CIRRELT

CIRRELT-2020-29

A Machine Learning-based DSS for Mid and Long-Term Company Crisis Prediction

Guido Perboli
Ehsan Arabnezhad

July 2020

Bureau de Montréal
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal (Québec) H3C 3J7
Tél : 1-514-343-7575
Télécopie : 1-514-343-7121

Bureau de Québec
Université Laval,
2325, rue de la Terrasse
Pavillon Palasis-Prince, local 2415
Québec (Québec) G1V 0A6
Tél : 1-418-656-2073
Télécopie : 1-418-656-2624

A Machine Learning-Based DSS for Mid and Long-Term Company Crisis Prediction

Guido Perboli[1,2,3,*], Ehsan Arabnezhad[2,3]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)
[2] ARISK s.r.l., Corso Castelfidardo 30/a - 10129 - Torino presso I3P Scpa, Turin, Italy
[3] ICT for City Logistics and Enterprises Center, Politecnico di Torino, Turin, Italy

Abstract. In the field of detection and prediction of company default and bankruptcy, many efforts have been devoted to evaluate the financial ratios as predictors with statistical models and machine learning techniques. This problem comes to importance where the big financial decision makers are provided with prediction models to act based on such models. Unfortunately, the current results are good predictors in the short term and are mainly focused on large and medium-large companies. In this paper, we focus this issues in the mid and long-term (up to 60 months), focusing our research on Small and/or Medium Enterprises. The key insight of the study is a substantial improvement in prediction of accuracy using machine learning techniques compared to state-of-the-art results in the short term (12 months), while making accurate predictions in the mid and long term (measure of area under the ROC curve of 0.88 with a 60 months prevision horizon). Extensive computational tests on the entire set of companies in Italy show the efficiency and the accuracy of the developed method, as well as the possibility to use it as a tool for the development of strategies and policies for entire economic systems. By considering the recent COVID-19 disease crisis, we show how our methods can be used as a viable tool for large-scale policy-making.

Keywords: Bankruptcy prediction, data analysis, machine learning.

Acknowledgements. While working on this paper, Guido Perboli was the head of the Urban Mobility and Logistics Systems (UMLS) initiative of the interdepartmental Center for Automotive Research and Sustainable mobility (CARS) at Politecnico di Torino, Italy and R&D Director of ARISK, a Spin-off of Politecnico di Torino

---

* Corresponding author: guido.perboli@polito.it

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2020

1 Introduction

Because of the huge impact of company crisis in economy, society and even global debt it has been always interesting, and nowadays more, to accurately study the consequences of bankruptcy and finding ways to avoid it. To reduce the effects of the crisis, companies apply for economical help/fund from financial institutions, while decision-makers in the financial system try to identify those companies, which are highly possible to declare bankruptcy state in the future. For this reason, company crisis/bankruptcy prediction tries to assess the financial health and future performance of a company. The literature mainly focused its research on the financial aspects, obtaining results with a good prediction rate in the short term, normally 12 months (Altman, 2014; Altman et al., 2016). Moreover, due to the different regulations, the methods tend to be more accurate for large and medium-large companies. This is true in particular in countries where the number of the SMEs is larger and they are characterized by a plethora of small companies (Altman et al., 2020, 2013).

This paper contributes to the literature along two axes. First, we introduce state of the art insolvency prediction model as a decision support system (DSS) based on Machine Learning. One of the novelties in our work consists of a two-round tuning algorithm for the Machine Learning module to be used when the dataset is highly unbalanced, where in the first round we choose (by threshold) outperforming companies to make our data set more robust to distinguish between the companies that will face a crisis and those which will not, with an improvement of up to 11 percent in accuracy with respect to previous works that have been previously done (Son et al. (2019)). The Machine Learning module is tuned using the financial statement data of more than 160,000 Italian SMEs that are live and operational by the end of 2018, joint with about 3,000 bankrupted company's data covering the period 2001-2018. Extensive computational results show the accuracy of our method not only in the short term (12 months) but also in the medium (36 months) and long one (up to 60 months). Second, we illustrate how our system might be used by company owners and decision-makers as a viable strategic tool. We apply our DSS to two different settings: the Italian SME system before the COVID-19 disease and the post-COVID economy, using the DSS to evaluate the financial policies of the Italian Government and testing different variants of the policies on the total set of SMEs of the Piedmont region.

This paper is organized as follows. In Section 2, we see different literature trails in this field, also highlighting the main gaps. Section 3 is devoted to present the overall DSS, while Section 4 describes the data and the machine learning module, whose performances are discussed in Section 5. The application to the Italian SME system is discussed in Section 6, including the usage of the DSS to the post-COVID-19 economic crisis. Finally, Section 7 summarises the results and presents the possible future direction of the research.

2 Literature Review

Financial institutions, fund managers, lenders, governments, and financial market players started to develop models to efficiently assess the likelihood of companies default almost 1 century ago, when in 1932, Patrick (1932) performed a multi variable analysis on 20 companies. Researchers and practitioners developed several quantitative approaches. In 1967, Beaver (1966) applied a t-test to get the significance of each ratio for each company. Altman (1968) used the multiple discriminant analysis (MDA) however, false statistical assumptions underlying the MDA approach, led researchers to concentrate their efforts on the development of conditional probability models (logistic regression) on the data sets (Ohlson, 1980). We analyze the literature along three axes. The prediction method type, the horizon of the prediction and the data types that are incorporated in the prediction model.

2.1 Prediction method

Traditional methods rely on statistical models. Altman generates a score by which to classify observations between good and bad payers (Altman, 1968). following his work, other applications were developed by specializing the model to specific sectors and segmentations (Altman et al., 2020, 2016; Altman, 2014). In contrast to Altman, Ohlson was one of the first researchers to apply logistic regression analysis to default estimation (Ohlson, 1980). Ohlson's model determines the default probability of the potential borrower. Several subsequent studies have sought to perform similar tests, thanks to the relative ease of running discriminant analysis and logistic regression (Hillegeist et al., 2004; Upneja and Dalbor, 2001; Chen et al., 2010). The advantages of these models are first, deriving an analysis of the certainty (probability) of the results and ,second, to evaluate the effect of each feature individually. Despite their wide adoption in research and industry, these classes of models had become inaccurate and suggested the need for enhancements in the modeling of default risk (Begley et al., 1996). Moreover, they require, to be accurate, to be tuned in different markets (e.g., having different tuning of the parameters for industry and services) and have a limited ability to enhance the predictive results, normally not over 12 months (Altman, 2014; Altman et al., 2016). Moreover, they cannot incorporate automatically into large time-series of data and rely on the standard mean-value theory, while in the most part of the cases extreme events are the key factors, and the extreme-value theory might give a better insight (Baldi et al., 2019; Perboli et al., 2014).

To overcome the limitations of the statistical models, studies that use pattern recognition methods have been developed actively in the field of machine learning (Linden, 2015; Barboza et al., 2017), showing how machine learning models can outperform traditional classification methods. Some of these works rely on artificial intelligence systems as neural

Table 1: Summary of machine learning models

Algorithm family	Linear	Accurate	Easy to interpret	Scalable	Algorithms
Linear Models	Yes	No	Yes	Yes	Linear regression Logistic regression
Basic Models	Possibly	No	Yes	Possibly	Naive Bayes DT, KNN
Ensemble Models	No	Yes	Yes	Yes	RF, AdaBoost Gradient Boosting
SVM	No	Yes	No	No	SVM
Deep Learning	No	Yes	No	Yes	MLP classifier MLP regressor

networks and genetic algorithms (Odom and Sharda, 1990; Coats and Fant, 1993; Boritz et al., 1995). Several new works also showed the power of ensemble models to deal with imbalanced data set (Brown and Mues, 2012; Kim et al., 2015). Figini et al. (2017), discussed in-depth particularly the difference between parametric and non-parametric methods to analyze the credit risk of SMEs where they used multivariate outlier detection technique to enhance the results. In an interesting case study of neural networks, Brédart (2014) used a limited number of features/ratios on Belgian SME's, improving the performances of the previous works (Shah and Murtaza, 2000; Becerra et al., 2005).

As the summary in Table 1 we can see that generally ensemble methods (bagging Breiman (1996), boosting Freund et al. (1999), stacking) outperform other methods. Gradient boosting is a powerful ensemble method which recently absorbed many attention for researches on company insolvency and it turned out that it is one of the best indeed (Friedman, 2001). This model is an additive model which operates on weak learners (e.g. decision tree) until the model doesn't improve the results based on a loss function.

To the best of our knowledge, presently the best results are obtained in Son et al. (2019), where the authors applied XGBoost to a dataset, audited by a Korean credit rating agency. Despite the good accuracy, the main problems rely on the difficulty to understand the prediction capacity in the mid-term (over 24 months), which is a classic problem in all the models thought as a financial rating, and the accuracy highly dependent on external factors, as the presence of a regulation which obliges the companies to have an external audit (which is compulsory only for a subset of the SMEs and it is dependent on the single country regulations).

2.2 Time horizon of the prediction

Traditional models are accurate up to 12 months, with some cases in which the prediction maintains a sufficient accuracy (around 70%) up to 24 months (Altman, 2014; Altman et al., 2016, 2013; Hillegeist et al., 2004; Upneja and Dalbor, 2001; Chen et al., 2010; Altman et al., 2020). Even in the case of machine learning, the prediction is at a short term (Son et al., 2019; Barboza et al., 2017). The ideal prediction model should be able to make mid-term forecasts. In fact, many studies show how the failure process symptoms can be traced back to 5-8 years before failure (Argenti, 1976; Hambrick and D'Aveni, 1988; Luoma and Laitinen, 1991; Ooghe and Prijcker, 2008). Thus, there is a need for studies with a longer horizon than a few months.

2.3 Data incorporated in the prediction model

The general improvement over time of the traditional model accuracy should be linked to the selection of ratios and indexes to be included in the statistical model. However, as highlighted by Balcaen and Ooghe (2006), who reviewed business failure studies over last 35 years, there is a little consensus on which variables are the best in discriminating between failed and non-failed firms. Moreover, the most of the literature focused on financial data, disregarding the non-financial ones. For a detailed discussion about this topic the reader can refer to the recent paper by Altman et al., where a deep review of the topic is given (Altman et al., 2016). The literature shows, in any case, how the introduction of non-financial data can improve the performances and the time horizon of both the traditional and the machine learning models (Altman et al., 2016; Son et al., 2019). Unfortunately, up to now the related studies try either to see if there is a correlation between bankruptcy and non-financial variables as done by Altman et al. (2016), or just adding one-two variables related to the organization of the company (normally the industry type and the presence of an external audit, as in Son et al. (2019)).

2.4 Literature gaps and paper contribution

From the analysis of the relevant literature, it emerges how there is a gap between the best practices available methods and the market needs. In fact, there is no model, in the literature or in the market, able to be accurate both in the short (one/two years) and the mid-term (up to 5 years), adaptable to different markets with a standard (and possibly automated) tuning and able to incorporate and analyze the effects on financial and non-financial variables. In this paper we try to give a first answer to these needs, by introducing a machine learning-based DSS able to give accurate previsions both in the short and mid-term and a new method for the tuning of machine learning methods in the case of unbalanced data able to improve the overall performances of machine learning methods.

3 Decision Support System

Our DSS considers, but is not limited to, financial data. It can collect, catalog and incorporate several types of risks. The present version collects info related to budget and financial data, company organization data, family risk matrices related to cash flows, supply chain management and ... The overall DSS structure is shown in Figure 1 and 2. The DSS is developed by ARISK, a fin-tech spin-off of Politecnico di Torino providing business interruption prediction services to SMEs, It is split into two different sections: a training and tuning module and a prediction server.

The training and tuning module (see Figure 1) collects data from public databases as public financial data (in Italy the Italian Camera di Commercio), a set of indexes and ratios from AIDA Bureau van Dick, as well as, whether available, data from the proprietary interface by ARISK to collect additional data. Then the data are cleaned, normalized and merged. Data are thus split between core and non-core sets. The core data represents the features of the machine learning module, while the non-core data are data that are not directly incorporated in the machine learning. An example can be qualitative data coming from specific industrial sectors.

Core data are then managed by the machine learning pipeline for reducing the features first (feature selection procedure) and then the Machine Learning algorithm is chosen and tuned. At the moment, our system considers a wide set of Machine Learning systems, including Random Forest, XGBoost, Logistic regression, and Neural Networks. The outputs of this pipeline are the binary files of the predictors then passed to the prediction module.

Non-core data are considered as secondary data that are not directly incorporated in the Machine Learning predictor, but whose effects are simulated as perturbations of the Machine Learning features. This is done by a specific pipeline. The non-core data are first classified by a tree-based taxonomy, based on the SHELL-based taxonomy by Cantamessa et al. (Cantamessa et al., 2018). The methodology adopted for the analysis of startup failure is based on the SHELL model, originally implemented to classify aviation accidents and errors, and here adapted to the entrepreneurship sector. The SHELL model, whose name derives from the initial letters of its components, Software, Hardware, Environment, Liveware People and Liveware Environment, was developed by Hawkins in 1975 basing on the original work proposed by Edwards in 1972 under the name SHELL model. Specifically, the SHELL model requires analyzing how each person acted and interacted with the other four components. The different interactions between the person and each of the other components are considered as the human possibility, while a mismatch between the central Liveware and any other four components leads to a source of human error. Moreover, the SHELL methodology adapted to the analysis of the startup failures presented excellent behaviors compared to other results in the literature (Cantamessa et al., 2018). For the aforementioned reasons, we decided to adopt the basic framework of the SHELL for Startups model and to incorporate it in our system. The output of the SHELL model is then joint with expert-based rules mapping the effect of the different components on the core features and creating a risk impact matrix to be used for perturbing the Machine Learning module in the prediction module.

The prediction of a company's business interruption and bankruptcy risk phase is performed by the system depicted in Figure 2. Given the data of a single company and its risk Matrix obtained by applying the risk impact matrix, a request to the prediction server is sent by REST APIs. The server checks the data, gives the core data to the Machine Learning module, while the non-core data are processed by the Risk's Impact Matrix and then the corresponding oscillations of the core data are introduced in the Machine Learning obtaining the effect on the prediction given by the non-core data prediction. For each set of core and non-core data 5 predictions are created (12, 24, 36, 48, and 60 months), plus a series of performance indexes related to national and international regulations and are then merged in a report. The report gives to the user (entrepreneur, bank, assurance, policy-maker) a detailed description of the company situation, as well as the key aspects that should be considered for reducing the business interruption and bankruptcy risks within a continuous improvement process.

4 Machine-learning prediction

This section describes our Machine Learning applied to financial data and used to predict the company bankruptcy, but the process is generic and can be repeated and applied to other data types too. The process works as follows:

- Data cleaning. Data of companies that went in bankruptcy and those which are still active are collected and cleaned;
- Data fusion and first balanced dataset creation. Data from the two sets of companies are joint. Due to the heavy unbalancing between bankruptcy and active companies, a balanced dataset is obtained by sampling the active dataset;
- Data split. To evaluate the performance of a machine learning algorithm, we need to split the data. One part (train) is used for the algorithm to learn how to predict future instances and another part (test) is to examine how good is our algorithm about predicting future samples. This is done by using the Python's Scikit-learn

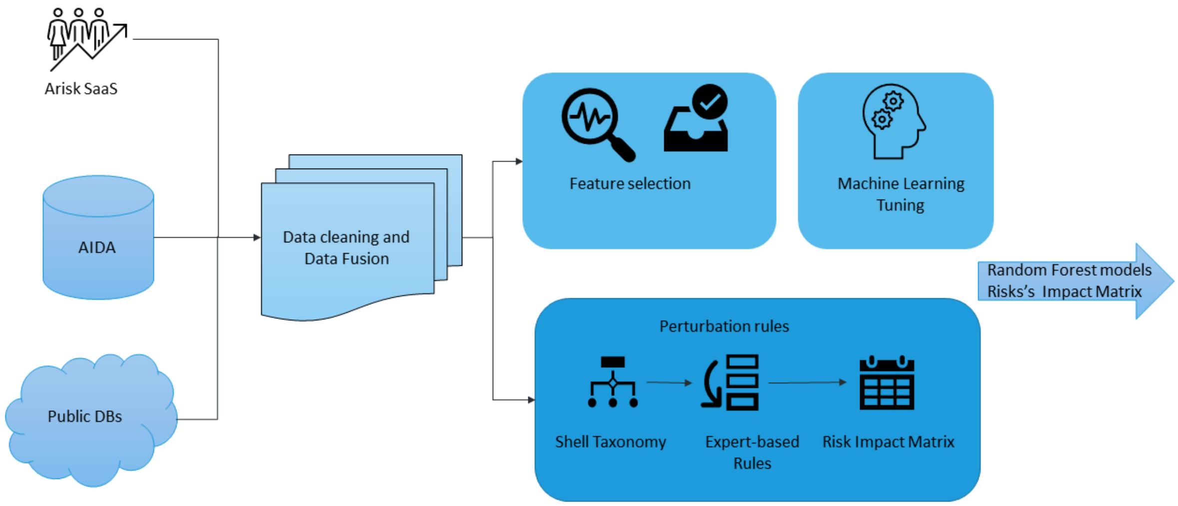

Figure 1: Decision Support System - System training and tuning

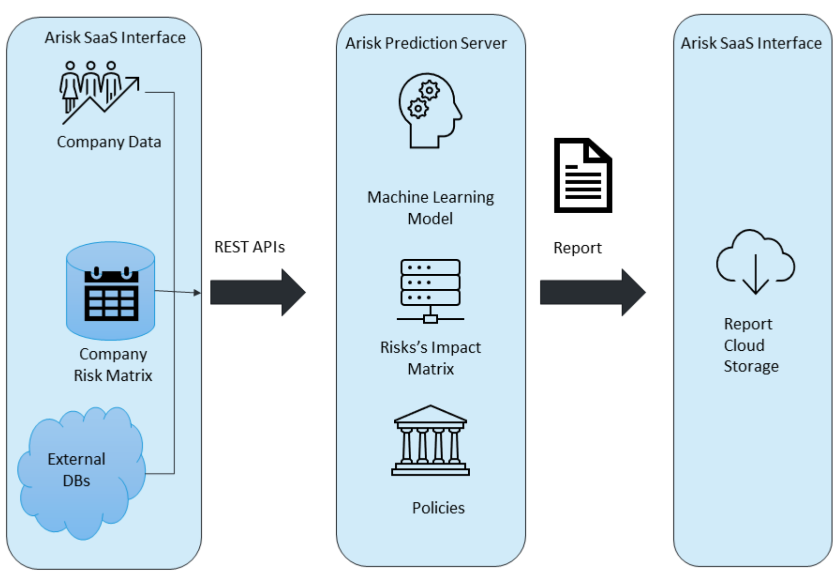

Figure 2: Decision Support System - Machine Learning prediction module

library, setting the test datatset equal to the 20% of the total. Further validation approaches to spot over-fitting or under-fitting have been tested on the data(k-fold,k = 10) (Burnham and Anderson, 2004; Cai, 2014).

- Feature reduction. Being the initial set of financial features composed by about 170 indexes and ratios, this number is reduced with an iterative procedure.
- Hyper Parameter Tuning. The parameters of the Machine Learning method are tuned. To enhance the performance of the Machine Learning module, we need to tune these parameters to get the best results. These parameters may be different From one classification task to another. For this step, we used an exhaustive tuning approach based on Grid-search. (Bergstra and Bengio, 2012).
- Final dataset creation. The dataset used in the previous step was built to be representative of some attributes related to geographic dispersion and the industry type. However, as previously stated, one of the contributions of this paper is the introduction of a procedure to obtain a sample that increases the performances of the Machine Learning. In this step, the final dataset is created.

In the following, we give more details concerning the main phases.

4.1 Data cleaning

The company data that we used in this work was made of financial information on Italian companies from 2001 to 2018. This data is extracted from the AIDA database, the largest financial and organizational database managed by Bureau van Dijk/Moody's (Bureau van Dijk, 2020). All types of companies are either limited companies or joint-stock companies. Out of which we collected bankrupted companies such that, all of them had revenues between 1 million to 40 millions of euros in at least one of the last 5 years of life before they go bankrupt and a company lifetime of at least 10 years. For each company, we collect the last 5 years of financial data of such companies and save them into 5 different data sets which are roughly made up of 3000 companies. If a bankrupted company has less than 5 official financial reports, it is removed from the dataset.

The active companies are composed of all active companies in 2018 and we collected, again, all financial data of those companies which their revenue is the same as bankrupted companies in the last 5 years of life (i.e., the last 5 years before 2018). For these companies, we kept the last year's information. The number of all active companies in 2018 is more than $160,000$. We will refer in the following to the companies that went bankrupt as Class 1 and the active companies as Class 0. Pre-processing data is an important task and it is necessary for improving machine learning metrics. This is because data is mostly noisy, it sometimes has missing and also false values. We applied missing value imputation Barnard and Meng (1999)Hilbe (2009) and standardizing to our train and test. We replaced all missing values with zero and we applied standard scaling to both. As shown by figure 3, standardizing of data seems to cause the distribution of data to be close to a normal distribution and this will result in improving the prediction. Just for the oversampling part, it is done only on the training set by the Imblearn Python library.

We should notice that if we build our data set in this way the data set would become highly imbalanced and this will affect the result of our machine learning model regarding the recall of confusion matrix. So just to remedy the negative effect of this, first we sampled $6,000$ active companies out of $160,000$ and then start to merge. By doing this we still preserve the imbalanced nature of the data set but in a controlled way meaning that by doing so we somehow sacrifice precision in favor of recall since finding all companies that most likely will declare bankruptcy status is more important. Now for each year of information of bankrupted companies, we add the same sampled of active companies and build our final data set which has 5 different parts (year1, year2, year3, year4, and year5) and is composed of 8959 companies.

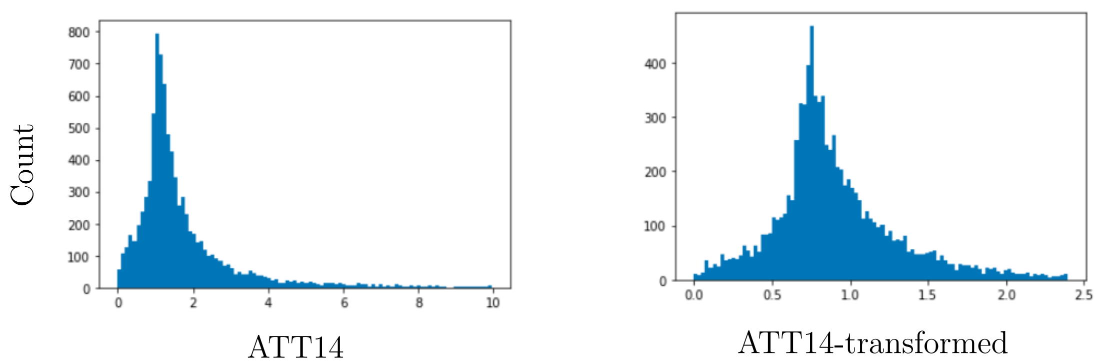

Figure 3: Standardizing

4.2 Feature reduction

At the beginning of the process when we collected financial information of companies, there were more than 170 different financial and operational features for each company. So we reduced the dimensions of our data set by an iterative feature removal process. More in details, at every step we remove one feature and we discard it if the precision score of a simple classification task didn't change more than 1 percent. By repeating this procedure we removed more than 150 features from our data set and we were left with 15 most important financial features. Table 2 reports a summary of the features. As in many other papers we cannot give the detailed list of the feature sets after the tuning, being under a Non-Disclosure Agreement. On the other side, and differently from the majority of the other works, we give to the reader an idea of the feature data types by

Table 2: Data set features

Feature	Feature Value	Feature Type
ATT10	Absolute Value	Revenue/Profit
ATT11	Index/Percentage (%)	Revenue/Profit
ATT12	Absolute Value	Revenue/Profit
ATT13	Absolute Value	Revenue/Profit
ATT14	Index/Percentage (%)	Revenue/Profit
ATT15	Index/Percentage (%)	Cost/Debt
ATT16	Absolute Value	Cost/Debt
ATT17	Index/Percentage (%)	Cost/Debt
ATT18	Absolute Value	Cost/Debt
ATT19	Index/Percentage (%)	Cost/Debt
ATT20	Absolute Value	Production
ATT21	Absolute Value	Production
ATT22	Index/Percentage (%)	Revenue/Profit
ATT23	Absolute Value	Production
ATT24	Index/Percentage (%)	Cost/Debt

providing some feature information. In more detail, data are split into category types (Profit, Cost, and Production) and feature value type (index or absolute value).

4.3 Evaluation

There are many metrics to evaluate a machine learning classification method (which is the effort to assign each future sample to its correct class). Depending on the nature of classification the trade-off between false-positives and false-negatives must be taken into account. In Powers (2011) some metrics for classification problems are introduced. In literature AUC is the most widely used metric for evaluation of a machine learning classifier, however, it is not a perfect metric when different classifiers are used (Hand, 2010, 2009). Moreover, the AUC curve is the metric used in previous works and so we adopt it to obtain comparable results (Barboza et al., 2017; Son et al., 2019) After training the data on the training set, we used AUC and confusion matrix in addition to Matthews score to evaluate the results on the test set. Since we used cross-validation, in each run, on the train and validation sets, we evaluated those metrics but rather than reporting all of them we averaged them to compare with the result of the test set to see if we are running into over-fitting or not.

4.4 Final dataset creation

In this section, we are going to explain the difference between our approach and all other previous researches which are done in this field, where we add another round of prediction based on the output of the previous round and build our data set again to further differentiate between active and failed companies. Active companies (Class 0) are companies still active, but they contain companies that will go in bankruptcy in the next years. Thus, to consider this aspect, we use a two-step procedure.

In the first round, along with getting the results that inform us which companies will go bankrupt in the five consecutive years, we also get the probability of going bankrupt. A classifier will output a company as bankrupt if the probability of bankruptcy for that company is more than 50%. Note that this is not true for all classifiers where some of them do not operate on the probabilities and instead on discrete zero or one values.

So, for each company and each year, we have a probability for which that company will go bankrupt. At this point, we tried to find a threshold that tells us which companies are less likely to go bankrupt and sample those companies and build our active companies again as we did in section 4.1. We sampled those companies as active if for all 5 consecutive years the probability of going bankrupt is less than the threshold. For example, if we set the threshold to 20%, we will have $80,000$ companies for which the probability of going bankrupt in all 5 years is less than 20% (the outperforming companies). We again sample 6,000 active companies out of this number $(80,000)$ and we build our data set as described in section 4.1.

We set the threshold to 20, 30, 40, 50, 60 % of bankruptcy and for each, we extract the new active companies from the active companies in the first round and we test the Machine Learning method.

5 Computational results

In this section we will discuss the results of our model which is described in section 4. We also compare our results with the ones by Son et al. (2019) and Figini et al. (2017). The performance metrics namely AUC, Matthews coefficient and log loss are computed on the test set. After this step and with all results in hand, we save the model for the second step of our approach to be done on data.

Concerning the two-phase dataset creation, we found a good threshold equal to 60 percent. By setting this threshold in fact we both maintain the companies which are doing very well to be active for the next 1 to 5 years and at the same time we do not bias our data set toward bankrupted companies. Besides, empirically we should consider that total number of companies going to bankruptcy at each year is roughly around 3 percent, so again 60 percent threshold will confirm this fact as well. The results of the method applied to different Machine Learning are presented in Table 3, where "First round" presents the results of the standard sampling, while "Second round" presents the performances of the two-phase approach. For the sake of simplicity, we present just the results of the best threshold (60%). As witnessed by the results, our two-phase procedure gives better performances than the standard stratification.

In our work, Gradient Boosting (GB) algorithm outperformed Logistic Regression and Neural Networks, while has similar (but slightly better) performances than Random Forest. The GB algorithm achieves better Log-Loss (GB = 0.25) than the other three models (Log Loss RF = 0.29, NN = 0.3, LR = 0.41). This result also confirms other historic results presented in Son et al. (2019) and Figini et al. (2017). From Figure 4, it is clear that logistic regression is the worst classifier as it also shown in table 3. Neural network suffers from over-fitting, even its accuracy is not as high as RF. And we can observe, in this particular task, the Gradient Boosting model generalizes the best, followed by Random Forest, showing a general better behaviour of the ensemble methods. In addition, as can be seen both in the Table 3 and Figure 6, doing the second round will improve AUC metric at least 7 percent, where for example for random forest model, AUC of the first round is 78% and it has been increased to 88% in the second round. Among ensemble models, the GB perform better than random forest (see Table 3). If we consider the best result of our work, GB and Random Forest, to the best results of the other two papers, the AUC score of our solution (after second round) achieves comparable or better results than that of achieved in Son et al. (2019) and Figini et al. (2017). This will be bolder because we used only 15 features as an independent variable regarding that they used more than 40 variables and not only financial data. This shows the effectiveness of our procedure for feature selection and the importance of financial data on prediction. Moreover, differently from the other papers, where at most 18 months predictions were considered, we present an almost constant prediction rate up to 5 years.

A study on the type of industrial activity and geographical location of companies have been done to observe the effect of these features on the company crisis. In particular, being those classifications normally based on arbitrary and historical reasons, we want to test if our Machine Learning module is independent of them. This is crucial in particular for industrial activity. It might not catch the real activity of the company. Imagine, for example, a company doing precision agriculture. It will be classified as agriculture (which is normally identified by a middle level of automation, low level of innovation and low or middle level of digital revolution), while, for the type of activity, it will be probable more similar to a company of the same size in the Industry 4.0, which means high level of innovation and digital penetration. The industrial activity is coded according to the ATECO code which represents the type of activity of a company. This code is a six digit code in which the first 2 digits reveal the greater area of activity. We divided this

Table 3: Summary of First and Second Round Prediction

Algorithm		First Round			Second Round		
	After	AUC	Matt	Log-loss	AUC	Matt	Log-loss
Random Forest	1st year	0.78	0.41	0.53	0.88	0.65	0.40
	2nd year	0.79	0.42	0.51	0.89	0.66	0.37
	3rd year	0.82	0.49	0.48	0.90	0.69	0.35
	4th year	0.85	0.53	0.45	0.91	0.71	0.32
	5th year	0.87	0.58	0.41	0.93	0.74	0.29
XGBoost	1st year	0.77	0.41	0.54	0.88	0.66	0.37
	2nd year	0.79	0.41	0.51	0.90	0.67	0.35
	3rd year	0.82	0.48	0.48	0.92	0.71	0.30
	4th year	0.85	0.53	0.44	0.93	0.75	0.27
	5th year	0.88	0.58	0.40	0.94	0.79	0.25
Logistic Regression	1st year	0.71	0.30	0.61	0.78	0.38	0.53
	2nd year	0.74	0.31	0.60	0.81	0.42	0.50
	3rd year	0.76	0.37	0.57	0.83	0.48	0.48
	4th year	0.79	0.43	0.55	0.86	0.51	0.44
	5th year	0.81	0.46	0.52	0.88	0.57	0.41
Neural Network	1st year	0.72	0.32	0.61	0.85	0.59	0.44
	2nd year	0.73	0.34	0.62	0.88	0.63	0.40
	3rd year	0.79	0.45	0.54	0.90	0.67	0.35
	4th year	0.81	0.47	0.52	0.91	0.69	0.34
	5th year	0.84	0.50	0.50	0.93	0.76	0.30

area into four subareas namely: Industry, Commerce, Public and Services and assigned a relevant code to each company. The mapping between ATECO codes and industry is done according to the Italian Ministry of Economy classification. The second feature is the geographical feature where the company established to work and are done according to the province and the region in which the Company has its headquarters

In figure 5, we can see that with these additional features there is almost no improvement in the AUC of GB model (also the same for other models). Maybe this makes sense because most of the companies tend to slightly change the activity type during their lifetime and also for the location, it is the case that they just register the company in one place (mostly because of heavy bureaucracy procedures) and start to work in another place (city).

As already discussed before, a model to accurately interpret the results is highly preferred in this field. In this research, to explain the decision made by the Machine Learning model, Shapley Additive Explanation (SHAP) method is used, as it has been described in Lundberg et al. (2020) , Lundberg et al. (2018) and Lundberg and Lee (2017). The results are shown in Figure 7 and Figure 8, where the overall importance of variables that affect the model from top to bottom. The SHAP value might be the only method to deliver a full explanation. In situations where the law requires explainability – like EU's "right to explanations" – SHAP might be the only legally compliant method because it is based on a solid theory and distributes the effects fairly. In figure 8 we evaluate the outcome of our analysis for a particular random company. We see that base value (average target probability without any prediction) is around 0.35. Also, we observe that what are the features that most affecting that company's target evaluation. As can be seen, these features push the outcome (0.28) from base value toward non-bankruptcy. It means that meanwhile the average probability of any company that would become bankrupt by chance is around 35% but this particular company has 28% probability to go bankrupt based on our evaluation.

6 Application to a real economic system: the Italian case study

In the following we show how our machine learning algorithm can be used as a predictive tool for an entire economic system, focusing on Italian SMEs. In more details, we consider all the 160000 companies with revenues between 1 and 40 millions of euros in at least one of the fiscal years in the interval 2013-2018 (2018 is the last fiscal year for which the official annual balance sheets were available, due to the COVID-19 crisis) and that are active at the end of 2019, i.e., no bankruptcy, merger & acquisition or displacement event is in the official Italian records. The results have been validated by a group of experts, led

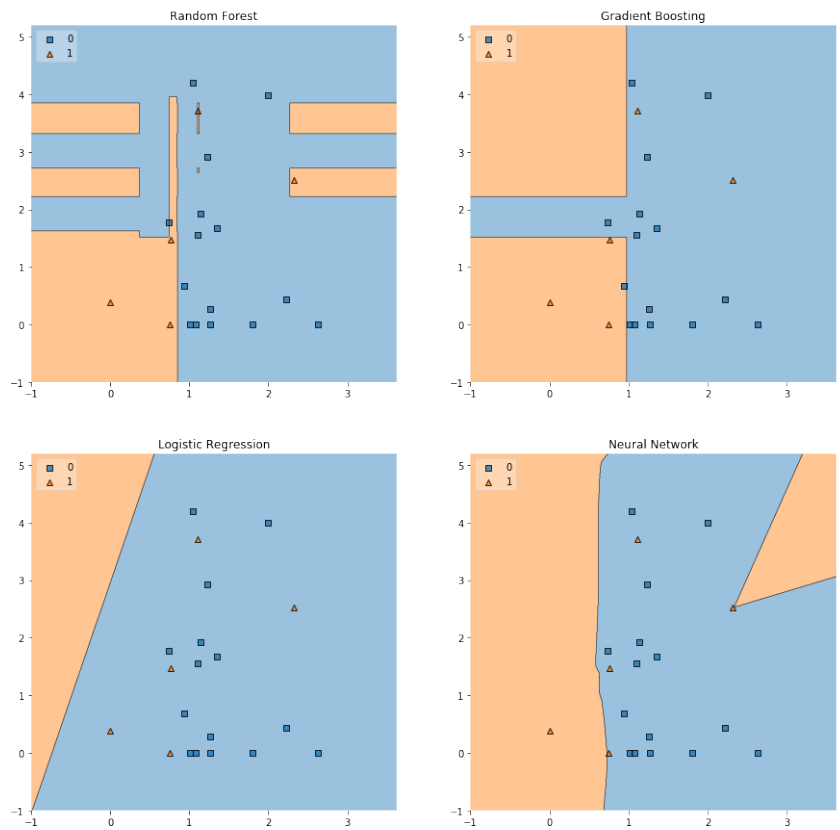

Figure 4: Decision Boundary of Different Classifiers on our Data Set

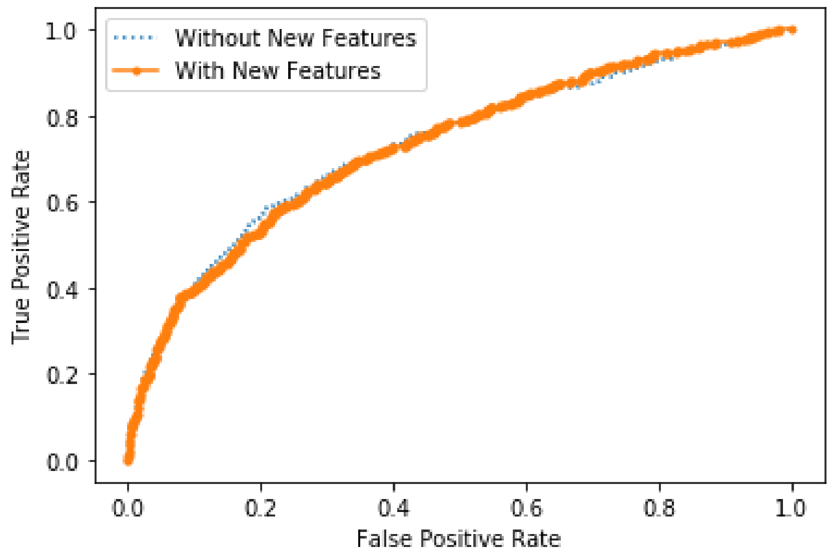

Figure 5: XGBoost on the "1st Year" data set with and without additional features of activity type and geo-location

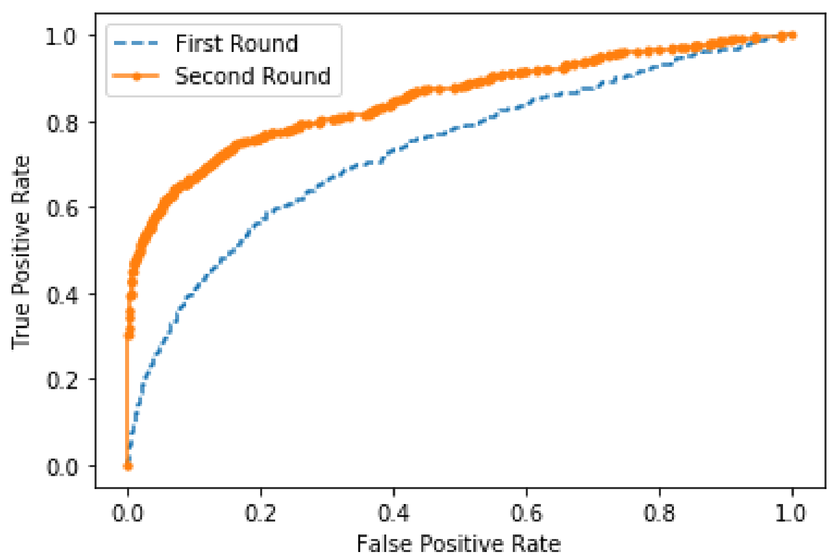

Figure 6: The figure shows the ROC curve of the Random Forest on the "1st Year" data set for the First vs Second Round

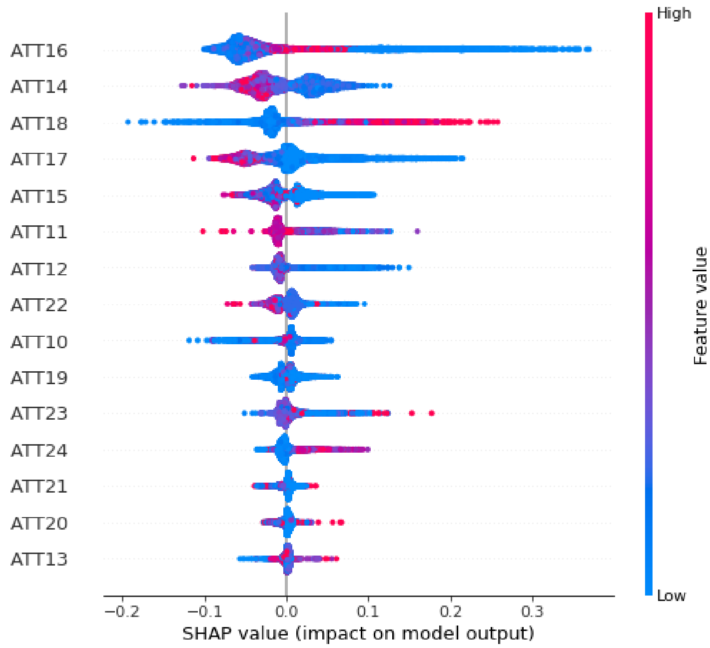

Figure 7: Attribute Importance

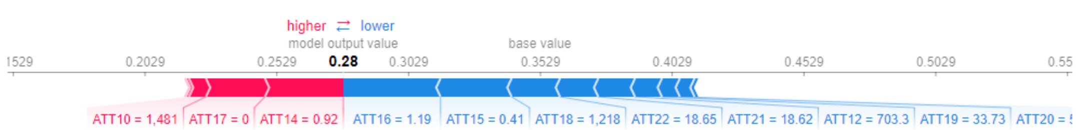

Figure 8: Attribute Importance Affecting One Company

by the former President of the Italian Companies and Exchange Commission (CONSOB), i.e., the authority of vigilance on the stock market and the banking system in Italy. Some classifications of the companies according to the geographic area of the headquarters, the type of activity according to the ATECO codes, and their revenues are presented in Table 4. The largest part of the companies is settled in the North of Italy and has revenues between 1 and 5 million euros. The activity type is more distributed, with a slight predominance of the ”Industry” category.

Table 5 gives the results of the prevision of bankruptcy related to the activity as provided by the ATECO codes. The probability of a company crisis is computed for the short (1 year), medium (3 years) and long (5 years) period, while the severity of the probability as low if the probability is under the 50%, medium if between 50% and 70% and high when greater than 70%. For each row, we report the total number of companies for each cluster (Industry, Commerce, Public, Service). Then, for each probability range is given the percentage related to the total number of companies in each cluster having the given probability of company crisis. The companies working in the ”Service” cluster are the ones with the largest percentage of medium and high risk, in particular considering the long term prevision, while the other sectors present better performances, in particular in the short term. Comparing the short term with the long term previsions, there is an increment of about 4% in terms of companies with a high probability of company crisis in the next 60 months. By considering their direct revenues and the indirect effects of an eventual crisis, we can estimate the impact of these companies of 30 billions of euros of direct revenues and 80 billions by also including the indirect effects, equal to about 3% of the Italian GDP. This proves the economic value of having middle and long term previsions in terms of economic impact.

Table 6 reports the same data, but clustered by geographic location. The companies are grouped according to the 4 standard Italian clusters (North East, North West, Center, and South). The largest number of companies with high company crisis risk are the ones in the Center and South of Italy, while the ones with the lowest are located in the North East. It is also worth mention how the companies in the North present also the largest number of companies with a low probability (up to 50%).

Regarding the size of the company in terms of yearly revenues (see Table 7), the companies are grouped in 4 clusters (less than 5 millions of euros, up to 10 millions, up top 15 millions and over 15 millions of euros). The results show how the companies up to 5 millions of euros are the ones with the largest percentage of high risk and also the ones with the lowest number of low risk. This aspect is particularly relevant moving from the short to the middle and long term, with about 5% with the companies moving to the middle and the high probability.

Given the recent COVID-19 crisis, we applied our Machine Learning DSS to the Italian case, by simulating the effect of the lock-down and the effects of the Italian Government

Segmentation	Cluster	Number of Companies
Region	North East	39775
	North West	53045
	Center	36724
	South	32058
Activity	Industry	69351
	Commerce	46524
	Public	3630
	Service	42097
Revenue	< 5 Mil	127009
	5 <= X < 10 Mil	17965
	10 <= X < 15 Mil	6519
	>= 15 Mil	10109

Table 4: Italian Companies Demographic/Revenue Info

Risk of bankruptcy	Activity	Count	Prob < 50%	50% <= Prob < 70%	Prob >= 70%
Short Term	Industry	69351	58%	18%	24%
	Commerce	46524	56%	19%	25%
	Public	3630	51%	23%	26%
	Service	42097	47%	21%	32%
Middle Term	Industry	69351	53%	19%	28%
	Commerce	46524	51%	21%	28%
	Public	3630	48%	23%	29%
	Service	42097	43%	22%	35%
Long Term	Industry	69351	52%	19%	29%
	Commerce	46524	50%	22%	29%
	Public	3630	46%	23%	31%
	Service	42097	42%	22%	36%

Table 5: Italian Companies Bankruptcy with regard to the activity

Risk of bankruptcy	Location	Count	Prob < 50%	50% <= Prob < 70%	Prob >= 70%
Short Term	North East	39775	60%	18%	22%
	North West	53045	57%	18%	25%
	Center	36724	49%	20%	31%
	South	32058	48%	22%	30%
Middle Term	North East	39775	56%	19%	25%
	North West	53045	53%	20%	27%
	Center	36724	44%	22%	34%
	South	32058	44%	23%	34%
Long Term	North East	39775	55%	19%	26%
	North West	53045	52%	20%	28%
	Center	36724	43%	22%	35%
	South	32058	42%	23%	35%

Table 6: Italian Companies Bankruptcy with regard to Company Location

Risk of bankruptcy	Revenue	Count	Prob < 50%	50% <= Prob < 70%	Prob >= 70%
Short Term	< 5	127009	51%	20%	29%
	5 <= X < 10	17965	64%	18%	18%
	10 <= X < 15	6519	65%	16%	19%
	>= 15	10109	65%	16%	19%
Middle Term	< 5	127009	47%	21%	32%
	5 <= X < 10	17965	60%	19%	21%
	10 <= X < 15	6519	61%	17%	22%
	>= 15	10109	60%	18%	22%
Long Term	< 5	127009	46%	22%	32%
	5 <= X < 10	17965	60%	18%	22%
	10 <= X < 15	6519	61%	17%	22%
	>= 15	10109	59%	18%	23%

Table 7: Italian Companies Bankruptcy with regard to Company Revenue (millions of euros)

law for financially supporting the companies. In this case, we focused on the Piedmont area, due to the possibility to have direct data and check the results with the help of the group of experts led by the former President of CONSOB and integrated by some policy-makers of the Regional Council of Piedmont. Moreover, the sample is representative in terms of company mix and revenues and presents a very favorable pre-COVID situation. Table 8 summarizes the characteristics of the sample. The first row reports the number of companies. As for the Italian case, the companies are SMEs with revenues between 1 and 40 millions. The other rows report the mean revenues and EBITDA (in K euros), the mean number of employees, and the mean number of shareholders. Our sample is responsible of the 49% of the GDP of Piedmont (65 billions of euros over 132) and has about 270000 direct employees in total. To simulate the situation pre and post COVID-19, we applied a simulation in which we decreased the revenues of each company of a percentage equal to 30% (estimation made by CONFINDUSTRIA, the main Company Association in Italy). As done by the Piedmont Regional Council and the Italian government, no differentiation concerning the revenues per sector is applied. We then apply the policy of providing the companies a financial support in the form of a loan granted by the Government equal to a given percentage of the previous year's revenue. We simulated a percentage equal to 10%, 20%, and 30% of the previous year's revenue. The risk is computed in the middle term (three years). The results of the simulation are reported in Table 9. The table reports, for the pre-COVID, the post-COVID without any public policy and the policy of the loans granted by the government with the three different percentages the number of companies (in the percentage of the total) with low (risk under 50%), medium (risk between 50% and 70%), and high (more than 70%) risk of bankruptcy, as well as the mean risk of all the 12707 companies. It is worth noticing that how the initial situation was quite good, with just 0.6% of the companies with high risk were less than 1% before the COVID. After the COVID, the high-risk ones triple, but the worst result is that the low-risk ones become just the 13.8% form the original 70.7%. This is due to two effects: the loss of revenues and that almost half of them were already on the border between low and medium risk. The policy of giving loans for a certain percentage of the revenues of the company not working properly if the percentage is low (10%), while with 20% and 30% the effect is more consistent. The 30% policy is giving a low deterioration of the general situation, due to the increase of the mid-term debt, being the financial support given as a very low interest rate to be refunded in a fixed time (5 years in Italy). Also in terms of mean risk over the full set of the companies, the best policy is the 20% one, with a mean risk coming back almost to the pre COVID values. We also performed an analysis of the effects on the economic system of the lock-down and the policy with a loan equal to 20% of the revenues, which we summarized in Figure 9. The index of unemployed workers is forecasted to increase 12%, with the Italian Temporary Lay-off increasing of 160% with respect to the previous year. Moreover, most of the companies would be out of the Basel III and the other short-term financial stress tests, leaving them in the impossibility to receive a loan form the banks and the bank system blinded in terms of evaluation of the future performances of their customer portfolios (Georg, 2011; Altman, 2020). The policy providing the grant for a 20% of the revenues would bring the high-risk companies back to

Companies	12707
Revenues (K euros	5005
EBITDA (K euros	452
Employees	22.58
Shareholders	3.03

Table 8: Piedmont companies main characteristics - companies with revenues between 1 and 40 millions

Risk of bankruptcy	Prob < 50%	50% <= Prob < 70%	Prob >= 70%	Mean Risk
Pre COVID-19	70.7 %	28.7 %	0.6 %	29 %
Post-COVID-19	13.8 %	84.6 %	1.6%	39%
Loan 10% revenues	15.8%	82.7%	1.5 %	40 %
Loan 20% revenues	27.7 %	71.5%	0.7 %	33 %
Loan 30% revenues	20.1%	79.0 %	0.9%	38%

Table 9: Piedmont Companies Bankruptcy pre-COVID, post-COVID, and after the financial support policy

the pre-COVID situation, but needing about 5 years to be back in the same situation than 2018, under the hypotheses of a reduction of the GDP of 10% in 2020 with an increase of the GDP of 6% in 2021, 5% in 2022 and the loan payment in 10 years. A pay-back of the loan in 5 years, as in the hypotheses of the Italian Government, might vanish the effect of the financial support, bringing the point of return of the investment to 8.5 years and stressing so much the companies to increase back the high and medium risk companies.

7 Conclusions and future developments

In this paper we considered the challenge of forecasting a company crisis by Machine Learning. The Machine Learning training is enhanced by a two-phase training procedure able to improve the performances of the Machine Learning. We showed how we are able, starting from operational and financial data, to accurately forecast the presence of a crisis up to 60 months. Moreover, we introduced our Machine Learning module in a DSS and we applied it to the Italian SMEs in order to analyze the Italian economic system and using the DSS as a support tool for validating public policies related to the economic shock due to the COVID-19.

Future developments include the introduction of additional data coming from other risk sources, as cybersecurity and seismic data, and to explicitly include in the Machine Learning module the dynamic evolution of the system and to include the presence of a

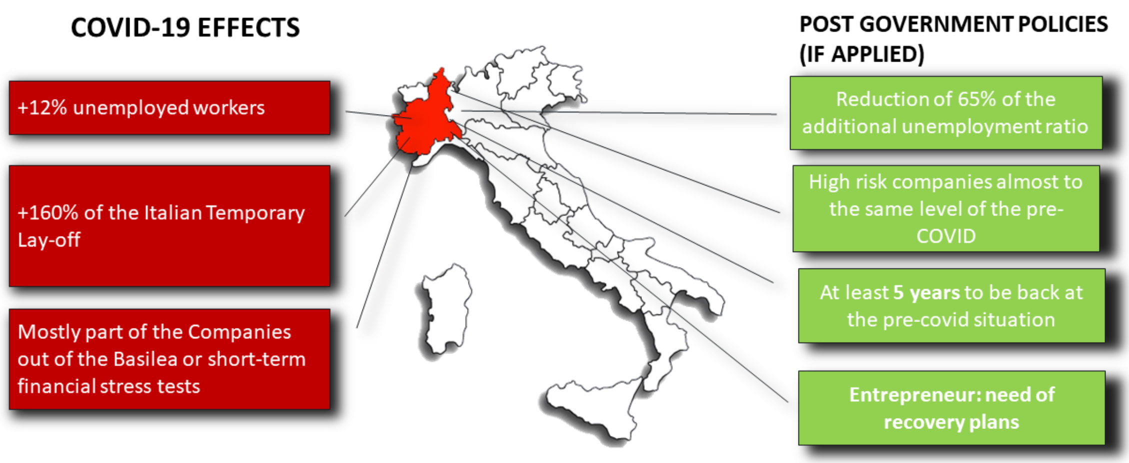

Figure 9: Summary of the post-COVID and the post Government policy (20%)

certain level of uncertainty by incorporating Extreme Value theory (Perboli et al., 2014).

Acknowledgments

While working on this paper, Guido Perboli was the head of the Urban Mobility and Logistics Systems (UMLS) initiative of the interdepartmental Center for Automotive Research and Sustainable mobility (CARS) at Politecnico di Torino, Italy and R&D Director of ARISK, a Spin-off of Politecnico di Torino.

References

Altman, E., Esentato, M., Sabato, G., 2020. Assessing the credit worthiness of italian smes and mini-bond issuers. Global Finance Journal 43.

Altman, E., Iwanicz-Drozdowska, M., Laitinen, E., Suvas, A., 2016. Financial and nonfinancial variables as long-horizon predictors of bankruptcy. Journal of Credit Risk 12 (4), 49–78.

Altman, E. I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The journal of finance 23 (4), 589–609.

Altman, E. I., 2014. Predicting financial distress of companies: revisiting the Z-Score and ZETA models. In: Bell, A. R., Brooks, C., Prokopczuk, M. (Eds.), Handbook of Research Methods and Applications in Empirical Finance. Edward Elgar Pub, pp. 428–456.

Altman, E. I., 2020. Covid-19 and the credit cycle. SSRN paper n. 3604361, DOI: 10.2139/ssrn.3604361.
URL https://ssrn.com/abstract=3604361

Altman, E. I., Danovi, A., Falini, A., 2013. Z-score models' application to italian companies subject to extraordinary administration. BANCARIA 04, 24–37.

Argenti, J., 1976. Corporate collapse; the causes and symptoms. McGraw Hill.

Balcaen, S., Ooghe, H., 2006. 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.

Baldi, M. M., Manerba, D., Perboli, G., Tadei, R., 2019. A Generalized Bin Packing Problem for parcel delivery in last-mile logistics. European Journal of Operational Research 274 (3), 990–999.

Barboza, F., Kimura, H., Altman, E., 2017. Machine learning models and bankruptcy prediction. Expert Systems with Applications 83, 405 – 417.

Barnard, J., Meng, X.-L., 1999. Applications of multiple imputation in medical studies: from aids to nhanes. Statistical methods in medical research 8 (1), 17–36.

Beaver, W. H., 1966. Financial ratios as predictors of failure. Journal of Accounting Research 4, 71–111.

Becerra, V. M., Galvão, R. K., Abou-Seada, M., 2005. Neural and wavelet network models for financial distress classification. Data Mining and Knowledge Discovery 11 (1), 35–55.

Begley, J., Ming, J., Watts, S., 1996. Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. Review of Accounting Studies 1 (4), 267–284.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. Journal of Machine Learning Research 13, 281–305.

Boritz, J. E., Kennedy, D. B., Albuquerque, A. d. M. e., 1995. Predicting corporate failure using a neural network approach. Intelligent Systems in Accounting, Finance and Management 4, 95–111.

Brédart, X., 2014. Bankruptcy prediction model using neural networks. Accounting and Finance Research 3 (2), 124–128.

Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140.

Brown, I., Mues, C., 2012. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications 39 (3), 3446–3453.

Bureau van Dijk, 2020. Aida - italian companies database.
URL http://aida.bvdinfo.com/

Burnham, K. P., Anderson, D. R., 2004. Model Selection and Multimodel Inference. Springer New York.
URL http://link.springer.com/10.1007/b97636

Cai, E., 2014. Machine learning lesson of the day – overfitting and underfitting.

Cantamessa, M., Gatteschi, V., Perboli, G., Rosano, M., 2018. Startups' Roads to Failure. Sustainability 10 (7), 2346.

Chen, J., Chollete, L., Ray, R., 2010. Financial distress and idiosyncratic volatility: An empirical investigation. Journal of Financial Markets 13 (2), 249–267.

Coats, P. K., Fant, L. F., 1993. Recognizing Financial Distress Patterns Using a Neural Network Tool. Financial Management 22 (3), 142–155.

Figini, S., Bonelli, F., Giovannini, E., 2017. Solvency prediction for small and medium enterprises in banking. Decision Support Systems 102, 91–97.

Freund, Y., Schapire, R., Abe, N., 1999. A short introduction to boosting. Journal-Japanese Society For Artificial Intelligence 14 (771-780), 1612.

Friedman, J. H., 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics 29 (5), 1189–1232.

Georg, C.-P., 2011. Basel iii and systemic risk regulation - what way forward? Working Papers on Global Financial Markets 17, Graduiertenkolleg 'Konstitutionelle Grundlagen globalisierter Finanzmärkte - Stabilität und Wandel'.
URL http://hdl.handle.net/10419/94477

Hambrick, D. C., D'Aveni, R. A., 1988. Large corporate failures as downward spirals. Administrative Science Quarterly 33, 1–23.

Hand, D. J., 2009. Measuring classifier performance: a coherent alternative to the area under the roc curve. Machine learning 77 (1), 103–123.

Hand, D. J., 2010. Evaluating diagnostic tests: the area under the roc curve and the balance of errors. Statistics in medicine 29 (14), 1502–1510.

Hilbe, J. M., 2009. Data Analysis Using Regression and Multilevel/Hierarchical Models. Journal of Statistical Software 30 (3).

Hillegeist, S. A., Keating, E. K., Cram, D. P., Lundstedt, K. G., 2004. Assessing the Probability of Bankruptcy. Review of Accounting Studies 9 (1), 5–34.

Kim, M.-J., Kang, D.-K., Kim, H. B., 2015. Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. Expert Systems with Applications 42 (3), 1074–1082.

Linden, H., 2015. Synthesis of research studies examining prediction of bankruptcy. Aalto University School of Business, 88.

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. In: Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Vol. 2017-Decem. Curran Associates, Inc., pp. 4766–4775.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable ai for trees. Nature Machine Intelligence 2 (1), 2522–5839.

Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D. K.-W., Newman, S.-F., Kim, J., et al., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering 2 (10), 749.

Luoma, M., Laitinen, E., 1991. Survival analysis as a tool for company failure prediction. Omega 19, 673–678.

Odom, M. D., Sharda, R., 1990. A neural network model for bankruptcy prediction. In: 1990 IJCNN International Joint Conference on neural networks. IEEE, pp. 163–168.

Ohlson, J. A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. The Journal of Accounting Researc 18, 109–131.

Ooghe, H., Prijcker, S. D., 2008. Failure processes and causes of company bankruptcy: a typology. Management Decision 46, 223–242.

Patrick, P., 1932. A comparison of ratios of successful industrial enterprises with those of failed firms. Certified Public Accountant 2, 598–605.

Perboli, G., Tadei, R., Gobbato, L., 2014. The Multi-Handler Knapsack Problem under Uncertainty. European Journal of Operational Research 236, 1000–1007.

Powers, D. M. W., 2011. Evaluation: From precision, recall and f-measure to ROC, informedness, markedness & correlation. Journal of Machine Learning Technology.

Shah, J. R., Murtaza, M. B., 2000. A neural network based clustering procedure for bankruptcy prediction. American Business Review 18 (2), 80.

Son, H., Hyun, C., Phan, D., Hwang, H., 2019. Data analytic approach for bankruptcy prediction. Expert Systems with Applications 138, 112816.

Upneja, A., Dalbor, M. C., 2001. An examination of capital structure in the restaurant industry. International Journal of Contemporary Hospitality Management 13 (2), 54–59.