, pages 90–102. Edward Elgar Publishing, 2009.

Vernon Henderson, Ari Kuncoro, and Matt Turner. Industrial development in cities. *Journal of Political Economy*, 103(5): 1067–1090, 1995.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*, volume 112. Springer, 2013.

Gary King and Langche Zeng. Logistic regression in rare events data. *Political Analysis*, 9(2):137–163, 2001.

P Ravi Kumar and Vadlamani Ravi. Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European Journal of Operational Research*, 180(1):1–28, 2007.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

Deron Liang, Chia-Chi Lu, Chih-Fong Tsai, and Guan-An Shih. Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(2):561–572, 2016.

Alfred Marshall. *Principles of Economics*. Macmillan, London, 8 edition, 1890.

Philippe Martin, Thierry Mayer, and Florian Mayneris. Spatial concentration and plant-level productivity in France. *Journal of Urban Economics*, 69(2):182–195, 2011.

Yaw M Mensah. An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study. *Journal of Accounting Research*, pages 380–395, 1984.

Steffen Mueller and Jens Stegmaier. Economic failure and the role of plant age and size. *Small Business Economics*, 44(3): 621–638, 2015.

James A Ohlson. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, pages 109–131, 1980.

David E Rumelhart, Richard Durbin, Richard Golden, and Yves Chauvin. Backpropagation: The basic theory. *Backpropagation: Theory, architectures and applications*, pages 1–34, 1995.

Fabio Sforzi. The geography of industrial districts in Italy. In *Small Firms and Industrial Districts in Italy*, pages 153–73. Routledge London, 1989.

L Federico Signorini. The price of Prato, or measuring the industrial district effect. *Papers in Regional Science*, 73(4):369–392, 1994.

Antoine Soubeyran and Shlomo Weber. District formation and local social capital: a (tacit) co-opetition approach. *Journal of Urban Economics*, 52(1):65–92, 2002.

Sadahiko Suzuki and Richard W Wright. Financial structure and bankruptcy risk in Japanese companies. *Journal of International Business Studies*, 16(1):97–110, 1985.

Carlo Trigilia. Social capital and local development. *European Journal of Social Theory*, 4(4):427–442, 2001.

Yan Wang and Xuelei Sherry Ni. A XGBoost risk model via feature selection and Bayesian hyper-parameter optimization. *arXiv preprint arXiv:1901.08433*, 2019.

Ching-Chiang Yeh, Der-Jang Chi, and Yi-Rong Lin. Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences*, 254:98–110, 2014.

Dong Zhao, Chunyu Huang, Yan Wei, Fanhua Yu, Mingjing Wang, and Huiling Chen. An effective computational model for bankruptcy prediction using kernel extreme learning machine approach. *Computational Economics*, 49(2):325–341, 2017.

Maciej Zieba, Sebastian K Tomczak, and Jakub M Tomczak. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, 58:93–101, 2016.

C Zopounidis. A multicriteria decision-making methodology for the evaluation of the risk of failure and an application. *Foundations of Control Engineering*, 12(1):45–64, 1987.

Printed by
Gi&Gi srl - Triuggio (MB)
October 2019


9788834341117