UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M.FANNO"

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

"Data Analysis for Bankruptcy Prediction: The Wholesale and Retail Sector in Italy"

RELATORE:

CH.MO PROF. MICHELE FABRIZI

LAUREANDA: LAURA MARCHESINI

MATRICOLA N. 1206478

ANNO ACCADEMICO 2019 – 2020

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

Laura Marchenini

____

Index

INTRODUCTION

The world nowadays has become more and more dynamic, with the product life cycle shortening and starting a business in general becoming more accessible; thus, incrementing competition and hence the risks. Consequently, the framework in which entities are is less stable with respect to the past, with more firms arising and ceasing to exist at the same time. Moreover, with the development of globalization in the last century firms of different countries are highly connected, and this negatively reflects the shock of a particular market within one country to the whole world; this is proved by the several crises of the last decades. All this to say that running a business entails different kind of risks and difficulties that an entrepreneur must consider. Therefore, it is necessary for an entity to be sound and ready to face shocks and stakeholders need to be able to verify this soundness. In this context enters the bankruptcy prediction topic, as the tool to foresee situations of financial distress in time for a reorganization of the firm. This is useful not only for managers of the company, but also for stakeholders, such as lenders, that have to decide whether to give credit or not to an entity; or for investors, to decide if it is worth investing on a company that may in the near future be failed. It is for this specific reason, the purpose of evaluating in advance the condition of companies, that the focus of my empirical research is to be able to predict financial distress four years ahead, or considering a relatively long time horizon, and not just the year before, when the circumstances may be so severely compromised that nothing could be done.

This thesis focuses on the development of a model with the purpose of predicting bankruptcy, that can be defined as the legal process through which an entity declares its condition of distress, and that may lead to the cessation of the activity. The work is organized as follows. Chapter 1 is a dissertation about the bankruptcy procedure in Italy and its main alternatives, in light of the new changes introduced by D.lgs. 12 gennaio 2019 n. 14. Chapter 2 is divided in two main sections: the first is the classic literature review on the bankruptcy prediction topic, while the second is a theorical explanation of the most widely used models and those implemented for the empirical analysis, and subsequent literature related to each of them. In Chapter 3, data and methodology, the dataset creation and the source of data is elucidated, together with the choice of the predictors and the relative descriptive statistics on the sample. To conclude, Chapter 4 reports the results of the empirical analysis, or better, of the four different algorithms executed, namely Logistic Regression, K-Nearest-Neighbour, Random Forest and Neural Network; the last section, in particular, provides a comparison between models.

CHAPTER 1: BANKRUPTCY AND OTHER INSOLVENCY MEASURES

Bankruptcy has become a matter of high importance nowadays, especially after the several periods of crisis that hit our world in the last decades, both at country level and globally. For this reason it seems appropriate for the aim of this dissertation to make clear what this term means and what are the pre-emptive measures that can be taken to avoid this final conclusion of an entity; with a particular focus on the Italian framework, since the empirical work is based on Italian firms, and a rapid highlight of main differences with respect to other countries on the topic.

First of all, I would like to clear the distinct meaning of two terms that are very often confused and improperly used as substitutes in the spoken language: Bankruptcy and Default. The latter is a specific situation in which a debtor fails to meet its obligation, for example fails to pay a debt within the deadline; it can therefore indicate a once in a while condition. Bankruptcy, on the contrary, is a far more serious and severe status; it is the legal process with involvement of legal authorities, through which a firm declares its inability to pay off its debts and its general condition of financial distress, and the aim of which is to liquidate the debtor's properties in order to share the proceeding among creditors, to fulfil, at least partly, their claims.

This word comes directly from the Italian "banca rotta", literally "broken bank" but has its translation into what was called "Fallimento" in the Italian Law. Indeed, another distinction has to be made here, "Fallimento" is what we described above as Bankruptcy, while "Bancarotta" in the Italian language identifies a completely different situation, since it is considered a crime and leads to imprisonment. Indeed, it does not represent a status of an entity, but it involves some wrong and illegal action taken by the entrepreneur (or other subjects, for which we speak of "bancarotta impropria" opposed to "bancarotta propria" committed by the entrepreneur itself). The Italian Law distinguishes between two different events: "bancarotta semplice"(Art. 323 of Codice della Crisi d'impresa e dell'Insolvenza or CCI), the most simple situation in which the entrepreneur is guilty of excessive personal expenses, operations of pure luck denoting imprudence in the administration of the properties or delay in the opening of the bankruptcy procedure, thus worsening the financial conditions. These actions are punished with a period of imprisonment ranging from six months to two years. Opposed to this, we find what is called "bancarotta fraudolenta" (Art. 322 of CCI), which entails an opportunistic and fraudulent behaviour and is more severely punished, with imprisonment that can vary from three to ten years. Among the actions included in the article we find the concealment, dissimulation or dissipation of goods, in order to prejudice creditors, and the falsification or destruction of accounting books or other business records, with the aim of providing an unjust profit to himself or others. Thus, the main difference between the two relies on the behaviour and intention of the entrepreneur, which is "just" gross negligence in the first case, but becomes wilful misconduct in the second.

Bankruptcy procedure

The following section analyses the procedure of Bankruptcy in Italy, that as we stated before, coincides with "Fallimento". In this regard, it is worth to point out a new enabling act, Legge delega 19 ottobre 2017 n. 155, translated into the D.lgs. 12 gennaio 2019 n. 14, with which the legislator aims at accelerating, simplifying and uniforming all Bankruptcy procedures, from which the name "Codice della Crisi d'Impresa" or CCI that can be translated as Entity's Crisis Code (this substitutes the old "Legge Fallimentare"). This is more a formal than substantial reform, because most of the specific rules regarding the insolvency legal institutes are kept invariant, but it is introduced a unique procedure for the initial opening of Bankruptcy and similar practices, that once initiated, will refer to the old specific rules of the practice. The formal innovation is the substitution of the vocabulary "Fallimento" with the new "Liquidazione giudiziale" (that therefore coincides with Bankruptcy in the English language) in all articles of the Code; this also to cancel the negative connotation attached to the term "Fallimento". The new Code becomes effective from August 15th 2020, and the additional novelty that it introduces is the purpose of identifying a financial distress situation before it becomes too late, and before the only choice left is to file for Bankruptcy. Alternatives will be briefly analysed later.

To start a Bankruptcy procedure two conditions have to be met according to Art 121 of CCI: first a subjective prerequisite, stating that this regulation applies to commercial entrepreneurs or to those that carry out a commercial, craft or agricultural activity, being it a natural person or legal entity. Excluded from the Bankruptcy regulations are the State, public authorities, minor enterprises and agricultural enterprises. On the other hand, the objective prerequisite requires the existence of the state of insolvency, that is defined in Art. 2 of CCI as the status of the debtor that is expressed with non-fulfilment or other fact that demonstrate that the debtor is no longer able to regularly satisfy its obligations. This insolvency condition must not be confused with the "crisis" status, defined by the Italian legislator as the financial or economic distress that makes the debtor's insolvency likely.

The initial petition has to be made to the Court district where the debtor has its "main place of business", that is assumed to be where the legal head office results to be, according to the Registry of Businesses of the firm, or, alternatively, where the effective offices of the usual activity are. The petition can be proposed either by the debtor himself, or by one or more creditors, or by the Public Prosecutor, in any case he learns about the insolvency status of the firm; moreover, the new regulation also gives this power to all the administrative authorities that have the function of control and surveillance on the company. Several documents must be attached when filing the petition, such as the compulsory accounting and fiscal books of the last three years, a detailed description and evaluation of assets and debts, and a list of creditors with their claims, specifying also the classification (privileged or not and why). The Court then proceeds to the evaluation of the existence of the prerequisites for the Bankruptcy procedure and the formal validity of the petition, and, if so, declares the process open and contextually fixes the date for the hearing of all the parties involved; in this moment it also appoints a Receiver (what is known as "curatore" in Italian). The Bankruptcy declaration has immediate effect and commands the Automatic Stay, in other words the suspension and cancellation of any enforcement action of creditors on the debtor's estate.

The discipline has emphasised with the new Code the role of the Receiver and its professionalism, indeed it has to be chosen among those enrolled in the Register of Lawyers or Public Accountants or similar highly skilled figure. The Receiver is given important powers and duties, such as:

- Review and verify the creditors' claims
- Prepare and submit to the Court the Relation on the Causes of insolvency, the Relation about the rise of the crisis, in addition to periodic relations (every six months)
- Dispose (and liquidate) of all debtor's assets involved in the process
- Get access to all databases in order to acquire all the necessary tools to verify the firm's situation, that means not only the financial and economic situation, but also having the lists of clients and suppliers related to the entity
- Execute all the operations of the procedure under the supervision of the Judge and of the Creditors' Committee
- Keep informed the different parties involved

The whole process is under the jurisdiction of the Courthouse that appoints the Receiver and it is entitled to relieve it in any moment for just cause; it monitors each phase of the procedure and has the power to summon all the parties whenever considered appropriate in order to acquire information and clarifications from the figures engaged; lastly, it decides in regard of all the litigation that may arise. An intermediary figure between the Court and the Receiver is a delegated Judge (in Italian "Giudice Delegato", but no correspondence exists in other countries) that embodies the function of surveillance and control on the regularity of the whole procedure. Among the others, he has the duty of reporting to the Court, promulgate urgent measures for the preservation of the assets, summon the Receiver and Creditors' Committee when necessary, distribute the proceeding, revoke the role of the engaged parties.

The last important party engaged in a Bankruptcy procedure is the Creditors' Committee, nominated by the Judge within 30 days from the sentence opening the process. It can be composed of three or five members, chosen among creditors so as to represent fairly the quantity and type of creditors; for what concerns the type, it is worth mentioning that two main categories of creditors exist: the privileged ones, that have some reason of pre-emption, that are allowed to be satisfied first, and are identified usually with those that claim pledge or lien on one of the debtor's goods; and the non-privileged ("chirografari" in Italian), that can be satisfied only after the privileged once. Within the two categories, the legislator imposes the "par condicio creditorum", the equity in the monetary satisfaction, in other words each creditor has the same right on the debtor's holdings. The Committee has the fundamental task of controlling the Receiver and voting to express agreement or denial to his proposals, or when necessary; decisions are taken with majority of the Committee within 15 days from the receiving of the issue and must be extensively justified. As stated above, from the moment of the opening sentence of the bankruptcy procedure, each action of a creditor against the properties of the debtor are cancelled and any new enforcement is considered void. In regard of the order of distribution of the proceeds it must be highlighted that priority is given to post-adjudication claims ("crediti prededucibili"), then come preferred creditors, unsecured creditors and other subordinated claims, and at the end, if there is something left, this goes to shareholders.

Important are the effects of the Bankruptcy declaration on the debtor: in general, the rule provides the dispossess of the properties, since, as already mentioned, the administration of all assets and holdings related to the firm is under the Receiver's responsibility. An exception exists to this provision, indeed properties that are strictly personal and those contributing to essential needs of the debtor and his family, remain in property of the debtor and he remains entitled of their management.

Normally, the declaration of the starting of the liquidation procedure automatically implies the interruption of the enterprise's activities; but the legislator has introduced a waiver to this rule recently, namely that the usual activity can be carried on by the Receiver if authorized from the Court, and only if this will not cause any damage to the satisfaction of creditors; on the contrary, sometimes the continuation of the activity can effectively contribute to this final purpose. To conclude, the Bankruptcy procedure, as prescribed in Art. 233 of CCI, ends naturally when all creditors have been paid, or in alternative, when the usefulness of the continuation has been certified (this usually means that there are no funds enough to cover the expenses or to satisfy creditors). The final act is the cancellation of the firm from the Registry of Businesses, with which it is considered extinct.

Alternative tools to face crisis and insolvency

Now I would like to briefly go through other insolvency procedure, focusing on those that are seen as tools to avoid the severe consequences and costs of Bankruptcy declaration, so those tools that should help the firm recovering from financial distress, and that should be undertaken at an early stage, before the insolvency becomes severe and irreversible. I will exclude from this dissertation the "Forced Administrative Liquidation" ("Liquidazione Coatta Amministrativa") and the "Extraordinary administration of Large Enterprises in a state of crisis" ("Amministrazione Staordinaria delle Grandi Imprese") since these are treatment for the state of crisis of a minority of firms; in particular the former applies only to financial, banking and insurance companies, while the latter applies only to large enterprises, and considering that Italy is mainly composed of Small and Medium Enterprises, these procedures are used in very few cases. In addition to that, my empirical analysis is based on the trade sector, thus excluding most of the companies that the two legal institutes take care of. I will therefore focus on those treatments accessible for the vast majority of firms, and that represent a sort of alternative to Bankruptcy.

It is important to say that the new regulation has introduced a single, uniform petition for all these procedure in order to simplify and accelerate the timings; as I stated above, from a single format of petition, then follows a different regulation according to the insolvency procedure that is applicable, and that mostly remained unchanged. The other key and fundamental innovation of the Code is that these measures I am going to explain, are privileged by the legislator with respect to the opening of a Bankruptcy procedure, thus confirming the purpose of identifying at an early stage the crisis of firms, in order to be able to appropriately intervene and restore the situation. Practically this means that if two different petitions arrive, for example a bankruptcy filing from creditors and a pre-bankruptcy composition from a debtor, the latter has the priority. These important tools are:

- Pre-Bankruptcy Composition ("Concordato Preventivo")
- Debt Restructuring Arrangements ("Accordi di Ristrutturazione dei Debiti")

- Certified Turnaround Plans (“Piani Attestati di Risanamento”)

The procedure more similar and linked to the Bankruptcy process is clearly the pre-Bankruptcy composition, which is a judicial procedure that aims at avoiding the bankruptcy declaration through a sort of deal between debtor and creditors, that establishes the way in which creditors will be satisfied. Two different variations exist: the one in continuity of the activity and the one with liquidation of assets; the former is privileged from the legislator, and provides the continuation of the enterprise’s activity by the debtor (direct continuity) or by another appointed subject (indirect). Prerequisites for this procedure are that the debtor must be a commercial entrepreneur and all the supplementary dispositions that we saw for the bankruptcy procedure (subjective prerequisite), and the existence of the state of crisis or insolvency, recalling that crisis does not always imply insolvency (objective prerequisite). This procedure, as all the other tools to face a crisis or insolvency state, is differentiated from the Bankruptcy procedure for its willingness feature, in other words the debtor is the figure that can propose these kinds of petition, to show its own willingness to start such a procedure. The aim of the process is the total recovery of the firm and the return on the market as a competitive entity. After the demand of the opening of the procedure, the Court has to establish a deadline within which the debtor has to present a formal proposal for the plan of recovery, and in the meantime appoints a Judicial Commissioner.

The plan proposed by the entrepreneur must include an analytic description of the means and timings of fulfilment of the proposal, and must be feasible; indeed, the feasibility of the plan has to be certified by an independent expert (he must also verify, in case of ongoing activity, that this alternative is the best way to achieve reimbursement of credits, and that it does not cause any damage). The plan can provide for the repayment of creditors in any way (liquidation of assets, distribution of proceeding of the activity, but also transfer of shares) and can include as content: the restructuring of debts, the delegation of the administration to a third figure, the division of creditors in classes and the consequent different treatment among classes. The peculiarity of this procedure is that it is admitted a plan that does not provide for the entire satisfaction neither of privileged creditors, but a minimum percentage must be stated. In addition to the plan, the debtor has to present several documents such as a report of the financial and economic situation of the firm, description of assets, liabilities, creditors’ claims and mandatory accounting books. Alternative solutions of the recovery of the entity can be offered by creditors, and in the poll, all of these will be examined and taken into consideration. The new Code has extinguished meeting of creditors to favour the online voting. The alternative with majority of preferences will continue the procedure; in case of parity, the debtor's proposal has the priority. The approved proposal is then subject to the homologation by the Court, that is followed by the inclusion on the Registry of Businesses and determines the start of the execution phase of the pre-bankruptcy composition, in which the debtor has to manage the firm to ensure that the dispositions included in the plan are followed, all this under the supervision of the judicial Commissioner.

The procedure can be interrupted and turned into a Bankruptcy procedure in three cases:

- If the initial petition does not meet the criteria required or is considered not feasible by the Court;
- If no proposal reaches the majority of votes;
- If the homologation is denied by the Court;

It is useful to clarify the duties of the judicial Commissioner, as he carries out a fundamental role in this procedure: he verifies the list of creditors and debtors resulting from the accounting books, makes the appropriate modifications, writes an extensive and detailed report of the causes of distress highlighting and justifying if there exist crisis or insolvency, compiles an inventory of the debtor's holdings, and writes a full report of the voting meeting; in addition, he is the one that controls and supervises the debtor in the management and administration of the entity, during the whole execution phase.

The second tool for the recovery of firms in state of crisis is the Debt Restructuring Arrangement, which is a particular agreement between debtor and creditors concerning the restructuring of debt for the rehabilitation of an entity, that only requires homologation by the Court. Indeed, one of the differences with the pre-bankruptcy composition is that there is no supervision of the Court through any appointed figure, and after the homologation the Court has no role in the development and execution of the agreement; for this reason, it is considered by most a private deal between multiple parties. Two other elements that contribute to the private nature of the arrangements are that the procedure does not open the right of creditors on debtor's properties, and that there is no kind of provision for the "par condicio creditorum". Since it is considered a private agreement, the negotiation between parties have to be carried out in "bona fide". For it to be valid, the deal has to be signed by at least 60% of creditors and must provide for the full repayment of non-adhering creditors. The payment must occur within 120 days from the homologation for those credits already expired, or within 120 from the deadline for the others. As in the pre-bankruptcy procedure, it is on the debtor's initiative that the process has a start, which has to deposit the same kind of documents in addition to the plan, the existence of crisis or insolvency state as objective criteria for eligibility is required here too, and also the verification of feasibility of the independent expert is requested. Three main innovations are worth mentioning: the first is that the entrepreneurs entitled to propose the Debt Restructuring Arrangement can run either a commercial activity, or an agricultural or crafts one: the only activities excluded from this regulation are those of minor entities. The second innovation is that the final aim of the procedure is of course the recovery of the firm to its ordinary activities, but can also be the liquidation of the assets. Third, two different variations of the Agreement were introduced: "Facilitated Agreement" ("accordo agevolato"), that requires only 30% of approving creditors, but does not include the moratorium for credits; and "Agreement with Extended Validity" , which produces effect also on non-adhering parties, that must satisfy the following conditions:

- All creditors must have been properly informed;
- It must be the case of a recovery for continuation of activity (not a liquidation purpose, new criteria introduced by the legislator);
- The adhering creditor in each category must represent the 75%;
- The effect is extended to the non-favourable, only if in this way they will get no less than what they would get in a Bankruptcy process;

Last thing to specify, if the homologation is accepted, the executive phase starts and the procedure and execution of the agreements are left to the parties; if the homologation is denied, creditors have the power to demand the opening of a Bankruptcy procedure, but this does not happen as a matter of course after the homologation refusal.

The third instrument to face and solve financial distress is the Certified Turnaround Plan, that is a completely extrajudicial tool which purpose is to recover the debt exposure and to rebalance the financial situation. The differences with the previous measures are clear and evident: this is a private agreement between the debtor and one or more creditors (the effects of which only apply to the subscribers), indeed the Court has no role in the procedure, not even for homologation; in addition, the publication on the Registry of Businesses is not compulsory, as to avoid negative publicity. It is worth specifying that this procedure applies only to debtors that are in a state of crisis that is not irreversible, they have to be in temporary and not severe financial distress, thus the simplicity of the procedure. It is then evident that the aim of this tool is to restore the ordinary activity of the company, and for this reason cannot be used if the aim is the mere liquidation of assets and holdings. Some aspects of the previous procedures apply here too, for example the minimum element of the content of the plan, the attached documents to be presented and the feasibility verification by means of an independent expert, this time appointed by the debtor. With the new Code of Crisis and Insolvency (CCI) the legislator has introduced the obligation of the plan to be compiled in written form and to have firm date and analytic content. There is no indication about the timings in the Code, but in practice these plans are executed within five years, and this is understandable, since it applies to temporary crisis.

In all the three cases presented above (especially pre-bankruptcy composition and debt restructuring agreements), the debtor is allowed to obtain urgent finance, if he demonstrates that it is necessary to the ordinary activity of the firm, and that it helps to achieve the satisfaction of creditors; these funds, the receiving of which is subject to approval by the Court, will have 100% priority.

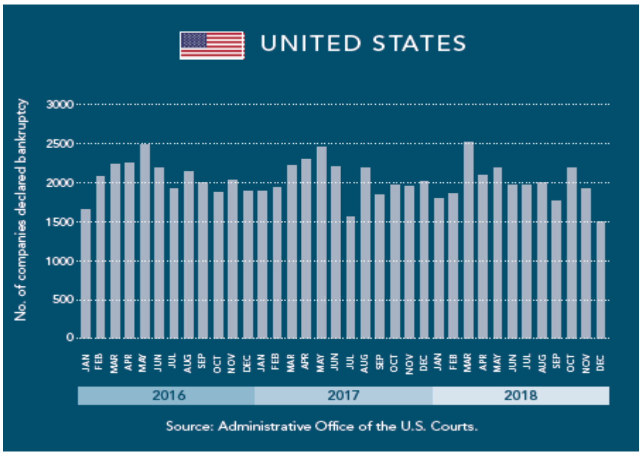
International framework

According to an article titled “The 5 largest economies in the world and their growth in 2020” by Prableen Bajpai published on nasdaq.com on Jan 22, 2020 (and confirmed by several other sources), the five world’s largest economies representing the 55% of world’s economy, classified by nominal GDP are, in order:

- U. S. (nominal GDP of $21.44 trillion, decline in growth rate expected, but growing)
- China (nominal GDP of $14.14 trillion, consistent growth expected)
- Japan (nominal GDP of $5.15 trillion, growth expected)
- Germany (nominal GDP of $3.86 trillion, growth expected)
- India (nominal GDP of $2.94 trillion, consistent growth expected, coming very close to Germany)

Now I will go through some bankruptcy data about some of these countries and Italy, to have a more complete and clear view of bankruptcy all around the world. I will, for each of the counties taken into consideration, report a graph taken from the most recent Global Bankruptcy Report, issued by “Dun & Bradstreet Worldwide Network” in 2019 , which data referred up to end of 2018, showing the evolution in the number of companies filing for bankruptcy in the last years. The company analysed data from 45 markets, and highlighted that “Forty-nine percent (49%) of all countries in the Global Bankruptcy Report saw a drop in the number of business failures in 2018 against the previous year". However, there is not a widespread positive trend, since 18 nations experienced a growing number of bankruptcies.

United States



The graph shows the number of business failures by month from 2016 to 2018 included. The total number of bankruptcies in 2018 was 24.000, representing a 3,1% decrease with respect to the previous year. Estimates expect this number to decrease even more in 2019.

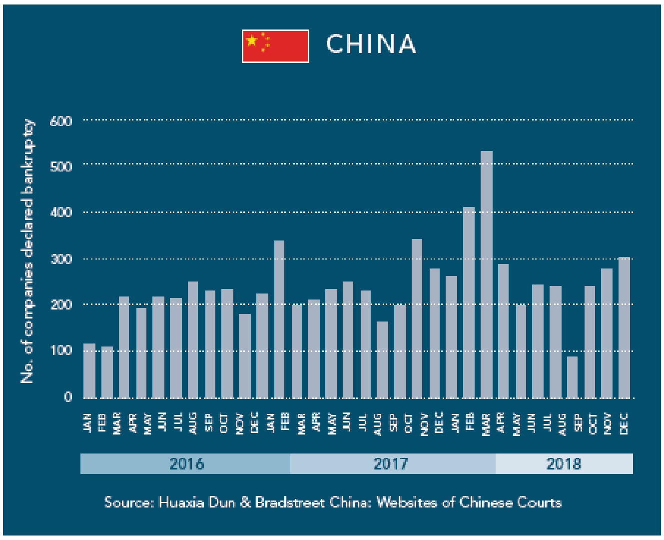
For what concerns the concept of Bankruptcy in the U.S., it is important to highlight that this term can apply both to businesses and individuals. All bankruptcies procedures (six in total) in the United States are managed at federal courts, and are regulated by a dedicated chapter in the U.S. Bankruptcy Code. Three of these are worth mentioning:

- Chapter 7 Bankruptcy, also known as "Liquidation bankruptcy", mainly used by individuals or small business; it only involves the liquidation of assets for the repayment of debts. All non-exempt asset (properties necessary to maintain basic standard of living are excluded) are liquidated to ensure the repayment of debts, in order: unsecured priority debts, for example tax debts and child support; secured debts and, lastly, normal unsecured debts. Under this procedure, most debts are discharged within some months, meaning that the debtor is released from any personal liability for payment. This is the simplest and fastest process, that offers to individual freedom from debts, but remains in records for 10 years, compromising the ability to access credit if necessary.
- Chapter 13 Bankruptcy, also known as "Wage Earner's Plan", used both by individual and businesses to reorganize finances under the supervision of the Court. The debtor can propose a detailed repayment of all debts based on the income he receives; the plan usually lasts for three or five years. The debtor each month commits a substantial amount of his income for the reimbursement, transfers the money to the trustee, that is entitled to distribute and take care of the repayment. This procedure is often used by individuals that want to be sure to maintain their home: indeed, no asset liquidation is

required, for this reason the fundament prerequisite is a stable, verifiable income and a level of debt under a certain threshold (varying every two or three years).

- Chapter 11 Bankruptcy, also known as “Reorganization”; this is the true bankruptcy procedure, the most complex and expensive, it can take several years and for these reasons is filed mostly by large businesses or those that can afford it. It still has the aim of restoring the ordinary activity of the firm and offer it a fresh start. It involves complicated restructuring plans, but during the whole procedure the debtor himself is allowed to run the business for ordinary operations, while extraordinary activities are subject to the Courts approval. There is no required income or debt limit to start this procedure.

China



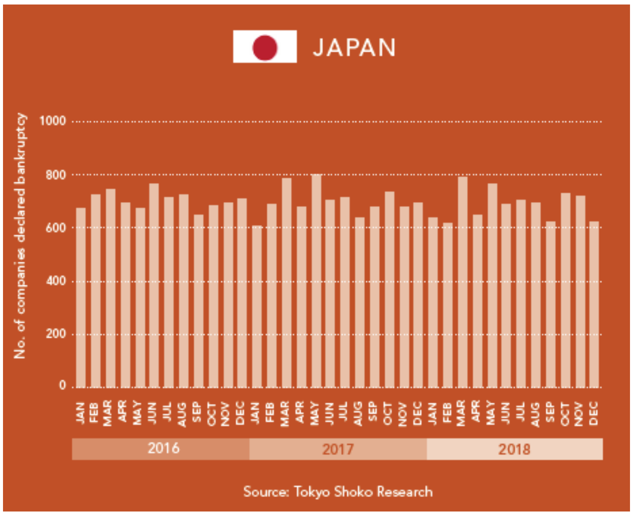
Bankruptcy in China is a relatively rare trend, resorting to formal bankruptcy procedures in courts is indeed uncommon, as we can see form data in the graph. However with the worryingly “trade war” with U.S.A. numbers has started to increase in 2018; moreover, it is reported by more recent sources that hundreds of thousands of firms (around 240.000), mainly in hospitality and retail industry, had to declare bankruptcy due to the severe effects of Covid-19, especially in some regions, thus we expect these numbers to grow a lot in 2020 (see Feng, 9th April 2020).

In regard of the available procedures, the People’s Republic of China (PRC) company law distinguishes between Dissolution procedures and Bankruptcy procedures. The first apply in some specific situations, such as the expiration of the terms of business operation, the decision taken by shareholders, as a consequence of mergers or demerges, or if the Court orders the dissolution. Generally, the company remains solvent during the whole process, and the administration of the firm is given to the liquidation group. Timings vary between six months and one year, but longer processes can take place. Moving to the pure bankruptcy procedures, characterised by an insolvency state, or better the inability of the debtor to pay off its debt and the insufficient assets to satisfy creditors, the PRC law provides for three different variations:

- Liquidation; the liquidation group is entrusted the management of assets and will declare the debtor bankrupt after the sale of all the holdings and if these are not sufficient to repay creditors. The firm will then cease to exist.
- Settlement; can only be initiated by the debtor, that can propose a settlement agreement subject to the creditors' approval and verification by the Court. The difference with respect to the previous case relies on the fact that if the plan is successful, the firm is restored to usual activity and the bankruptcy procedures are interrupted.
- Revival; in which debts are put under moratorium with the aim of reviving the firm. The debtor must act in accordance with the plan established either by himself or the administrator, the refusal leads to the opening of the liquidation procedure. This is the only legal arrangement that allows the debtor to manage the company himself, even if under the supervision of the Court; another advantage is the deferral of debt payments, as to give him a chance to get out of financial distress.

Japan

The number of bankruptcies for this country has been declining for nine years, and 2018 is no exception to this trend with 8.235 failed firms. However, this descending behaviour will eventually end, and 2019 may be the conclusive year since data for January indicate a rise of 4.8%.



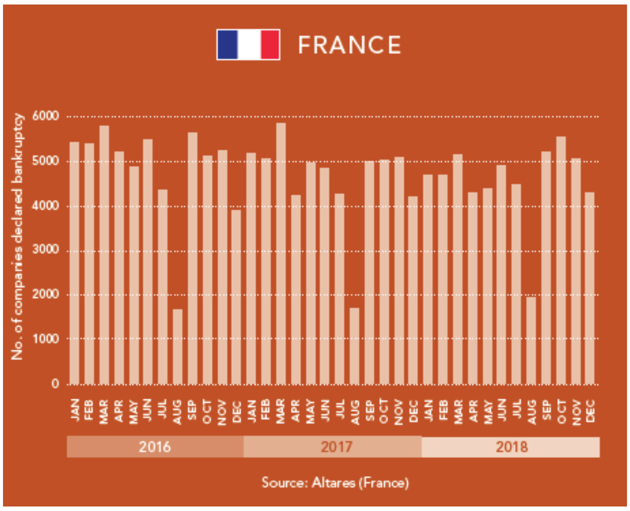
Speaking of the insolvency procedures available for Japanese firms, two are of main importance and can be applied to all kind of firms (as opposed to "Corporate Reorganization" and "Special Liquidation" that can only be used by stock firms):

- Bankruptcy itself, as the process of liquidating the assets of an insolvent debtor with the purpose of fulfilling creditors' claims, if possible. In this case, as in all other cases of proper bankruptcy in other countries, the debtor does not remain in possession of its estate, the management of which is given to a trustee, nominated by the Court.

Moreover, the business activity of the entity cannot continue, since the aim is the liquidation and not the recovery. To be eligible, the debtor can be either balance-sheet insolvent, meaning that the entrepreneur has a net worth deficit; or cash-flow insolvent, which identifies the inability to regularly repay debts. The process can take from several months to one year to be completed.

- Civil Rehabilitation; opposed to bankruptcy as its main purpose is to make the firm survive. In order to do this, a rehabilitation plan must be accepted by creditors and approved by the Court, and followed thereafter. Since this is a less severe procedure, different is the requirement for the debtor's status: indeed, there only must be the risk of bankruptcy, that will occur if no provisions are taken; the aim of this measure is to avoid this worsening to the liquidation. Another key difference is that the management of the entity, in general, keeps the control of the firm; the figure of the trustee is appointed only in special circumstances.

France



For what concerns Europe, I will skip Germany and rather focus on France, that among the most economically strong countries is ranked seventh (after UK), with a nominal GDP in 2019 of $2.707 trillion. Of all the European countries in the top ten it is the one having the highest number of firms going bankrupt per month, peaking sometimes up to and over five thousands. Numbers were relatively stable in the last years, even if a mild declining trend can be seen, but these are still remarkably high.

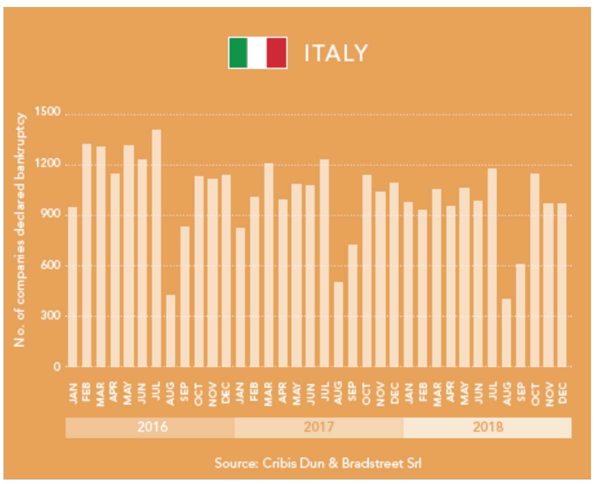
The French system in terms of crisis and insolvency procedures provides for four different variations, that entail different levels of severity in the initial condition of the debtor and in the consequences:

- Conciliation; the simplest and less severe case, applies to debtors generally not insolvent and can be initiated only by them. It is an out of court agreement between debtor and creditors, the only supervision consists on the presence of a Conciliator, but the entrepreneur keeps full powers for the management of the firm. Procedure can last up to five months.

- Safeguard procedure; a court process, that can be started only by the debtor, for the financial or corporate restructuring that ends with the approval of the plan by the Court. The debtor can continue to run the company, under surveillance of the judicial administrator. The opening declaration implies the automatic stay for creditors over the debtor's estate. It can last up to 18 months. To be eligible, debtor must not be cash insolvent, but it may be facing some difficulties that he is not able to overcome without help.
- Rehabilitation proceedings; the purpose of which is to restructure cash insolvent but viable entities; it leads either to the court approving the rehabilitation plan, or to the sale of the firm as a going concern. Even in this case the debtor still owns the company, and is helped in the administration by the judicial administrator. Maximum duration is 18 months.
- Liquidation; the proper bankruptcy procedure, from which there is no recovery and that determines the cancellation of the firm. It is indeed applied to insolvent companies for which there is no possibility of rehabilitation. It is carried out either through the individual sale of assets or (rarely) through the sale of the company as a going concern (in this case it continues to exist). The timings of the process may vary a lot, normally the goal is to conclude it within two years. In this case the debtor does not remain in possession of the firm and has no involvement in the management, that is given to the liquidator

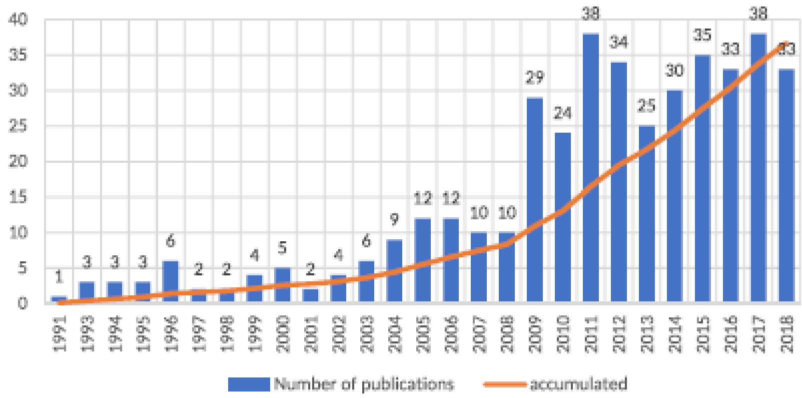
Italy

Even if I have already extensively spoken about the different insolvency procedures in Italy, I add here the graph showing the number of bankrupt firms, which at the end of 2018 could be considered in decline. Of course, with the recent emergency we will have higher number reported for the year 2020.



CHAPTER 2: LITERATURE REVIEW

Bankruptcy prediction is a highly addressed topic nowadays, indeed, lots of studies and researches have been published all over the world from the early 30s to present. These studies involve a variety of differences, either in the focus of the study in terms of type of firms and geographical area, or in the methods used for the prediction, the development of which is still ongoing with the creation of new techniques in order to reach an higher accuracy. Even if, as already said, the analysis of the topic began in the 1930s, authors started to focus on bankruptcy prediction after the 2008 financial crisis, as reported by Shi and Li (2019).



Graph 1: evolution in number of publications in bankruptcy prediction. Source: Shi and Li (2019)

The authors collected and analysed studies, researches and publications from 1968 (this year is a meaningful choice, as it is the release date of one of the main works on the subject by Altman) to 2018 using the Web of Science database. Having gathered fifty years of studies, they were able to perform different kind of analysis, including the evolution in number of publications per year. The last three decades are reported in the graph, and it is evident how the interest on the topic rose exponentially in the years right after the financial crisis. In particular, a peak is shown from 2008 to 2009, moving from 10 to 29 publications. This is not surprising, since with the financial crisis the need for a model able to foresee bankruptcy or, in general, financial distress became urgent. The models were created with the aim to reach a wide use, obviously by lenders that have to give credit to firms, but also by managers, so they can intervene on time, and investors or other stakeholders related to the companies. I will now go through the main studies on bankruptcy prediction, focusing on their importance and on the contribution they gave for the development of the following research.

The first study worth mentioning is the one by Beaver (1966), that as his predecessors used univariate analysis, which is the practice of focusing on individual ratios to predict bankruptcy, comparing ratios of failed and non-failed firms and deducing thresholds according to which companies are then classified. In particular, Beaver worked on a sample of 79 failed firms of 38 different industries, matched with as much non-failed firms according to asset size and industry. Failures for the first group occurred between 1954 and 1964; he then collected financial statement data of the five years preceding the bankruptcy declaration, computing for each entity 30 different ratios, that can be classified in six groups; of these 30, only one per group was selected (Cash-flow to Total Debt, Net Income to Total Assets, Total Debt to Total Assets, Working Capital to Total Assets, Current Ratio and the No-Credit interval).

Ratio	Prediction[a]
Cash flow to total debt[b]	Nonfailed > failed
Net income to total assets	Nonfailed > failed
Total debt to total assets[b]	Failed > nonfailed
Working capital to total assets	Nonfailed > failed
Current ratio	Nonfailed > failed
No-credit interval	Nonfailed > failed

Table 1: predictions of mean values of bankrupt and non-bankrupt firms. Source: Beaver (1966)

As we can see from the table above, the author was able to identify different characteristics among the two groups, in particular insolvent firms tend to have, five years prior to failure, lower cash-flow, lower reserves of liquid assets and more debt. The differences in mean values among the two groups become larger in the years closer to the failure, representing the deterioration of the situation. But this was just an analysis of the status quo, Beaver then computed a dichotomous test, with the aim of predicting the status of a company, not just observing it.

Ratio	Year before Failure				
	1	2	3	4	5
Cash flow / Total debt	.13 (.10)	.21 (.18)	.23 (.21)	.24 (.24)	.22 (.22)
Net income / Total assets	.13 (.12)	.20 (.15)	.23 (.22)	.29 (.28)	.28 (.25)
Total debt / Total assets	.19 (.19)	.25 (.24)	.34 (.28)	.27 (.24)	.28 (.27)
Working capital / Total assets	.24 (.20)	.34 (.30)	.33 (.33)	.45 (.35)	.41 (.35)
Current ratio	.20 (.20)	.32 (.27)	.36 (.31)	.38 (.32)	.45 (.31)
No-credit interval	.23 (.23)	.38 (.31)	.43 (.30)	.38 (.35)	.37 (.30)
Total Assets	.38 (.38)	.42 (.42)	.45 (.42)	.49 (.41)	.47 (.38)

Table 2: Dichotomous classification test. Source: Beaver (1966)

The table above shows the percentage of incorrect prediction performed by each ratio, provided that classification into one or the other group stems from the comparison of the own ratio of a firm with a threshold, identified as to give the minimum percentage of incorrect grouping. Thus, it can be seen as a measure of the predictive power of each of these six ratios: the lower the percentage of incorrect classification, the higher the accuracy of the ratio. He found that the best predictive ratio was Cash-flow to total Debt, that reported the lowest number of misclassifications throughout all the years prior to failure. This method is surely subject to many limitations, the first of which is that it adapts well to the sample collected, but it may not work in other circumstances; indeed, of the various studies carried out with this procedure, each author reports different results, and different ratios as to be the best for bankruptcy prediction. However, the reason why this research is extremely important is that, Beaver, aware of the disadvantages and limitations of these kind of process, was the first to suggest the use of multivariate analysis in failure prediction. He was able to understand that the use of several financial ratios together could have a higher predictive ability than a separate use. He also tried to approach this method, but with no encouraging results.

The person who seized the opportunity following his suggestion was Altman (1968), who was the first to use multiple discriminant analysis (MDA), and for this reason this publication is considered to be at the basis of all subsequent works. MDA is a statistical technique used to classify an observation into one of two or more pre-established groups, according to a set of individual characteristics; in this framework groups are bankrupt and non-bankrupt firms and the characteristics are the ratios computed from financial statements data. Altman focused on thirty-three manufacturing bankrupt firms, failed in the period between 1946 and 1965 in U.S.; which he associated with thirty-three non-failed firms by industry and size. The author then used a discriminant function, the one that has the power to classify one observation into one or the other group; the theorical version of the function is: $Z= V_1X_1 +V_2X_2+...+V_nX_n$; where $V_1,V_2,\dots,V_n$ are the coefficients and $X_1,X_2,\dots,X_n$ are the independent variables (ratios). The variables selected by the author, that will be known and used worldwide as reference thereafter, are:

- X_1= Working Capital / Total Assets
- X_2= Retained Earning / Total Assets
- X_3= Earnings Before Interest and Taxes / Total Assets
- X_4= Market Value of Equity / Book Value of Total Debt
- X_5= Sales / Total Assets

MDA then computed the related coefficients, so the final discriminant function used was:

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$$

This formula was then used to compute what will be later known as "The Z score" for each firm in the sample, by multiplying the coefficient for the corresponding value of the firm's ratios. In this way, as opposed to previous studies, the five ratios were considered simultaneously, and not separately, thus constituting an innovation.

	Number Correct	Per cent Correct	Per cent Error	n
Type I	31	94	6	33
Type II	32	97	3	33
Total	63	95	5	66

Table 3: Empirical results 1 year prior to failure. Source: Altman (1968)

This table represents the results of his study, reporting the accuracy of the model but also distinguishing between Type 1 and Type 2 errors. Overall, the model shows a very high accuracy, equal to 95% of correct classification of firms into the two groups, indeed of the 33 firms originally belonging to the two groups, 31 and 32 were correctly classified. Of the others, the author wanted to highlight the difference among two types of mistake: the first happens when a bankrupt firm is wrongly classified as solvent (Type 1); the second when a solvent firm is wrongly classified as bankrupt (Type 2). In general, the first kind of error is considered to be more costly and the one that has to be reduced as much as possible. Here we see that both error percentages are low, thus confirming the goodness of the model (which was expected, since the function was tested on the data from which it was derived). However, it has to be noted that, when the same analysis is computed with data referring to two years prior to failure (the table reports results one year prior to failure), the percentage of correct classification decreases a lot, for a total accuracy of 54%. To test the predictability of the model, two new samples were introduced, in order to verify if the discriminant function derived from the original sample could be applied to different frameworks and data, so as to ascertain the possibility to practically use the model in different contexts. The first test was computed on a sample of 25 failed firms and the model turned out to be very powerful, since it correctly classified as bankrupt 24 out of the 25 total firms, thus achieving 96% accuracy. What is more interesting, in my opinion, is the second test, in which data of 66 companies were collected, half belonging to 1958 and the other half to 1961; all these are non-bankrupt firms, over 65% of which suffered for financial distress for two or three years prior to the reference years.

			Predicted	
		Bankrupt		Non-Bankrupt
Non-Bankrupt Group Actual		14		52
	Number Correct	Per cent Correct	Per cent Error	n
Type II (total)	52	79	21	66

Table 4: Secondary Sample of Non-Bankrupt firms. Source: Altman (1968)

Results show that 52 out of 66 were correctly identified as non-failed, even if they experienced financial distress, reaching 79% of correct classification, which is a good result considering that this is a secondary sample, used only for predictive purposes. To look deeply into the situation, Altman discovered that if a firm had Z score higher than 2.99, it was correctly classified as non-failed, while if it had a Z score lower than 1.81 it was correctly classified as failed. Values of the score between these two number were not always crucial to achieve right classification; in fact, this interval constitutes a “grey area” and could not assure correct grouping. Better analysing the grey area, Altman found that the Z score that, when used as a cut-off point, was able to give the lowest number of misclassifications was 2.67. This elucidation was done having in mind that potential addressees of the model could be lenders, managers and investors, that typically do not have access to such data and that need a simple model to be implemented and not too difficult to understand. He was therefore thinking at the practical application of his model into the real world, where 2.67 could have been used as a general discriminant number to be applied directly, to achieve best classification without implementing complex approaches.

In the following years, it became clear that this model could only be applied to publicly traded firms, as it was required the Market Value to compute X_4; this meant that application for experts such as managers and lenders was limited. For this reason, Altman himself (2000) re-estimated the model, substituting the Market Value in X_4 with the Book Value and computing new coefficients, from which the following discriminating function:

$$Z' = 0.717X_1 + 0.847X_2 + 3.107X_3 + 0.420X_4 + 0.998X_5$$

This model specification led again to a high accuracy, 91% for Type 1 and 97% for Type 2; in addition, the grey area became a bit wider, with lower bound 1.23 and upper bound 2.90. It can be concluded that the two models seem to be very similar in terms of predictive power, although this second model is slightly less efficient.

During the following years several different techniques have started to be used, in particular from the 90s Artificial Intelligent techniques became more and more important and were proved to be a valid alternative to the classic multivariate analysis or logistic regression. As proof of this, the work by Tam and Kiang (1992) is one of the first in which a Neural Networks approach was used for bank failure prediction. A Neural Network is a system of algorithms that have the purpose of recognizing distinctive characteristics among a multitude of data, and for the aim here required, can be used to classify data into several groups according to the individual characteristics they are able to distinguish. This process is said to resemble the functioning of the human brain, and in particular of aggregates of neurons, from which the name Neural Networks. Speaking of the mentioned study, the authors used five different techniques, both statistical (MDA, logistic regression and k Nearest Neighbour) and intelligent (neural networks and ID3, that stand for Iterative Dichotomiser 3 and is a decision tree technique), and compared them in terms of results. The sample was composed of Texas banks, failed in the period 1985-1987, matched with non-failed banks according to asset size, number of branches, age and charter status. Data were referred to one and two years prior to failure; a total of 19 ratios were computed for each company, but for the different methods involved, not all of them were used, only those that came out to be significant, therefore number of ratios involved varied in the five methods implemented.

	Percentage (%)					
	One-year Prior			Two-year Prior		
Model	I	II	T	I	II	T
DA	0.0	22.0	(11.0)	10.2	1.7	(6.0)
Logit	8.5	6.8	(7.7)	13.6	13.6	(13.6)
1NN	37.3	23.7	(30.5)	32.2	32.2	(32.2)
3NN	35.6	25.4	(30.5)	37.3	32.3	(34.8)
ID3	10.2	5.1	(7.7)	13.5	5.1	(9.3)
Net_0	5.0	11.0	(8.0)	10.2	11.9	(11.0)
Net_{10}	0.0	7.6	(3.8)	6.7	10.2	(8.5)

Table 5: Misclassification Rates of Different Techniques with Training Sample. Source: Tam and Kiang (1992)

	Percentage (%)					
	One-year Prior			Two-year Prior		
Model	I	II	T	I	II	T
DA	18.2	13.6	(15.9)	30.0	5.0	(17.5)
Logit	31.8	4.5	(18.2)	15.0	0.0	(7.5)
1NN	40.9	4.6	(22.8)	20.0	25.0	(22.5)
3NN	36.4	9.1	(22.8)	30.0	10.0	(20.0)
ID3	22.7	18.2	(20.5)	40.0	5.0	(22.5)
Net_0	31.8	4.5	(18.2)	20.0	12.6	(16.3)
Net_{10}	18.2	11.4	(14.8)	2.5	20.0	(11.3)

Table 6: Misclassification Rates of Different Techniques with Hold-Out Sample. Source: Tam and Kiang (1992)

1NN and 3NN correspond to the 1 Nearest Neighbour and 3 Nearest Neighbour techniques, meaning, for example, that one observation is classified into one specific group based on its 3 more similar observations; in other words, when a firm has to be classified, the model looks at its characteristics (ratios) and identifies the 3 firms that have the most similar features and thus puts them all together in the same group. Net_0 and Net_{10} are the two different neural networks employed, to be precise, in the first there is no intermediate level between the input and output layer, while in the second we have one intermediate layer between input and output, composed of 10 hidden units. In the first table, that reports the results based on the training sample, which were expected to be good, it was evident that Net_{10} outperformed all the other techniques, given that it showed the lowest level of misclassification in total, third column (the first and second column are displayed to distinguish among type 1 and type 2 error). The other key evidence is that all the models were able to achieve quite good and acceptable results, except for the two computed with the k Nearest Neighbour technique. To test the predictive ability of these models, a Hold-Out sample was collected, in order to verify if models could be used in different framework with respect with the one from which they were derived. Table 6 provides these results: again, Net_{10} outperformed all the others, both in term of total misclassifications and of type 2 errors. The second-best classifier turned out to be MDA, with 15,9% of misclassification, followed by Net_0, Logit and ID3. Again, 1NN and 3NN did not seem to provide a good predictive ability. For what concerns the two-year period, logit appeared to be the best, followed by the Neural Network method with 10 hidden units. Overall, from these results the authors concluded that Neural Networks, if appropriately specified, represented a valid alternative to the classic statistical methods. Moreover, in this particular study, they performed even better; still, classic statistical method cannot be considered overwhelmed. In line with these conclusions, several other researches were carried on, developing new algorithms for bankruptcy prediction and confirming the superiority of machine learning techniques over statistical methods, which anyway lead to acceptable prediction: MDA and logit regression indeed, in all the following studies, were proved to have high predictive powers (see Bellovary, Giacomino and Akers 2007, Table 6).

After the development and the diffusion of intelligent techniques, many researchers tried to assess if those were better than classic statistical methods; this is the case of Chen (2011), that used a variety of different models, including:

- Linear Discriminant Analysis, or LDA, (statistical)
- Logistic Regression, or LR, a regression in which the dependent variable is dichotomous (statistical)
- C5.0, a Decision Tree algorithm, meaning that the independent variables are used to split the sample in progressively smaller sub-groups. C5.0 in particular uses boosting. (artificial intelligent)
- Classification and Regression Tree, or CART, uses DT algorithms which are a set of if-then conditions (artificial intelligent)
- Self-Organizing Map, or SOM, a clustering method (artificial intelligent)
- Learning Vector Quantization, or LVQ, a relatively simple algorithm composed only of one input layer and one output layer (artificial intelligent)
- Genetic Algorithms, or GA, that performs an optimization process in four phases, looking for a global optimum (artificial intelligent)
- Particle Swarm Optimization, or PSO, which searches among a population of individual, called particles, that are updated in each iteration (artificial intelligent)

For the empirical phase, Chen collected data of 200 firms failed between 2000 and 2010 from the Taiwan Stock Exchange Companies (TSEC); he then selected a total of 8 independent variables to be used as inputs for each algorithm, including financial and non- financial ratios, and a macroeconomic index. He executed all the statistical and artificial intelligent techniques, computing for each model some measures in order to compare the results. These measures are:

- $Overall\ Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$ as the percentage of correctly classified firms;
- $Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ within group accuracy, the higher it is, the lower the type 2 error;

- $Sensitivity = \frac{\text{TP}}{\text{TP+FN}}$ representing how well a classifier recognizes abnormal records;
- $Specificity = \frac{\text{TN}}{\text{TN+FP}}$ representing how well a classifier recognizes normal records;

Results computed two, four, six and eight quarters before failure are displayed below:

Algorithms		Overall accuracy				Precision				Sensitivity				Specificity			
		2	4	6	8	2	4	6	8	2	4	6	8	2	4	6	8
Statistical	LDA	75.47	80.81	80.3	82.32	68.75	85.41	84.84	92.68	88	77.35	77.77	76.76	64.28	84.78	83.33	90.76
	LR	79.25	77.78	80.3	78.66	71.87	84.44	85.93	89.02	92	71.69	76.38	73.73	67.85	84.78	85	86.15
DT	C5.0	86.79	81.25	78.79	78.75	78.12	79.06	84.37	86.84	**100**	85	75	73.33	75	77.5	83.33	85.71
	CART	84.91	83.75	78.03	76.87	75.75	82.92	85.24	85.33	**100**	85	72.22	71.11	71.42	82.5	85	84.28
ANN	SOM	90	80	77.5	82.5	94.11	85.57	86.27	94.28	84.21	79.48	68.75	73.33	85.23	82.5	87.5	**94.28**
	LVQ	87.5	82.5	79.16	83.12	85	84.21	**95.34**	94.36	89.47	80	64.06	67.67	85.71	85	**96.42**	**94.28**
Other	SVM	90	87.5	85	84.37	94.11	87.5	89.65	89.15	84.21	87.5	81.25	82.22	**95.23**	87.5	89.28	87.14
Evolutionary	GA-SVM	92.5	91.25	86.66	91.87	94.44	90.24	87.5	93.25	89.47	92.5	87.5	92.22	**95.23**	90	85.71	91.42
	PSO-SVM	**95**	**93.75**	**87.5**	**93.12**	**94.73**	**92.68**	87.3	**94.38**	94.73	**95**	**88.7**	**93.33**	**95.23**	**92.5**	86.20	92.85

Table 7: Classification Results. Source: Chen (2011)

In terms of overall accuracy, it seems clear how the artificial intelligent techniques outperformed the statistical, with the evolutionary approaches that showed the best results in all quarters (PSO being the best among all). In terms of precision, again the PSO evolutionary approach turned out to be the best (in three out of four quarters) with highest precision rates, that implies a lower chance of misclassifying a solvent company as insolvent. For what concerns the sensitivity, we can see a 100% sensitivity rate for the Decision Tree approaches two quarters prior to failure, but going back in time, this rate decreases a lot, and since the aim of prediction analysis is to foresee financial distress before it is too late to find a remedy, focusing on the last columns, the Particle Swarm Optimization approach seems to be outperforming all the others. In regard of specificity, intelligent approaches in general performed better than statistical approaches, some in the short run, some in the long run. The author, after executing additional tests, concluded that the two evolutionary methods were the best; but also highlighted that statistical models apply better to large samples, while the development of a specific artificial intelligent algorithm is more suitable to small samples, as it can better capture the feature and peculiarities of companies. Therefore, it seems again that no general rule exists to establish that there is a better method that can be applied in any context, but it is sure that intelligent approaches are a valid and extremely effective alternative to statistical analysis.

For what concerns the Italian framework, and moving to more recent periods, two studies are important for different reasons. The first is a research by Giordini (2014), that considering that many of the previous studies were focused on large or medium firms and feeling the lack of small and medium enterprises (SMEs) inclusion, decided to conduct a study on these kind of firms, based on Italian database, since these represent the vast majority of the economy in Italy, but also on other countries. He highlighted that SMEs have peculiar characteristics with respect to large and publicly traded companies, and for this he found it appropriate to develop a model tailored to them. Moreover, he realized that for small entities, a type 2 error (classifying as defaulted a solvent firm) may be worst with respect to type 1, due to the fact that it would lead to high difficulties in accessing credit from lenders and would also ruin the image with no justified reason. In practical terms, he used both logit regression and machine learning techniques, namely genetic algorithms (GAs) and support vector machines (SVM), then he compared the results. Genetic algorithms are machine learning methods, adaptive algorithms capable of perpetual innovations that have the advantage of extracting rules and results that are easy to understand. They resemble the Darwinian evolution concept, and can be broken down in four phases: Initialization, that is the selection of the population; Selection of Better Individuals, that are the observations with the highest fitting values, so those that are considered to be the best and that will be copied onto the next generation to propagate their features; Crossover, the phase in which two parent strings (those with the best feature) are selected and combined into a child string; Mutation, the process used to maintain genetic diversity from one population to the other. With this complex process, genetic algorithms can create new solutions (child strings) that have features that were not present before in the original sample, and have the ability to innovate and develop to find the best solution, whenever new observations are used as inputs. In practice, since the inputs are financial ratios values, GAs find and express rules according to which observation are classified, using cut-off measures for each ratio. An example of these rules could be:

IF ratio1 is lower than (or higher than) threshold1, AND ratio2 is higher than (or lower or equal to) threshold2 …. AND ratio N is higher than (or lower than) threshold N, THEN FIRM IS DEFAULTED.

In order for a firm to be classified as failed, all the conditions must be simultaneously satisfied; the number of conditions is of course equal to the number of ratios used for the prediction.

The author collected an overall sample of 3100 companies from the CERVED database, 1500 of which were defaulted; data collected referred to three, two and one year prior to failure. This sample was randomly divided into a training sample (2170 firms, 1050 of which defaulted), and a hold-out sample to verify the predictive ability of the model (930 firms, 450 of which defaulted). It is worth to point out that a total of 8 financial ratios were selected as inputs and that GAs extracted six different rules used to classify companies as failed.

Model	Real group membership		Predicted group		Firms correctly (incorrectly) classified
			1	0	
Gas	Defaulting firms	1	78.8	21.2	71.5 (28.5)
	Non-defaulting firms	0	35.8	64.2	
SVM	Defaulting firms	1	77.1	22.9	69.5 (30.5)
	Non-defaulting firms	0	38.1	61.9	
LR	Defaulting firms	1	76.7	23.3	66.8 (33.2)
	Non-defaulting firms	0	43.1	56.9	

Table 8: predictive accuracy of the three model, 3 years prior to failure. Source: Giordini (2014)

The above results show clearly that genetic algorithms were found to be superior in terms of predictive power with respect to logistic regression and support vector machine; in particular the increased prediction accuracy rate is equal to 2% if comparing GA with SMV and equal to 4,7% if comparing it with logistic regression. The same result (superiority of GA over the other two techniques) was found considering two and one year prior to failure. In addition, Giordini also computed the same comparison when dividing the sample in four subsamples according to their asset size: again, he found that genetic algorithms had a higher predictive power than logistic regression and support vector machine; thus leaving no doubt on the superiority of this method, at least in small and medium enterprises context.

Another important study based on the Italian context is the one by Madonna and Cestari (2015), that is interesting because instead of trying to create a new intelligent technique as is the major trend in literature, they preferred to focus on the existing methods and tried to assess which was the best and why for their data. It is indeed common to develop new machine learning technique, but the authors believed that there was already a sufficient number of models with high predictive ability, so there was no reason to spend time implementing new ones that could lead approximately to the same result, but that was better to try to adapt the existing ones to their sample. In particular, they decided to compare results of three different multivariate models, namely: Altman's Z'-score (cited above, the one designed for private companies); Alberici's Z-score (1975),who tried to replicate Altman's model in the Italian context; and Bottani, Cipriani and Serao's discriminant function (2004). The authors performed their analysis on Emilia Romagna's firms, in two phases: in the first they used sample containing only bankrupt companies, while in the second they used a mixed sample of healthy and bankrupt firms. Hence, phase one was computed on a sample of 323 firms, that had gone bankrupt in the years 2012-2014; five years prior to failure data were gathered. Results of each model are shown below:

		Altman's Z score		**Alberici's Z score**		**Bottani, Cipriani, Serao**	
Years prior bankruptcy	Sample	% correct classifications	% erroneous classifications	% correct classifications	% erroneous classifications	% correct classifications	% erroneous classifications
1	113	98,23%	1,77%	3,54%	96,46%	83,19%	5,31%
2	301	98,01%	1,99%	0,00%	100,00%	68,11%	7,31%
3	323	97,83%	2,17%	0,00%	100,00%	55,11%	10,84%
4	308	94,48%	5,52%	0,00%	100,00%	49,68%	13,31%
5	286	94,41%	5,59%	0,00%	100,00%	47,20%	16,43%

Table 9: reliability results of, in order, Altman, Alberici and Bottani. Source: Madonna and Cestari (2015)

Looking at the result the striking evidence is that Alberici's model could not be used for prediction in this sample, since it gave 0% correct classifications most of the time, thus this model was discarded for the second analysis. Among the other two, Altman's model performed better.

For the second phase, the mixed sample was used, and only two of the three models, those being Altman's computed both with cut-off and considering the grey area and Bottani's (only grey area); three measures were computed in each of the three cases, namely total accuracy, specific accuracy (specific to the group, thus dividing bankrupt and non-bankrupt firms) and type of error. Not reporting all the table of the result since too dispersive, it suffices to say that:

- For Altman's model using the cut-off point, total accuracy varied from 94.5% in the first year prior to failure, to 74.2% in the fifth year; specific accuracy was higher for insolvent firms (from 99% to 95.8%) with respect to healthy firms (from 90% to 52.6%); type 2 error was significantly higher than type 1 error, the former increasing from 10% to 47% five year prior to failure, while the latter remained always within 5%. Altman's model using the grey area was always underperforming with respect to the cut-off method, both in terms of total and specific accuracy; however, type 2 error was less severe.

- For Bottani, Cipriani and Serao's model, that only imply the grey area specification, a 90% total accuracy was found in the first year prior to failure, but this dramatically decreased in the long run, reaching only 46% in the fifth year; on the other hand, both erroneous classifications and uncertain classification increased in the long run, the latter reaching 34% five years prior to failure, representing the lower precision of this method in early prediction. In regard of the specific prediction, this models seemed quite accurate in classifying healthy firms, with accuracy varying from 97% to 72%; the most important difference with Altman's model is that type 2 error here was less severe than type 1 and was relatively low, ranging from 1% to 12%; on the contrary, type 1 error was better is Altman's model.

The authors therefore concluded that Altman's model computed with cut-off and Bottani's model are the best when taken out of their framework, meaning that their result and threshold can be generally applied, also in such a different context as the Italian. In addition, it is worth to point out that these methods were able to achieve high level of accuracy in bankruptcy prediction, thus supporting the idea that there is no need for development of new techniques, since there exist methods that can be widely used.

Lohman and Ohliger (2020) addressed the problem of bankruptcy prediction from a totally different point of view: with their recent study, they aim at proving that also qualitative information, in addition to quantitative ones represented by ratios, can be successfully used in models. What they focus on is the distinction between bankrupt firms and firms that are financially distressed but that will remain solvent; indeed, the majority of previous researches had the purpose of discriminating among insolvent and healthy companies, and often resulted in not being able to classify a portion of the sample, those being part of the grey area. In practice, they collected data of 117 failed German companies, and matched them with as much companies in financial distress, but that remained solvent, according to six measures taken from financial statements (applying propensity score matching). They then used the qualitative information in Annual Reports to perform the distinction, for example structural and linguistic characteristics or the time passed between the balance sheet date and the date on which the Annual Report was released. All the independent variables used are shown below since they are quite interesting:

Words_Risk	Number of words in the risk report
Words_Rest	Number of words in the annual report, excluding the risk report
Words_Rel	*Words_Risk*/(*Words_Rest* + *Words_Risk*)
LIX_Risk	LIX of the risk report
LIX_Rest	LIX of the annual report, excluding the risk report
LIX_Rel	*LIX_Risk*/*LIX_Rest* − 1
Sentiment	Weighted average of the positive and negative words
Pos_Words	Proportion of positive words among the total number of words in the annual report
Neg_Words	Proportion of negative words among the total number of words in the annual report
Delay	Number of days between the balance sheet day and the release date of the annual report
Risk_yes	Dummy variable; *Risk_yes* = 1 if a company states in its annual report that there are existential risks endangering the company as a going concern
Risk_no	Dummy variable; *Risk_no* = 1 if a company does not state in its annual report that there are existential risks endangering the company as a going concern
Restr	Dummy variable; *Restr* = 1 if the annual report mentions recently implemented, ongoing, or planned restructuring
Risk_Syst	Dummy variable; *Risk_Syst* = 1 if the annual report mentions the existence of a risk management system
Envir	Dummy variable; *Envir* = 1 if the annual report mentions the company's efforts to protect the environment

Table 10: list of independent variables. Source: Lohmann and Ohliger (2020)

The authors found out that the Annual Report of bankrupt companies was on average longer, that the complexity in language of their report was lower with respect to solvent firms, and that they showed less negative sentiment, containing fewer negative words. This is a bit puzzling, but is consistent with the "Management Obfuscation Hypothesis", that is the idea that when close to bankruptcy, management of companies try to hide the severe situation by concealing the true risk the firm is facing, while in solvent firms, the situation is truly reported and thus reflects the distress they are facing. This to say, that more negative words appeared in solvent but distressed firms than in companies really close to bankruptcy, and that the description of the level of risk was often hidden by management of these latter companies. (Indeed, only 22% of bankrupt firms stated the existence of risk, while 19% stated they were incurring no risks for the continuation of the activity). After that, Lohman and Ohliger decided to apply five different GLMs (generalized linear model, meaning a regression), to assess the effect of these variables in discriminating between bankrupt and solvent companies. Some of these include the qualitative variables, some do not, in order to understand if and how much the presence of these variables helped the analysis. They do concluded that these qualitative information, derived analysing the Annual Report, helped to classify firms more accurately, in contrast with the model only containing accounting-based ratios, that was found to be not able to classify these firms; in fact, as already said, usually those information have been used to distinguish among healthy and bankrupt companies, leaving aside the category of the distressed but solvent entities. The authors concluded that in future, to better predict what the status of a firm will be, both accounting-based and qualitative linguistic information shall be used.

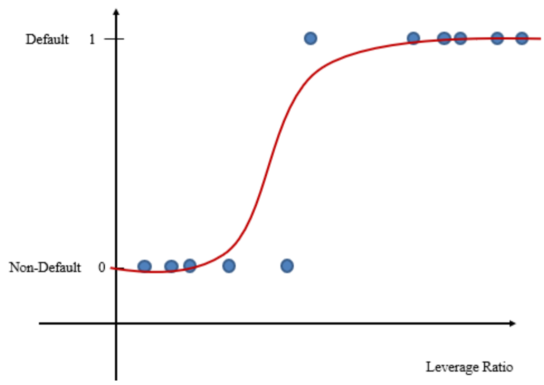
A REVIEW OF THE MAIN MODELS IMPLEMENTED

In this section, a brief review of the main models implemented in literature is provided, so as to be able to better understand in detail the functioning of the techniques. It starts with logistic regression, to which follows an introduction to machine learning (ML) and their general features, to then continue with the description of some of these ML techniques. In this way, it will be easier for the reader to appreciate the subsequent empirical sections.

Logistic Regression (LR)

One of the methods most widely used to perform prediction is for sure logistic regression, which is mainly used when the dependent variable is categorical, meaning that it can take only few options, two or more, for example Default or Non-Default, True or False, Success or Failure etc.. For this reason, LR is implemented when observations have to be classified into two or more groups. For the purpose of the thesis, I will present here the characteristics and features of this method when having a dichotomous variable, namely when it can take only two alternatives, Default or Non-Default. It is important to say that usually value 1 is associated with an alternative (Default) and value 0 to the other. In this way, it is possible to interpret the result of the response variable as the probability of the observation to belong to the default group (or to the one to which we applied value 1), given the independent variables. This is the main difference with the classic linear regression, and with the help of a simple graph, it will become evident.

Graph 1: Linear regression



Graph 2: Logistic regression

On the left, we have a visual example of a linear regression: it can be the case that income has to be predicted on the basis of the level of education. Income in this case is a continuous variable, and therefore can take all values on the y-axis (imagine that 0 is set to be the level of income that ensures basic standard of livings, so in this way it can also take negative values). The blue dots represent the observations collected, while the red line is the best fitting line found with the OLS method (ordinary least squares). On the other hand, on the right, we have an example of Logistic Regression: it is clear that the dots, representing real observations, can be either on the default line, or on the non-default line, no other possibility exists; thus, the variable y, that is represented by the probability of being defaulted, has values that belong to the interval (0,1). In this case it appears clear that the OLS method would not be the best, as a line can go up till positive infinity and down till negative infinity, but there are no values of the outcome variable corresponding to it. The logistic regression best fitting line is therefore represented by an S-shaped curve, which maximum tends to 1, and minimum tends to zero.

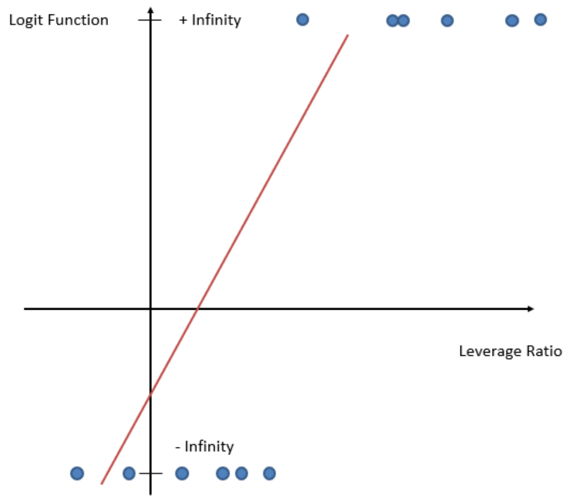
Once established that logistic regression better applies to dichotomous response variable analysis, in order to be able to use this method for different statistical purposes, it is better to transform the y-axis (that was represented by the probability of being classified as defaulted) into the logarithm of the odds scale, which main advantage is the normal distribution. Several clarifications have to be made: first of all, what are the odds. The odds are the ratio of the probability that an event happens, to the probability that that event will not happen (for sake of simplicity: odds are those used in bets, for example if a football team is said to win 3:1, this means that in an hypothetical situation in which 4 matches are played, that team would win 3 times, and lose 1 match). In this context, if the original variable was the probability of being classified as default, and was identified by p, the odds are the following:

$$Odd\ of\ default = \frac{p}{1-p}$$

So, the probability of defaulting divided by the probability of non-defaulting. It is important to specify that odds are not probabilities. Indeed, while a probability varies between 0 and 1, the corresponding odds can vary between 0 and positive infinity; in particular, if the odd is in favour of the event, this will range between 1 and ∞, while if it is against the event, it will vary from 0 to 1. Second important thing, the log(odds) are called “Logit Function”, that is at the basis of the logistic regression technique:

$$Logit\ Function = \log\left(\frac{p}{1-p}\right) = \log(odds)$$

This is the one having normal distribution, and that now represents the new y-axis (hence, the graph will be in log of odds scale). It is worth mentioning, that with p=0.5, the logit function equals 0, and therefore represent the origin of the new graph. Moreover, with p=1, the logit function tends to +∞, while with p=0, it tends to -∞. The new graph is shown below:



Graph 3: log(odds) scale graph for LR

Blue dots at the top, in the direction of +∞, represent observations of default firms, while the others stand for non-default firms. It is important to specify, that for this best fitting line we have the same construction as for the linear regression model:

$$y_{odds} = \beta_0 + \beta_1 X$$

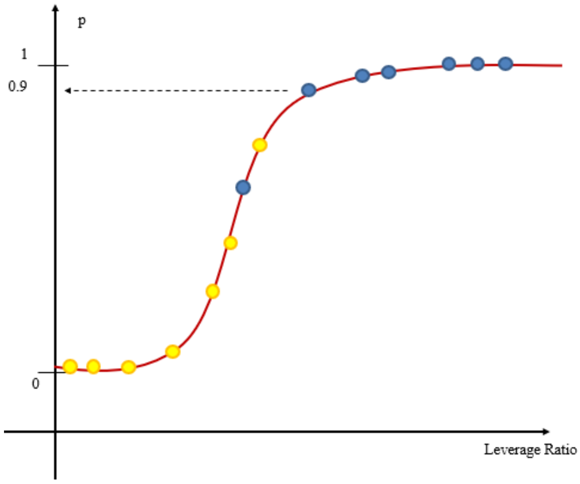
Where β_0 represents the intercept and thus will be negative, reflecting the fact that having a hypothetically null (even if this cannot be the case, imaging it being a very low value) leverage ratio is associated with odds against the possibility of being default. On the other hand, $\beta_1$represents the slope, namely that for one unit increase of the leverage ratio, the log of the odds of default increases by β_1. All this to highlight that the coefficients that are found with logistic regression are in log odds scale.

For what concerns the method with which the best fitted line is found (the line in the last graph, that is directly linked with the S-shaped curve), it is useful to specify that the Ordinary Least Squares (OLS) procedure cannot be applied in this case. The least square procedure consists of finding the line that best fits to the observations, by projecting the dots onto the line, calculating the squared distance ("residual") and summing it up with the squared distance of all the other residuals. The "best fitting" line is the one having the lowest value of sum of squares of residuals. The reason why this process cannot be replicated in logistic regression is that, as can be clearly seen in Graph 3, the residuals are of an infinite measure, in other words, the distance between the observations and the line cannot be computed, since the dots are at positive and negative infinity.

Therefore, in logistic regression Maximum Likelihood is used: to start, also here the observations are projected onto the line (speaking of the line in the log of the odds scale graph); then, the correspondent log of odds is derived and subsequently the related probability, in order to pass into the original plane. The formula to compute the probability from the log of the odds is the following:

$$p = \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$

As a consequence, a graph in the scale of the original one is obtained, with the difference that the observations are all along the S-shaped curve. Now, the likelihood of being classified as default and the likelihood of being classified as non-default is computed. In order to do this, the value of the y-axis is taken for each observation; it has to be noted that in this case the likelihood corresponds to the probability level that emerges projecting the dot on the S curve onto the y-axis. In this figure, default and non-default observations are of different colours, for sake of clarity (blue for the default observations and yellow for the non-default). To have the total likelihood of being classified as defaulted, the likelihood attached to each of the default observations are multiplied between each other. The same is done to compute the likelihood of being identified as non-default, with the difference that here it has to be kept in mind that the probability of being non-default is given by 1 minus the value on the y-axis. Finally, overall likelihood multiplies together the likelihood of the two groups. In a more formal way:



Graph 4: best fitting line with maximum likelihood

$$l(\beta_0, \beta_1) = \prod_{i:y_1=1} p(x_i) \prod_{i':y_i'=0} (1 - p(x_i'))$$

The line that is found to give the maximum likelihood is chosen as to be the best fitting line; therefore, coefficients and further statistical analyses with be based on that line.

For its peculiarities and differences with respect to linear regression, Logistic regression is one of the most widely used statistical methods when the purpose is to classify observations, hence, also in the context of bankruptcy prediction. To be precise, it was implemented by

Ohlson (1980) to overcome some of the problems that arose when using linear regression and the multilinear discriminant analysis; indeed, these require the variance-covariance matrix of all the group to be the same, and the normal distribution of the predictors. As an advantage, Logistic regression does not require any of these, and thus can be applied to a wider variety of samples and contexts. Ohlson, in particular, used LR for bankruptcy prediction of industrial U.S. firms failed between 1970 and 1976, for a total sample of 105 companies, with data from the 10K financial statements were collected up to three years prior to failure. On the other hand, 2058 solvent companies were used for his analysis. He implemented 3 different models, all using nine ratios as predictors: the first predicted bankruptcy within the year, the second within two years and the third within one or two years. In order to assess the goodness of fit of the models, he computed the likelihood ratio (equivalent to the R^2 in the linear regression): it reaches 0.83 for the first model and decreases to 0.72 for the third model. Being similar to the R^2, it means that with the implementation of the first model 83% of the variation of the data is explained through the model; a very good result. In regard of the predictive ability, it suffices to say that all the three alternatives of models reach a percentage rate higher than 90%, with the best being the model predicting bankruptcy within the year (96% correct). For the high predictive results obtained, Logistic Regression will be taken as example in many subsequent works; both as the main implementation method, or as one of the several alternatives with which to compare the results of new techniques. An example of this is the study by Tam and Kiang (1992) that, as already seen, used LR as a measure of comparison to test their artificial intelligent approach. However, in spite of what found by Ohlson, the authors find quite poor results implementing logistic regression on their samples, reporting especially high type 1 errors both one and two years prior to failure. Different is the result found by Chen (2011), who performed a comparative analysis between statistical and artificial intelligent approaches on firms in Taiwan: indeed, even if he concluded the superiority of artificial intelligent techniques, logistic regression turned out to have a high level of overall accuracy. For what concerns the Italian framework with SMEs, Giordini (2014) compared used logistic regression together with support vector machines as a benchmark to test the predictive ability of genetic algorithms. LR was discovered to be the worst performing, with a 66% of entities correctly classified; but it has to be mentioned that none of the methods implemented was able to reach 72% of correct classification; therefore, it can be said that, in general, methods applied to the sample selected were not able to achieve the level of success of other studies. Lastly, in more recent timings, Baboza et al. (2017) also implemented a wide variety of methods, both statistical and artificial intelligent, to predict bankruptcy of American and Canadian firms, focusing on a very large timespan with respect to previous works: 1985-

2013. Using 8 techniques and 11 predictors, including the 5 of the Z-score, they testified with the test sample (2006-2013) the better predictive ability of machine learning approaches, in particular boosting, bagging and random forest, over traditional approaches. The latter obviously included logistic regression, that reached 76% accuracy; even if this result is in line with other researches on the topic, it seems clear that the development of machine learning techniques in some circumstances appears to be crucial for the better realization of failure prediction.

Artificial Intelligence and Machine Learning

There is lots of confusion about these two terms, that are generally used as synonyms while they are not; both refer to techniques and tools used to overcome some problems or to perform a particularly complex task, with the use of machines, or computers. Artificial Intelligence (AI) is a broad concept, and is often defined as the technology used to make machines perform activities simulating human intelligence. Machine Learning, on the other hand, is a subset of AI, and refers to those algorithms that make a machine able to learn from experience without being specifically programmed; therefore, it can be seen as one of the ways to achieve artificial intelligence, if not the most famous and promising for the future. In particular, an algorithm trained through a set of training inputs "learns" the concept, and is then able to replicate the reasoning if applied to other sets of inputs (test set, used to verify the predictive ability of the algorithm). In brief, one of the differences between the two is that Artificial Intelligence also involves the programming of the machine, step by step, to act and behave as a human being, with the actions to be performed that are specifically coded; on the contrary, in ML the algorithm is fed with some inputs, from which it extrapolates the desired output, but from these feeding it learns how to behave in different context (different dataset but with the same output required) without being coded the specific actions to take. It is said to "learn from past experience". They work in different manners, and also have different purposes: AI aims at being able to work at a wider level, and the scope is to make machines able to perform complex activities as humans would do, and therefore, tends to maximize the chance of success; ML aims at accuracy, and is focused on solving one specific problem with the highest rate of precision, it is more limited, since it is designed to complete one specific task. Artificial Intelligence wants the machine to think and to reason as humans, and to be able to perform different actions according to the situation it faces, in order to solve very complex tasks. Examples of AI is Siri on I-Phones or online game playing; while we speak of

Machine Learning for Google Search algorithms or Facebook automatic suggestion of friendships.

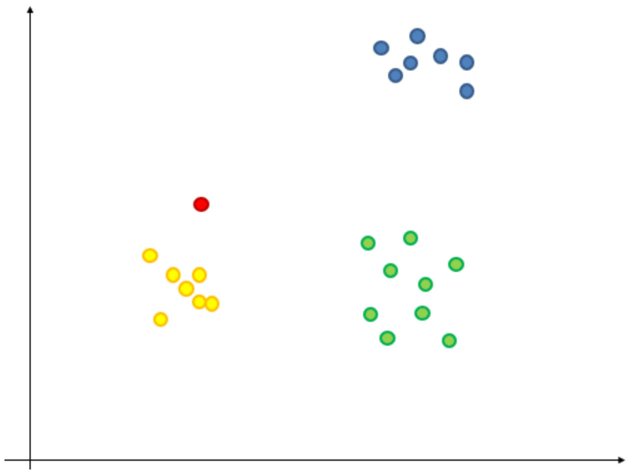
Moreover, Machine Learning can be further divided into three main subsets:

- *Supervised*; in practice, when inputs data are already labelled and the algorithm has the function of deciding which of the given label best applies to the new observation. This is the case of bankruptcy prediction: first, the algorithm is trained with a training set of observation, and these observations are already labelled as default or non-default; then, a test set is used and here the algorithm uses the information he learnt from the previous step to classify these new observation either as defaulted or not.
- *Unsupervised*; more difficult to achieve, it is best explained with an example. Imagine that an algorithm is created in order to find similarities between data, without specifying the number and characteristic of groups wanted. This can be the case for example of images of animals; where the algorithm has to classify these animals according to the similarities it finds. If we feed it with images of cats, dogs, ducks and gooses, the outcome could be composed of three groups: cats, dogs and ducks together with gooses, since they have similarities in their aspects. In addition, it has to be said that the outcome labels are not specified. Just for it to be clear, the correspondent example with supervised learning could be providing an input set of images, telling the algorithm which images represent dogs, which cats, and so on, then testing it with a new set of images; hence, it will classify new images is those pre-specified groups.
- *Reinforcement*; in which a software agent (algorithm) continuously interacts with its environment in order to get the maximum reward or the minimum risk; it works by trial and error with a reward-punishment system. The concept can be better explained if thinking about games: the environment is the grid in which the character moves; the reward can be identified with points gained through some actions (collecting coins or similar); punishment occurs when the game is over (death of character or loss of the battle to make some examples).

Among the different Supervised learning algorithms, we can distinguish among two broad categories: Regression, namely Linear regression, Non-linear regression, Bayesian linear regression; and Classification algorithms, such as Random Forest, Support Vector Machines, Decision Trees and Logistic regression. It must be noted that, in literature, the development of new advanced techniques associated with the term Machine Learning are mainly those of the Classification algorithms, with the exclusion of logistic regression; this because regression techniques are considered to be more traditional and have been in use for a long time, while ML is generally linked with the idea of innovation and a more complex use of algorithms and machines. For what concerns Unsupervised algorithms, the main methods are those of Clustering (with k-means clustering, k Nearest Neighbour and Hierarchal clustering) and Association (with Principle Component Analysis, Independent Component Analysis and Singular Value Decomposition). Speaking of Reinforcement learning, Q-learning and SARSA, that stands for State-Action-Reward-State-Action, are the most common algorithms.

K-Nearest Neighbour

K-Nearest Neighbour is among the most widely used methods for Classification purposes and one of the easiest to perform and understand: the idea is to classify new observations looking at a precise number (k) of nearest observations, and to the group to which they belong. It is worth and more intuitive to explain this technique with a graphical example, because the functioning is visually evident.

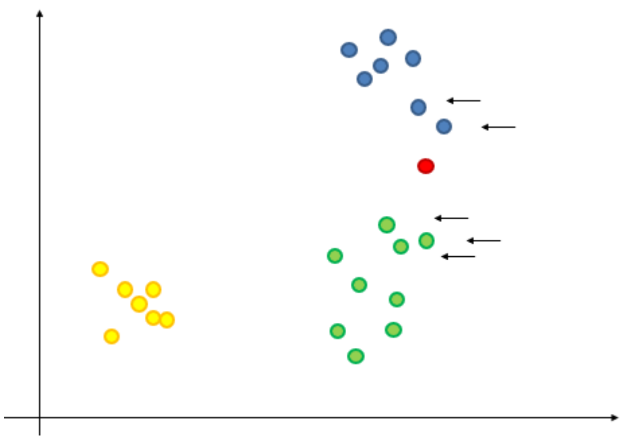


Graph 5: KNN first example

It can be said that this method works in 5 simple steps:

1. Select an appropriate k
2. Feed the algorithm with labelled observations, in the graph above these are the 3 groups of dots (yellow, blue and green)
3. Add a new observation to the graph, the red dot
4. Compute the Euclidean distance of the k nearest neighbours, and see to which group they belong

5. Classify the new instance into the group having the majority of nearest observations. In the graph above, it is evident that the red dot will end up in the yellow group. This is not obvious in other cases, as can be seen in the second example below.

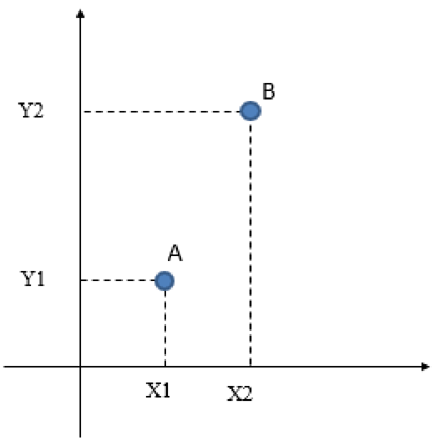


Graph 6: KNN second example

Consider the graph on the right: here the new observation falls between the blue and the green categories, thus its classification is not immediate. This example is to prove the importance of the k selection: if k is set equal to two, the red dot will be classified as a blue dot, since the two nearest dots are both blue. On the other hand, if k is set equal to five, the observation may end up being green, and this is because, after the two blue dots indicated by the arrows, the three subsequent nearest observations are green; therefore, the group having the majority of nearest instances is the green group. This is to say that the choice of the parameter k is fundamental, in particular the choice of an odd number in a binary classification, as the one I will later implement, helps avoiding cases in which the groups have the same number of neighbours. Moreover, it is worth mentioning that a low k can be noisy and subject to the effects of outliers, while on the contrary a high k may prevent classification into group composed of few observations (e.g. if I have 3 labels and one only has 3 observations in it, choosing k equal to 11 automatically excludes classification of new instances into that class).

Another thing that must be specified is the way in which the neighbours are identified: through the classic Euclidean distance between two points, computed as:

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$



Graph 7: Euclidean distance

For what concerns the application of this method in literature, as already said, it is commonly used for its simplicity, and in general an odd number is selected as the best number of observations to look at, in particular, five seems to be the most used nowadays. Yet, for example, Tam and Kiang (1992) in their study, focusing on the development of Neural Network, used 1NN and 3NN as comparison model; even if, it has to be said, the choice of one as the number of neighbours is controversial, indeed, these two models in comparison with the others show the highest percentages of type 1 and type 2 errors. In more recent timings, KNN is also used by Zhang (2017) in order to predict bankruptcy, as compared to Random Forest and Neural Network techniques. He tested different options of k, and found that three, five and seven were the best according to accuracy levels; specifying that k equal to three is relatively the best because the dataset was small. Over the years, this method was also combined with the Fuzzy Set Theory (defined by Shan et al. (2015) as "a research approach that can deal with problems relating to ambiguous, subjective and imprecise judgments, and it can quantify the linguistic facet of available data and preferences for individual or group decision-making"), into what is known as the Fuzzy k Nearest Neighbour (FKNN). An example of this practice is the study by Chen et al. (2011) in which the parameters k and m (representing the fuzzy strength parameter) are specified automatically by the continuous Particle Swarm Optimization (PSO). The result leads to very different k and m parameters for each different fold of data, with k varying from 1 to 100 for the first dataset (Polish) and from 9 to 100 for the Australian dataset. This highlights the importance of choosing the k value that best fits to the sample and to the model specification, including the other parameters.

Random Forest (RF)

This technique takes its name from the fact that it is composed of several decision trees, thus giving the idea of a collection of trees, namely a forest. The underlying purpose under the development of this method is to overcome the shortcomings of singular decision trees that can lead to inaccuracy if not correctly specified. Decision Tree has on its side the characteristic of being simple and well interpretable; therefore, in order to keep this feature while gaining accuracy, comes the creation of Random Forest by Leo Breiman. RF belongs to the so-called *ensemble* category, in other words a combination of multiple classifiers in order to improve the efficiency of single ones.

The process of a Random Forest algorithm appears to be a bit more complex with respect to others, but I will try to explain it in the clearest and simplest manner. At first, a Bootstrap dataset is created from the original one: this has the same size of the original dataset but its observations are randomly selected from the initial sample; hence, by construction, some observations will be repeated and some will be excluded. I want to emphasize the fact that the construction of this Bootstrap dataset influences all the following process (indeed, rerunning the algorithm may lead to slightly different result). The second important step is to create several decision trees from this bootstrap sample, all of them considering different combinations of the variables in the dataset: this means that if the variables that we selected for our prediction model are Sales, Income, Age, Distance and Area of Living, each tree will consider a pre-defined number of variables; therefore, one tree may only consider Income and Distance, another one only has Sales and Age and so on (if the number of variables wanted is 2). In this process lies the improvement with respect to simple Decision Trees, namely on the variety of trees that make the algorithm as a whole very effective. The next step worth mentioning is the final one, those referring to how new observations are classified. When a new instance is introduced into the algorithm, it is passed through all the different trees, and supposing we only have two final groups into which this can be classified (solvent or non-solvent, for example), it is eventually classified into the group that obtained the "majority of votes". This last sentence refers to the fact that if there are 10 different trees, and the new observation comes out to be solvent in 4 trees and non-solvent in 6 cases, it will be classified as non-solvent.

One of the advantages of this method is that, even if it seems a laborious process, it involves less computations and complexity than other techniques, as it is composed of Decision Trees. In addition, combining bootstrap and decision tree processes, made this method one of the most successful in different fields, in fact it is used in banking, but also medicine and biology (for example to predict and classify different type of cells). This is proved by the high accuracy level this technique always reaches in comparison with others, as in the study by Barboza et a. (2017), which can be seen as a collection and comparison of algorithms used in bankruptcy prediction field. As a result, Random Forest achieves the highest accuracy rate together with boosting and bagging, overwhelming other methods such as Support Vector Machines, Logistic Regression and neural Network. Moreover, RF shows the lowest type 2 error together with boosting; it also performs among the bests if considering the Area Under Curve (AUC) of the Receiving Operator Characteristic (ROC) curve. In brief, the ROC is obtained plotting the True Positive Rate over the False Positive Rate; therefore the AUC is the integral of this ROC curve representing the probability that the classifier (RF in this case) will rank a randomly chosen positive instance higher than a randomly chosen negative one. This is a very commonly used indicator when it comes to compare different methods, the higher it is, the better. In the paper by Barboza et al. (2017), RF, boosting, bagging ad Neural Network all achieved an AUC higher than 0.91 which is an extremely good result, considering that the maximum is 1 since it is a probability. Another research that testifies the goodness of Random Forest is that by Zhang (2017), in which two different specification of RF are provided, according to different truncation levels, and both perform extremely good with accuracy rate of 97,5% and 99%. These results leave no doubts on the predictive ability of this technique.

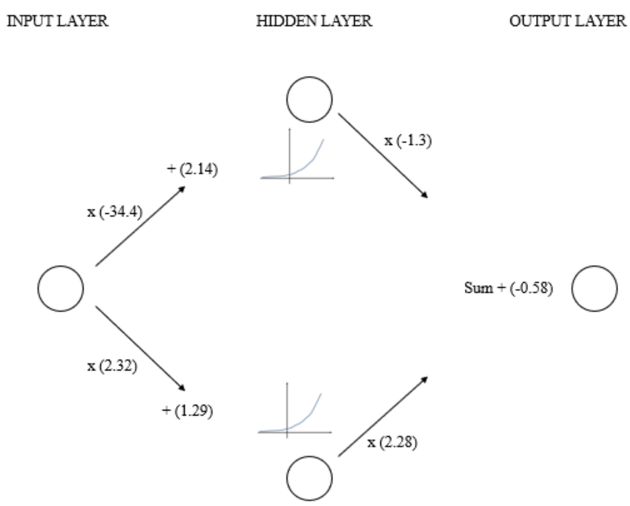
Neural Network (Multilayer Perceptron MLP)

The last category of algorithm that is used as predictor or classifier is that of Neural Network. To be clear with the terminology, a NN with only one hidden layer between input and output is called Perceptron; therefore, when speaking of Multilayer Perceptron, we speak of Neural Networks with several hidden layers. The basic composition of a NN is thus constituted of one input layer, one or more hidden layers and an output layer. They are ordinarily called Neural Networks because they are said to resemble the structure of the brain: in fact, each hidden layer is composed by a number of nodes that are associated with neurons, and the connections between nodes are associated with synapses.

To deepen into the structure of a Neural Network, it can be said that they are composed of four elements:

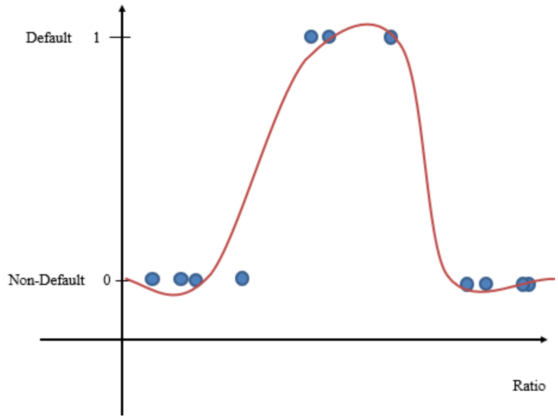
- Input Layer
- Weight and Biases
- Weighted Sum
- Activation Function

To better understand each element of the algorithm, suppose we want to classify solvent and non-solvent firms on the basis of one input only, that is for example a ratio varying from 0 to 1; moreover, to make it simple, we decide to implement the algorithm only with one input layer composed of two nodes. Suppose, in addition, that the algorithm has already been fit to our train set, and



Graph 8: NN with one input and one hidden layer

therefore, that weight and biases have already been estimated through backpropagation; in this way the use of the element will become clear when describing the process of classification of a new observation. We feed the input of the new instance and we see that, to reach the two nodes of the hidden layer, two different ways are possible. The nodes are reached by multiplying the value of the input (our ratio) by a number, that is the weight, and then summing the result with another number, the bias: this gives us the x-axis coordinate for the Activation Function. Since the Activation Function is pre-specified, plugging the x value into it gives the y-axis coordinate. The result is the graph of the activation function that in graph 8 is represented by the small graph next to the two nodes. Before explaining what an activation function is, it is better to conclude the process: next, the activation function is reshaped through other weights (those in the arrows) in order to obtain a squiggle. Repeating this process for the other way, we are left with two reshaped activation functions: for each x coordinate, the sum of the two y coordinates are computed, obtaining a single graph. This last figure is then shifted through a final bias to adapt to the data. The process is laborious, but leads to a final curve that adapts to different data. Just to have an idea, the graph shows the advantage of this method: as we said for logistic regression, with a binary variable as our dependent variable, it is not possible to fit data with a straight line, and Neural Networks are another efficient technique to shape and adapt curves to data. Taking into consideration the example on the right, we may have different level of the input that lead to the two values that the output can take alternatively! It is evident that both a straight line and a sigmoid of logistic regression are not the best to fit data, and here comes the advantage of Neural Networks, that are capable of creating multiple type of curves.



Graph 9: squiggle adapted to data

Coming back to the Activation function, it must be said that several functions can be used into the algorithm, for example:

- *Soft Plus*, $AF = \log(1 + e^x)$, the one represented in graph 8
- *ReLU*, the Rectified Linear Unit, $AF = \max(0, x)$,
- *Sigmoid*, $AF = \frac{e^x}{e^x+1}$
- *Tanh*, the hyperbolic tan function, $AF = \tanh(x)$

According to the activation function chosen, results may vary; it is worth mentioning that the default function for the algorithm implementation is the ReLU.

The use of Neural Networks started around the 90s, for example among the first to develop a NN algorithm for bankruptcy prediction were Tam and Kiang (1992), who implemented two different NN, one with zero hidden layers and one with one intermediate layer compose of 10 nodes. They then compared these models with several others such as KNN and LR in terms of type 1 and type 2 errors one and two years prior to failure. They reported different results, indeed, for the year preceding failure the model with no hidden layers showed a very low level of type 2 error but a high percentage of type 1 error, while on the contrary the other model showed discrete level of both measures. For the two years before failure analysis, the first model showed higher type 2 error but lower type 1, but the second model performed very well in terms of type 1 error, but a bit worse in terms of type 2 error. As already stated in the previous section, Neural Networks were also tested by Barboza et al. (2017), they found RF, boosting and bagging to be the best classifiers, but NN did not perform badly: it achieved an accuracy equal to 73% and an AUC of 90%, the type 1 error was relatively low but the main problem was with the type 2 error, that was found to be 27%. To conclude, Zhang (2017) applied NN to his dataset and discovered that Neural Network performs better when dropout is applied: in particular, with a dropout rate of 0.3, the model reaches 99,5% accuracy. To sum up, even if it seems that this method is surpassed by others, the work by Zhang (2017) is the proof of the high predictive ability of Neural Networks.

CHAPTER 3: DATA AND METHODOLOGY

The empirical side of this thesis aims at implementing different models in order to be able to predict bankruptcy or financial distress focusing four years prior to failure. The timeframe goes from 2009 to 2019, for a total of 11 years. The sample is composed of 628 Italian commercial default firms retrieved from the AIDA- Bureau Van Dijk database, each associated with five non-default companies, for a total of 3482 firms. These were used in different algorithms in order to assess which can be considered the best for bankruptcy prediction, namely Logistic Regression, Random Forest, K-Nearest Neighbour and Neural Network. In the following paragraphs I will discuss in detail how this sample was obtained, how it is composed and the whole process through which the predictors for the analyses were selected, from which it was then able to display descriptive statistics.

Origin of dataset and Propensity Score Matching

All firms included both in the default and non-default group belong to the commercial sector, or better, to the Ateco Code equal to 4. This choice is based on a preliminary analysis, where it was highlighted that this sector included the majority of failed firms in Italy, therefore, it seemed natural to focus on this in order to apply models to firms with similar characteristics. In particular, the most numerous and thus the selected groups are:

- *Ateco Code 43*; “Lavori di Costruzione Specializzati”, Specialised Construction Works,
- *Ateco Code 45*; “Commercio all’Ingrosso e al Dettaglio e Riparazione di Autoveicoli e Motocicli”, Wholesale Trade, Retail Trade and Fixing of Cars and Motorcycles,
- *Ateco Code 46*; “Commercio all’Ingrosso”, Wholesale,
- *Ateco Code 47*; “Commercio al Dettaglio”, Retail.

For what concerns the non-performing sample, this was obtained downloading data from the AIDA- Bureau Van Dijk database, selecting the Ateco codes wanted and eliminating all those firms that had missing data in the accounting statements for more than 5 years. These firms experienced default in the years 2013-2019; for this reason, the models later implemented will try to predict failure four years before failure, basically because it was possible to get data from the database only from 2009, and this means four years before. Firms included in this sample are those that filed for bankruptcy, but also those in a situation of real financial distress; therefore, firms that are in Pre-Bankruptcy Composition and in the other financial distress measures explained in Chapter 1 are included. After this screening, 696 default firms were selected; Propensity Score Matching (PSM) was applied to this sample.

In regard of the non-default group, again accounting data were obtained through the AIDA database, selecting firms that are considered to be Active. Among these, those that are non-defaulted but still in financial distress are included; thus, these were eliminated from the performing sample, since they are part of the non-performing one. Also here, only firms with five years or more of available accounting data were kept; arriving at around 70.000 companies to which PSM was applied.

It was decided to perform Propensity Score Matching in order to have a more balanced sample: for each default firm, five performing firms were selected. Before the execution of PSM, it has to be clarified that this method was implemented after exact matching in terms of sector (identified with the Ateco Code) and geographic area (firms were divided as belonging to North, Centre and South). Practically, this means that a failed firm of sector 46 located in the North of Italy was then matched with a non-failed firm from the same sector and geographic area; this was done since there are good reasons to believe that firms of the same sector and geographical area tend to have similar characteristics. It is a practice widely used in literature, both in the past and in more recent periods: Altman (1968) designed its first model only for manufacturing firms; whereas Chen (2011), for example, took into consideration only firms quoted in the Taiwan Stock Exchange: this is to say that it is common to restrict the sample in order to have entities with similar features. The restriction applied in this work is not believed to be problematic, since the sample is quite large with respect to the average used in literature; hence, it should not bias the implementation of models.

After this preliminary matching, Propensity Score Matching was applied: failed firms were matched with non-failed ones in terms of Sales and Equity over Total Assets; this was done in order to have an association in terms of size and leverage. This procedure works calculating a score, through the use of logistic regression, for each firm, that varies from 0 to 1, considering only the two mentioned variables; then, for each failed firm, 5 of the performing group with the closest score are selected. To specify, the dependent variable is represented by the status of the firm (1 for default and 0 for non-default), whereas the dependent variables are, in this case, Sales and Equity/Total Assets. Due to lack of some data, the process enables 628 of the default firms to be matched with 2854 performing firms (the performing group is not composed of 3140 companies because the process works "with replacement", meaning that non-failed companies can be used more than once if they are the best match for more than one failed company). The full sample is therefore composed of 3482 firms.

A conclusive step was executed in order to check if the PSM worked well: the ideal was to have no significative difference among the two groups (Default and Non-Default) in terms of Sales and Equity/Total Assets. One of the simplest ways to verify it was to run two different linear regressions having the dummy variable "Status" as predictor: the ideal result was to find the coefficient linked to this variable to be not statistically significant under the t-test approach.

	Coef	Std err	t	P>|t|
Const	40435.85	13266.35	3.048	0.002
Status	-21047.68	325079.84	-0.649	0.516

Table 11: Outcome of Linear regression of Sales over Status

As can be seen above, the coefficient related to the variable "Status" (a dummy variable taking value 1 if firm is default, 0 otherwise) is found to be not significant as desired: indeed, the t test has a value of -0.649 which is not significant. On the table below, the regression for Equity over Total Assets is displayed, and again the coefficient related to status is not significant. These variables are not significant because, for a 5% level significance, the t-test value should be higher than 1.96 or lower than -1.96 to be significant.

	Coef	Std err	t	P>|t|
Const	0.0532	0.016	3.234	0.001
Status	0.0450	0.040	1.117	0.264

Table 12: Outcome of Linear regression of Equity/Total Assets over Status

Thus, it can be concluded that the Propensity Score Matching was effective and therefore the sample obtained was kept.

Sample Composition

In this paragraph, some tables reporting the distribution of the firms in the sample are displayed.

Status	Sector 43	Sector 45	Sector 46	Sector 47
Default	137	75	294	122
Non-Default	563	347	1380	564

Table 13: Distribution of failed group by sector

What is evident from the above table is the major contribution of Sector 46, namely Wholesale Trade, both with respect to the Default and Non-default sample: it constitutes the majority of the sample, followed by Sector 47, that represents Retail Trade. This was expected, since these two sectors are of great importance in every economy.

Default	North	Centre	South	Tot
Sector 43	111	19	7	137
Sector 45	48	19	8	75
Sector 46	209	69	16	294
Sector 47	73	32	17	122
Tot	441	139	48	628

Table 14: Distribution of failed group by area

Non-Default	North	Centre	South	Tot
Sector 43	454	75	34	563
Sector 45	214	93	40	347
Sector 46	984	316	80	1380
Sector 47	330	152	82	564
Tot	1982	636	236	2854

Table 15: Distribution of performing group by area

The two tables above show the distribution of firms by geographic area, for the default and non-default groups respectively. The striking evidence is that the majority of failed and non-failed firms are located in the north, with Centre and South together not even reaching half of the sample. This is valid also in each of the sub-groups, represented here by the sectors. This difference in distribution was expected, since in Italy it is known that the majority of existing firms are located in the North; thus, it can be said that the sample derived from PSM, well represents the global framework of the country.

Default	2013	2014	2015	2016	2017	2018	2019	Tot
Sector 43	38	30	17	9	13	15	15	137
Sector 45	21	18	14	5	5	4	8	75
Sector 46	71	63	52	34	22	16	36	294
Sector 47	21	25	12	13	9	13	29	122
Tot	**151**	**136**	**95**	**61**	**49**	**48**	**88**	**628**

Table 14: Distribution of default group per year of failure

With the information displayed above, the trend of bankruptcy in Italy over the years is showed: as expected, the number of failed companies was higher in the first years of the timeframe, because Italy was still suffering from the 2008 financial crisis. Indeed, during the following years the number of firms in financial distress decreases steadily until 2018, when a significative jump can be observed from 48 firms in 2018 to 88 in 2019. The number is expected to be higher in 2020 due to the recent Covid-19 emergency.

Independent Variables Choice

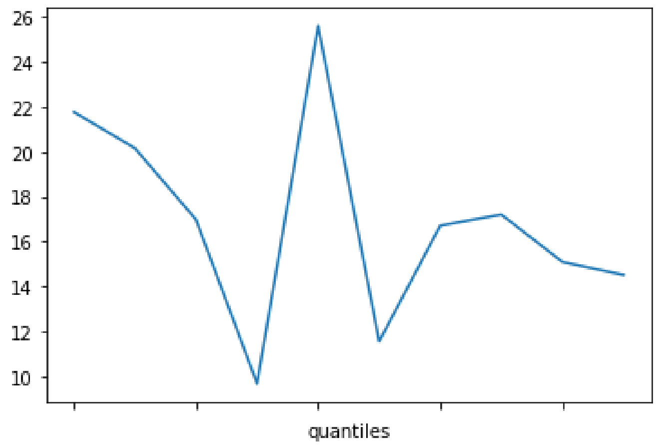
As reported by Winter (2007), the number of factors used by researchers to be included into the analysis as predictors of bankruptcies widely varies according to the different studies, from two factors to twenty-one; moreover the use of more explanatory variables did not assure a higher predictive ability. For this reason, the problem of choosing the appropriate ratios and a well-suited number arose during the analysis. In order to solve this issue, several approaches were taken into consideration and combined, such as skimming by correlation, binning, univariate regressions; each of these will now be briefly described.

Correlation Analysis

At the beginning, a total of 48 possible predictors were selected from two main sources: the first is Winter (2007), that in Appendix B lists all the ratios used in the literature collected according to the frequency, for example the first ratio cited is Net Income / Total Assets because it was found to be the most frequently used (54 times) in the studies taken into consideration by the author. On the other hand, other ratios were collected following the valuation method of enterprises taken from the book "Financial Statement Analysis and Security Valuation" by Penman (2013). In order to avoid multicollinearity problems, these 48 variables were skimmed according to their correlation, in particular, it was decided to keep ratios with a correlation lower than 0.8, a threshold widely used in literature (see e.g. Kennedy 2008). Among those with a correlation higher than this threshold, to decide which indices should be kept and which should be dropped, various criteria were used; for example, if one of the two had several other correlations with other variables over the threshold, then that was the one discarded; if one of the two was more frequently used in literature and seemed to be more significant for predicting bankruptcy due to its composition, that was the one kept. Another factor was considered, namely the amount of data available in the sample for the two candidates. At the end of this analysis, 30 ratios were selected as possible predictors.

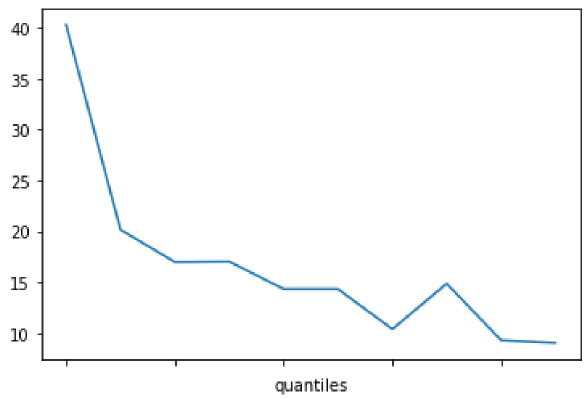
Binning

Binning is a method used to visualize continuous variables in a more useful way, by grouping observations of a variable into different "bins"; hence, it is often said to be an approach to categorize continuous variables. In the context of this work, binning was applied to each of the 30 ratios, in order to have an approximate and preliminary idea of which of the predictors could be more useful for the subsequent analysis. I will now provide two examples to better explain the functioning of this method.



Graph 10.: Binning of Net Sales to Cash from Sales

In this graph binning was applied to one of the 30 ratios that came out from the correlation skimming; on the x-axis are the quantiles, meaning that the variable was divided into ten groups; on the y-axis is the frequency of failed firms in percentage. The aim of these analysis is to understand if a variable could be a good predictor of bankruptcy: since this graph shows in each quantile a number of default firms varying from 10% to 25%, we cannot highlight groups in which default companies are evidently present and groups in which they are not; thus we could expect this ratio not to be very significant.



Graph 11: Binning for Current Assets to Total Assets

A different conclusion can be deduced from this second example: it can be clearly seen that failed firms tend to have a low value of this ratio, since the majority is included in the first half of the ten quantiles; therefore, this variable is expected to have at least some predictive ability. The above analysis was performed for each ratio and the relative figures can be found in Appendix 1. However, the interpretation of the figures was not considered conclusive for the choice of the independent variables for the prediction models, for it was not possible to order the ratios from the better to the worst and to find the best combination of them.

Univariate Logistic Regression

In an effort to validate the outcome of binning, univariate logistic regressions were performed: in other words, a total of 30 logistic regressions were executed, in which the dependent variable 'Status' (a vector of zeros for control firms and ones for failed) was regressed upon the single predictors, taken singularly. For each regression, measures of goodness were obtained, including Accuracy, Precision, Sensitivity and F1-score, computed as follows:

- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$; percentage of correctly predicted firms in total
- $Precision = \frac{TP}{TP+FP}$; percentage of correctly predicted positive observation over total predicted positive
- $Sensitivity(Recall) = \frac{TP}{TP+FN}$; percentage of correctly predicted positive to all the observation in that class
- $F1\ score = \frac{2*Recall*Precision}{Recall+Precision}$; weighted average of Precision and Recall

All these measures involve four measures:

- TP= number of failed firms correctly classified as failed
- TN= number of solvent firms correctly classifies as solvent
- FP= number of solvent firms wrongly classified as failed (Type 2 error)
- FN= number of failed firms wrongly classified as solvent (Type 1 error)

In particular, these measures are derived from the so-called Confusion Matrix, that is an easy and immediate way to understand if a model predicted well or badly, without having a concrete number of goodness of fit (the above measures of performance in fact help to compare results of different models):

$$\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}$$

In our context, the first row gives the sum of the control firms, those correctly predicted on the left; while the second row gives the sum of the failed firms, those correctly predicted on the right. Therefore, it can be naively said that a model with high numbers on the diagonal is good, since the diagonal line represents the number of observations correctly classified by the model. It is worth specifying that these results can be obtained thanks to the fact that the sample is always divided into a training set, used to let the algorithm learn the model, and into a test set, to which the algorithm previously learnt is applied in order to test its predictability onto a new sample. For this reason, results and goodness measures used to compare different models are those referred to the test set only (because extremely high results are expected from the train set, onto which the model is based). The use of these performance measures can be seen in various studies, such as the one by Chen (2011), who compared statistical and evolutionary methods according to Accuracy, Precision, Sensitivity and Specificity. Some authors, for example Giordini (2014) and Madonna and Cestari (2015), only use Accuracy as a comparative element, but in my opinion Accuracy alone is not enough to assess if a model is better than another, because it does not give any insight of the proportion of correctly predicted firms in the groups. To be clearer, from an accuracy level of 70%, one cannot retrieve if both default and not default firms are correctly classified at 70%, or if the high proportion of well predicted firms only comes from one of the two groups, which is fundamental in any predicted model to understand what are the flaws of the algorithm and maybe to modify it in order to achieve more equal results.

It is worth providing a practical example of the link between the confusion matrix and the performance measures, in particular, the following are the confusion matrix and the classification report of the univariate logistic regression related to Net Income to Equity.

$$\begin{bmatrix} 569 & 191 \\ 79 & 92 \end{bmatrix}$$

	precision	recall	f1-score	support
0	0.88	0.75	0.81	760
1	0.33	0.54	0.41	171
accuracy			0.71	931
macro avg	0.60	0.64	0.61	931
weighted avg	0.78	0.71	0.73	931

Table 15: Confusion matrix and Classification report of Net Income to Equity

Analysing the confusion matrix, the number of control firms correctly classified is 569 (True Negative), that of default firms correctly classified is 92 (True Positive), that of control firms wrongly classified as default is 191 (False Positive), while the number of default firm predicted to be non-default is equal to 79 (False Negative). Thus, as stated above, the number of firms correctly classified is displayed on the diagonal. For what concerns the Classification Report on the right, it must be said that "0" identifies the row related to the control firms, whereas "1" identifies the row related to default firms; indeed, on the last column the number of observations for each group is reported, 760 control firms and 171 default, for a total of 931 firms: this is the composition of the test set, by construction equal to 25% of the initial dataset (75% was used as the training set, in order to train the algorithm). The first measure to be taken into consideration is Accuracy, that does not stand on the 0-1 rows since it is a measure of overall performance: it is the sum of all correctly predicted firms over the total test set, therefore $(569 + 92)/931$= 0.7099, approximated to 0.71. This is the most intuitive measure, as it means that 71% of firms were correctly classified. The within group accuracy can be found in the Recall column (often called Sensitivity): 54% of default firms were predicted without error (92/171= 0.538), whereas 75% of the control firms are correct (569/760=0.7486). I would like to specify that this last computation, namely the number of correctly predicted control firms over the total control firms in the test set, is often called Specificity, and is exactly the other side of the coin with respect to the Recall, which in literature is ordinarily referred only to the default group. Next, as stated above, Precision is given by True Positive over the sum of True Positive and False Positive, this gives the percentage over the Positive class. For the default group, this is given by the right column of the confusion matrix, or better, 92/ (92+191) = 0.325. On the other hand, for the control group, we have 569/ (569+79) = 0.878, approximated to 0.88. Likewise, the F1score is the weighted average of precision and recall, therefore 2*(0.325*0.538) / (0.325+0.538) = 0.4052 for the default and 2*(0.748*0.878) / (0.748+0.878) = 0.807 for the control. The rows on the bottom are the average and the weighted average of the two groups. To conclude, it can be said that Net Income to Equity alone has some predictive power, but we will see later that the combination with other indices obviously leads to better results. The analysis was repeated for all the 30 indices, and the next table provides a summary of the measure for all of the predictors, while all the confusion matrices can be found in Appendix 2.

Ratio	**Accuracy**	**Macro Avg Precision**	**Macro Avg Recall**
CFO	0,7250	0,5714	0,5823
EbitToTotAsset	0,5371	0,5608	0,6009
TurnoverPayables	0,3631	0,5446	0,5532
Acid	0,4801	0,5841	0,6272

IntCov	0,2975	0,5761	0,5561
NetSalesToCashFromSales	0,3738	0,5248	0,5326
NetSalestoNAR	0,5510	0,5101	0,5165
EbitdaToEbit	0,7583	0,5590	0,5461
CFOtoEBIT	0,5747	0,5395	0,5650
TAXtoEBIT	0,5800	0,5456	0,5751
OtherRevToTotRev	0,6541	0,5399	0,5570
FixedChargeCash	0,8045	0,5776	0,5177
FixedChargesEbit	0,3201	0,5287	0,5292
RetToTotAsset	0,2793	0,5697	0,5450
CurrentRatio	0,4103	0,5630	0,5844
NetIncomeToTA	0,5779	0,5769	0,6281
TotDebtToTotAsset	0,5424	0,5218	0,5362
CurrentAssToTotAss	0,6069	0,5630	0,6029
NetIncomeToEquity	0,7111	0,6023	0,6440
CurrLiabToTA	0,7658	0,5849	0,5711
QuickAsstoTA	0,5639	0,5731	0,6219
CurrAssToSales	0,8335	0,9153	0,5468
InventoryToSales	0,6327	0,5829	0,6345
ROS	0,2137	0,5495	0,5116
LTDTA	0,4565	0,5789	0,6150
TLToEquity	0,2965	0,5562	0,5442
LnTotAss	0,5456	0,5681	0,6129
OpExpToSales	0,1944	0,5207	0,5020
NWCToEquity	0,6638	0,5519	0,5743
DeltaDef	0,5166	0,5396	0,5657
CashToCL	0,3974	0,5673	0,5856

Table 16: Accuracy, Average Precision and Average Recall for the 30 ratios

These measures (namely accuracy, precision, sensitivity) will be used to evaluate the different models implemented in the next chapter; for now, it is enough to say that these performance measures were plugged into an algorithm whose purpose was the selection of the best variables according to these features; in addition, it is fundamental to say that from the execution of this algorithm the appropriate number of ratios that better distinguish among solvent and non-solvent firms came out to be 25. Several other attempts were tried onto the models, and indeed 25 was confirmed to be the winning combination. These predictors were compared with the binning analysis for a quick check, and it was found that the 5 ratios that were discarded were exactly those not showing a clear trend in the distribution of non-solvent firms in the binning graphs. The next table represents the predictors selected and their composition.

# Factor	Name	Composition
1	CurrentAssetsToSales	Current Assets/ Sales
2	ROE	Net Income/ Equity
3	CurrentLiabToTotAss	Current Liabilities/ Total Assets
4	Acid Ratio	(Current Assets-Inventory)/ Current Liabilities
5	InventoryToSales	Inventory (Tot)/ Sales
6	LTDTA	Long Term Debt/ Total Assets
7	FixedChargesCashCov	(Delta Principal+ Financial Charges+ CFO)/ Current Liabilities
8	EbitdaToEbit	Ebitda/ Ebit
9	CFO	Cash Flow from Operations
10	ROA	Net Income/ Total Assets
11	IntCoverage	Ebitda/ Financial Charges
12	QuickAsstoTA	(Current Assets – Inventory)/ Total Assets
13	NWCToEquity	Net Working Capital/ Equity
14	OtherRevToTotRev	Other Revenues/ Total Production Value
15	CurrentAssToTotAss	Current Assets to Total Assets
16	TAXtoEBIT	Tax/ Ebit
17	RetToTotAsset	Retained Earnings/ Total Assets
18	LnTotAss	Ln (Total Assets)
19	CFOtoEBIT	Cash Flow from Operations/ Total Assets
20	CashToCL	Cash (Tot)/ Current Liabilities
21	CurrentRatio	Current Assets/ Current Liabilities
22	NetSalestoNAR	Sales/ (Tot Customer Receivables–Devaluation of Receivables)
23	EbitToTotAsset	Ebit/ Total Assets
24	TotDebtToTotAsset	Total Debt/ Total Assets
25	DeltaDef	Deferred receivables days – deferred payables days

Table 17: Variables selected as predictors and relative composition

Descriptive Statistics

In this section I will provide some information about the features of some variables of the sampled firms. I would like to specify that I chose to show descriptive statistics referred to four years prior to failure, even if ,on the next chapter, all the prediction models will be tested also on data referred to three years prior to failure and to the mean of the five year preceding the failure. This is because I believe that a model that is able to highlight situations of financial distress and to predict failure years ahead could be more useful for stakeholders, hence, the other two specifications of models (three years ahead and mean of five years ahead) will be used as comparison, in order to show the differences with the main model. Therefore, in the following tables data are collected four years prior to failure; in particular, I found it interesting not only to display the characteristics belonging to the full sample, but also the division among solvent and non-solvent firms.

	TotalAssets	Sales	Profit/Loss	EBIT	TotalDebts	CurrAss/Sales
count	3768	3768	3768	3768	3768	3760
mean	20735	36998	282	699	15081	17,42
std	172745	746967	9292	13794	143766	936,33
min	8	0	-85423	-60315	6	0,00
25%	997	1315	-9	9	816	0,34
50%	3306	4750	9	75	2626	0,53
75%	10680	16318	91	313	8333	0,78
max	8358603	45461542	541171	823165	8120877	57298,28

Table 18: Descriptive statistics full sample

	TotalAssets	Sales	Profit/Loss	EBIT	TotalDebts	CurrAss/Sales
count	628	628	628	628	628	626
mean	18181	19400	-430	-66	14663	0,97
std	49377	48166	3281	3424	39913	7,28
min	79	0	-52723	-60315	44	0,00
25%	2344	2344	-92	-18	1971	0,32
50%	5999	5634	2	83	4732	0,51
75%	14907	16164	26	276	11735	0,74
max	729303	591258	7756	23452	592743	179,75

Table 19: Descriptive statistics failed firms only

	TotalAssets	Sales	Profit/Loss	EBIT	TotalDebts	CurrAss/Sales
count	3140	3140	3140	3140	3140	3135
mean	21246	40517	424	852	15165	20,70
std	187943	817954	10067	15029	156478	1025,42
min	8	0	-85423	-45238	6	0,00
25%	840	1147	-4	10	677	0,34
50%	2805	4547	11	72	2168	0,54
75%	9720	16359	113	321	7583	0,78
max	8358603	45461542	541171	823165	8120877	57298,28

Table 20: Descriptive statistics solvent firms only

The first two variables were chosen in order to be able to classify the entities included in the sample in terms of size; indeed, according to the European regulation, firms are classified as follows:

- Micro firms: sales and total asset level lower than 2 million; number of employees lower than 10
- Small firms: sales and total asset level lower than 10 ml; number of employees lower than 50
- Medium firms: sales lower than 50 ml; total assets lower than 43 ml; number of employees lower than 250

Since many data about the number of employees are missing, I focused on the two other variables. In total, 1485 enterprises meet the asset criteria for the micro firms, in particular 127 in the failed group (representing 20%) and 1358 on the solvent group (43.9%). On the other hand, the number of firms that meet the criteria to be classified as small are 1268, with 286 in the failed group (42.5%) and 982 belonging to the solvent group (31%). Lastly, a total of 968 medium enterprises are present, of which 218 in the distressed group (34.5%) and 750 in the solvent group (24.3%). This means that no Large firms are included in the sample according to the asset criteria, result which is in line with the composition of firm framework in Italy, where 99% of the entities are in fact SMEs. In addition, looking at the above table, no clear distinction can be made about the composition of this variable among the two groups, the means are not distant; while it must be noted that the standard deviation of the performing group is higher, justified looking at maximum and minimum values.

For what concerns the sales criteria, fewer firms classify as micro (1233) and small (1230), and therefore the number of firms classified as medium is higher: 1258. Again, no firms qualified for being Large. However, it has to be noted that Sales was one of the two variables according to which non-solvent companies were associated with solvent ones through Propensity Score Matching, thus, it can be said that due to the composition of the failed group, it was not expected any large firm in the solvent group. Speaking about the statistics, if the total mean of sales is 36,99 ml in the full sample, looking at the two subgroups separately, it can be highlighted that even four years prior to failure is possible to notice that sales of the future default entities are lower than those of the performing group (19,4 ml versus 40,5 ml); meaning that it could be a potential wake-up call for entities, even if not alone (we must not forget that, for example, start-ups tend to have zero or very low level of sales in their first years of existence, but this means nothing in terms of bankruptcy).

Moving to the analysis of profitability, I decided to report both Profit or Loss and Ebit, to show that both variables may be important for the detection of serious distress problems. Looking at the full sample values, the means appear to be positive, but considering the two subsample is far more interesting: indeed, even if both report high negative minimum values, it is evident from the means the difference of features: means for non-solvent groups are negative, while they become positive in the solvent group. Moreover, it must be said that in the non-performing group the first quartile is still significantly negative both for Loss and Ebit, showing that is not just the case of one observation that is negative, but meaning that it is a common trend. The third quartile of Profit/Loss, on the other hand, is evidently higher for solvent firms, and this is in line with what one would expect, and also in line if considering the maximum value of the two groups. For what concerns the standard deviation, it is always found to be higher for the performing group in each of the variables considered, and this is because performing companies are five times more than the failed firms (as consequence of the PSM). We can conclude that the difference among default and non-default entities can be better deduced from the Profit and Loss account, but since Ebit is also widely used for prediction, it was reported to prove that it has distinctive power as well.

Next, the tables show the feature of Total Debts for the full sample and the two subsamples. This variable was reported since having to predict the failure of a firm, one could expect the level of debts to be higher for firms that actually failed; therefore, I found it interesting to show that this seems not a good criteria to distinguish firms, on the contrary at a first glance solvent companies have higher level of debts as indicated by the maximum value. Looking more deeply and considering the third quartile, it must be noted that this value is higher for the failed group than for the performing, therefore, this could mean the maximum showed for the latter is likely to be an outlier. Taking into account also the minimum value, the performing group reports a lower value; but in general, all in all, I would not say that the total level of debt is astonishingly higher for non-solvent entities, so high in comparison with the performing as to justify the subsequent failure.

The last variable I decided to include in this analysis is actually one of the ratios used as predictors in the models, in particular, the one performing better: Current Assets over Sales. This ratio on the univariate analysis showed high levels of accuracy (over 90%) and very high levels of precision as well (over 80%); hence, I wanted to verify if even considering only these simple statistics, the difference among subgroup was clear and noticeable. Indeed, it seems to be the case, taking into account both the mean and third quartile values, that this ratio is higher for the non-performing group (leaving aside the maximum values, since they are clearly due to outliers). A difference was expected after what was previously said about the level of sales, lower for failed firms, but since another ratio including sales was used in the analyses (namely Inventory to Sales), it can be concluded that the high performance of this index in dividing the two groups is due to a combined effect: the lower level of sales and a higher level of current assets for the insolvent companies.

Outliers and Missing Values

The management of outliers is fundamental when working with data, in particular there are some prediction methods or models that are sensitive to outliers; for this reason, it is important to take care of them. One common method, and the one used in this work, is using the z score to identify the outliers, the z score was computed as:

$$z\ score = \left(\frac{value - m}{s}\right)$$

Where m stands for mean of the variable and s stands for standard deviation. This is a classic way to standardize variables. Following a widely used rule of thumb, all those values of the z-score above 3 and below -3 were considered outliers and replaced with missing values; then all missing values were replaced with the mean of the variable. It has to be specified than it is common to replace missing values with the median instead of the mean, since it is not sensitive to outliers, but since all the outliers have been eliminated in the previous step, it seemed more significant to replace the missing values with the mean in order to make them not influential in the analysis.

CHAPTER 4: RESULTS AND MODEL COMPARISON

In this chapter, the results of the models are first reported separately, in order to highlight the trend and differences of the same model three and four years prior to failure with respect to the model implemented using the mean of the preceding five years. In this sense, four different sections will be presented, one for Logistic Regression, one for the K-Nearest Neighbour, one for the Random Forest and one for the Multi-Layer-Perceptron respectively. Then, a final section is dedicated to the comparison of these models among each other.

Logistic Regression (LR)

For logistic regression, and for all the other models implemented in this thesis, some preliminary steps were taken before using the data as inputs. First of all, as already said in Chapter 3, all the outliers were eliminated with the z score procedure, meaning that all values having a z score higher than three in absolute values were substituted with NaN ("not a number" in Python language, meaning their value was deleted and considered as missing values); subsequently, all missing values were replaced with the mean of the variable. Another thing worth mentioning is the division into train and test set: through a specific function available on Python, it was decided to train the model onto 75% of data, chosen randomly, so as to leave 25% for the validation of the algorithm; therefore, in this way train and test set were created for LR and for all the other models. Next, the Standard Scaler was applied to the independent variables, in order to make them comparable: in practice, this means that all the observations were transformed into 0-1 scale, thus assuming values in this interval only. When fitting LR to the training set, the 'lbfgs' solver was selected as the algorithm for the optimization problem (there are few choices among which one can choose); this was chosen for different reasons: it is one of the solvers used in multiclass problems, it is suitable for medium and large datasets (whereas for example 'liblinear' well adapts to small datasets only), and it is the default choice when the function is called on Python. Last thing to be specified is that, by default, Logistic Regression algorithm takes 0.5 as the threshold used to classify an observation: this means that if the algorithm, when predicting the outcome from the test set, predicts a probability of being bankrupt higher than 0.5, that instance is automatically classified as bankrupt. It is evident that this criterion cannot be applied to real cases, or at least that it is not possible that it well adapts to all situations. Indeed, when LR was first applied in this naïve manner, it was clear that 0.5 was not the best choice for my own sample; therefore, with the use of the ROC curve, the best threshold was found and applied for the prediction phase: if an observation was predicted to have a probability higher than this best threshold, it was labelled as failed. After this modification, the model was found to have solid results.

I will now report the results of the three logistic regression, four and three years prior to failure, and for the LR based on mean data. Of course, as expected, the of the LR based on the mean is higher than in the other two model specifications.

$$\begin{bmatrix} 550 & 223 \\ 69 & 100 \end{bmatrix}$$

	precision	recall	f1-score	support
0	0.89	0.71	0.79	773
1	0.31	0.59	0.41	169
accuracy			0.69	942
macro avg	0.60	0.65	0.60	942
weighted avg	0.78	0.69	0.72	942

Table 21: Confusion Matrix and Classification Report of LR 4 year prior to failure

First thing to notice is the size of the test set, looking at the last column: it is composed of 942 observations in total, of which 169 are failed firms (true failed firms) and 773 are control firms. Overall, the model classifies correctly 69% of firms (Accuracy), and in particular, 71% of the control companies versus 59% of the failed ones (Recall column). Therefore, it can be said that it better identifies solvent entities than default ones. Looking at these measures it seems not ideal that only 69% of firms are correctly classified, but it must be taken into account that this model is based on four years prior to failure, hence, this level of accuracy is relatively good. I decided to run this and all the other models on data referred to the four preceding years because identifying failure or signals of high distress four years in advance may be very useful, both for stakeholders in general, but most of all for managers and administrators of the company, in order to try to invert the trend before it is too late. On the other side of the coin there are the error rates, that can be retrieved by difference. The total error rate is the complementary of Accuracy, therefore it is 31%, but to be precise, Type 2 error is equal to 29% and Type 1 error to 41%, which are both quite high. Just to remember, type 1 error occurs when a failed firm is classified as solvent, whereas type 2 error occurs whenever a solvent firm is labelled as non-solvent. Even if, in general, type 1 error is considered more important and thus the one to be controlled and reduced as much as possible, in a context as Italy, mainly composed of Small and Medium Enterprises, as Giordini (2014) reports, type 2 error may not be neglected, and rather it may be the most important. In fact, labelling a solvent firm as default may cause severe problems to the entity, the most important of which is restriction to the access of credit, that may lead to severe consequences for a small firm. To conclude, looking at the average Precision we see it is equal to 60%, but looking in detail, 89% comes from the solvent group, while only 31% from the default group, highlighting the differences among the two.

$$\begin{bmatrix} 614 & 159 \\ 66 & 103 \end{bmatrix}$$

	precision	recall	f1-score	support
0	0.90	0.79	0.85	773
1	0.39	0.61	0.48	169
accuracy			0.76	942
macro avg	0.65	0.70	0.66	942
weighted avg	0.81	0.76	0.78	942

Table 22: Confusion Matrix and Classification Report of LR 3 year prior to failure

It is clear and evident just looking at the confusion matrix, that this model better classifies the control group (614 correct versus 550 of the 4-years model) but still lacks of accuracy when classifying the default group (103 correct versus 100 of the 4-years model). This is reflected in the measures, overall accuracy is equal to 76%, hence, in general, the model works better because it has more entities that are correctly labelled; Recall is equal to 61%, so higher than the previous model as expected; in addition, Specificity (recall for the control group) is higher as well and equal to 79%. On the other hand, the total error of the model is 24%, lower than the precedent; both errors are lower, with type 1 being equal to 39%, and type 2 error being 21%, which is a good signal. In conclusion, average Precision is a bit higher (65%), and the same trend can be highlighted: 90% of the control group and 39% of the default one, this time the values are closer.

$$\begin{bmatrix} 687 & 86 \\ 30 & 139 \end{bmatrix}$$

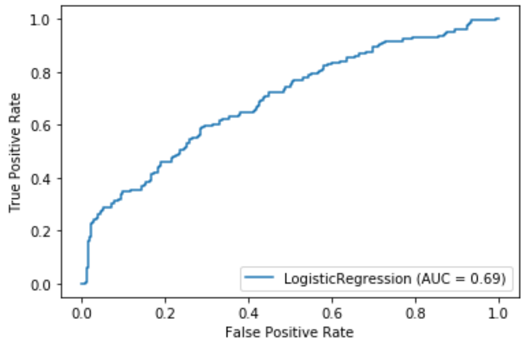
	precision	recall	f1-score	support
0	0.96	0.89	0.92	773
1	0.62	0.82	0.71	169
accuracy			0.88	942
macro avg	0.79	0.86	0.81	942
weighted avg	0.90	0.88	0.88	942

Table 23: Confusion Matrix and Classification Report of LR based on the mean of years preceding failure

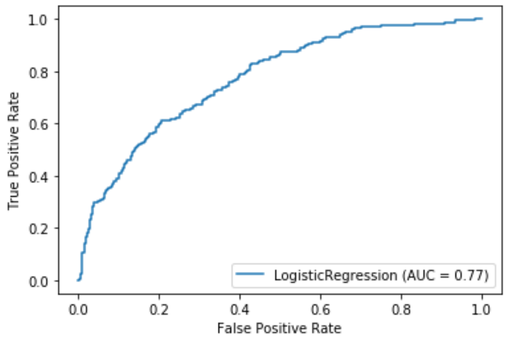
This is the model that performs best from all points of view: 687 correctly labelled solvent firms associated with 139 correctly predicted failed firms lead to an accuracy level of 88%, which is very high. Moreover, recall is equal to 82% and 89% for the non-solvent and solvent groups respectively, both higher than 80%, thus being a very good indicator of the predictive ability of the model with respect to the previous two. As a consequence, total error rate is 12%, with 11% being the type 2 error and 18% being the type 1 error. For what concerns precision, the average is equal to 79% and higher than both the previous models, and we see that this comes from a higher percentage related to the default group, that is in this case 62% opposed to the others that did not reach 40%. These results all confirm that this model works better than the previous two variants, and this is understandable since it includes in the mean also the years right before the bankruptcy declaration, when the situation of a distress firms is surely different and more identifiable.

Another way to compare models is the AUC of the ROC curve, which is a visual method used for comparisons. A Receiving Operating Characteristic curve (ROC curve) is a graph showing the performance of a model; in particular, it plots the True Positive Rate TPR, that is exactly the Recall (or sensitivity) on the y-axis, and the False Positive Rate, that is exactly 1-Specificity (FP/(FP+TN)) on the x-axis. The curve shows these two rates for different thresholds. Therefore, the Area Under Curve (AUC) is the portion of plan below the ROC curve, and is equivalent to the idea of integral: it gives an aggregate measure of performance considering all the possible thresholds. It is usually interpreted as the probability that the model ranks a positive observation higher than a negative observation, our positive observation being the failed firm. Since it is a probability, it varies between 0 and 1, the higher, the better. The main difference between this method and the analysis of the classification report and the reason why I decided to show it, is that this curve is independent of the threshold, whereas the report is the result of the logistic regression implemented with the best threshold. These graphs give a general idea of which model works best, if all the thresholds are considered: it could be the case that once chosen the best threshold, model 1 gives better outcomes that model 2, but if all thresholds are considered, in aggregate, model 2 has a higher AUC than model 1.

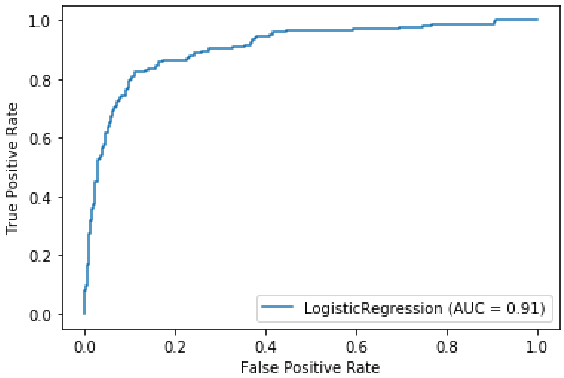
The following two are the ROC curves of the three models, and AUC is indicated at the bottom right. As can be seen, the first two models, those based on values four and three years prior to failure, have AUC values equal to 0.69 and 0.77 respectively.



Graph 12: AUC of LR 4 years



Graph 13: AUC of LR 3 years



Graph 14: AUC of LR mean

On the other hand, the graph on the left, which is the one referred to the mean values, is evidently preferable than the other two, indeed, the AUC is equal to 0.91 which is a very good result. Hence, according to this visual analysis, the model performance is in line with what was predictable, with the models working better when approaching the failure moment.

K-Nearest-Neighbour (KNN)

The k-nearest neighbour was performed after the same preliminary steps discussed for the logistic regression, namely the outlier's identification and the standardization of the predictors. After some trials, it was decided to keep and show the results of the KNN implemented using k=5 and k=7. As explained in Chapter 2, this method classifies new observations on the basis of their k nearest neighbours, identified with the Euclidean distance (according to the inputs given to the algorithm, for example to get the Euclidean distance it must be used 'minikowski' as input for the metric); therefore, the choice of k is important and leads to different results, as I will now prove.

These on the right are the two classification reports of the KNN performed with 4-years ahead data, with k equal five and seven respectively. Looking at the accuracies, the model with 7 neighbours seems better for its 73%, this is due to the fact that it classifies better the group of solvent firms, which is the most numerous among the two. On the contrary, the model with 5 neighbours predicts better the status of observations belonging to the default group, which, being the minority, do not contribute much to the overall accuracy. For what concerns precision, the difference among the two models is not so evident. Among the two, it can be said that the second one is more desirable, since a model with only 58% of overall accuracy may be considered not reliable, even if, as already said, accuracy is not the only measure that should be taken into consideration. However, here the wide difference is determinant.

	precision	recall	f1-score	support
0	0.90	0.55	0.68	773
1	0.26	0.71	0.38	169
accuracy			0.58	942
macro avg	0.58	0.63	0.53	942
weighted avg	0.78	0.58	0.63	942

Table 24: Classification report KNN 4years with k=5

	precision	recall	f1-score	support
0	0.88	0.77	0.82	773
1	0.33	0.52	0.40	169
accuracy			0.73	942
macro avg	0.61	0.65	0.61	942
weighted avg	0.78	0.73	0.75	942

Table 25: Classification report KNN 4years with k=7

Moving to analyse the model with data referred to three years before the bankruptcy declaration, the opposite trend is noticeable: the model with a lower number of neighbours better classifies the control group, and thus reaches higher accuracy, equal to 79%, but only 54% of the non-performing entities are correctly labelled.

	precision	recall	f1-score	support
0	0.89	0.84	0.87	773
1	0.42	0.54	0.48	169
accuracy			0.79	942
macro avg	0.66	0.69	0.67	942
weighted avg	0.81	0.79	0.80	942

Table 26: Classification report KNN 3years with k=5

	precision	recall	f1-score	support
0	0.90	0.77	0.83	773
1	0.36	0.60	0.45	169
accuracy			0.74	942
macro avg	0.63	0.69	0.64	942
weighted avg	0.80	0.74	0.76	942

Table 27: Classification report KNN 3years with k=7

Comparing these two models (3-years based) to the corresponding previous ones (4-years based), comparing those having the same number of neighbours, it is evident the better performance of these last displayed: for example for k=5, accuracy 3-years head is higher, and so are the average recall and average precision. The same reasoning can be applied to the case of k=7, with average precision, recall and accuracy being higher, even if only slightly. The superior performance of the 3-year-base model with respect to the 4-year-based one, suggests that the distress situation is more identifiable when reaching the failure year, and this is in line with common sense; indeed, one would think that a firm would declare bankruptcy when it has no other choice, and therefore it is understandable that the years preceding this moment are the ones in which the situation of an entity that can be derived from financial statements is more noticeable. This is also the reason why, in each model, we expect the model based on the mean to be the best performing.

These are the last two reports, referred to the KNN computed with the mean of the five years preceding failure. Both models show a good performance in a wide sense; here as well, the model with less neighbours better identifies the solvent group; for this reason, it achieves a higher accuracy equal to 82%. Considering average recalls the models are equal, even if in the first one both groups reach a recall of 80%, whereas looking at the average precision, the first one works best. If, as Giordini (2014) believes, type 2 error leads to more severe consequences for entities and is the one that must be taken under control, the best model among these two is the first one, with five neighbours, as it better classifies solvent companies.

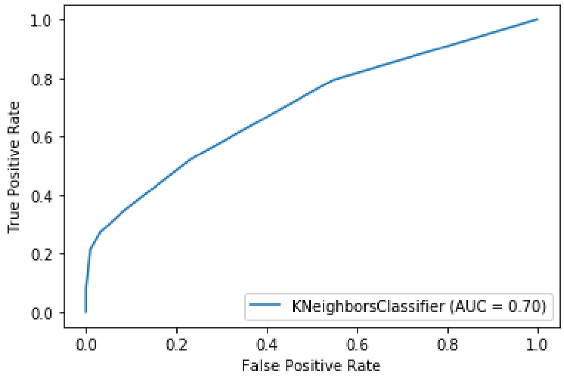
	precision	recall	f1-score	support
0	0.95	0.83	0.88	773
1	0.50	0.80	0.62	169
accuracy			0.82	942
macro avg	0.73	0.82	0.75	942
weighted avg	0.87	0.82	0.84	942

Table 28: Classification report KNN mean k=5

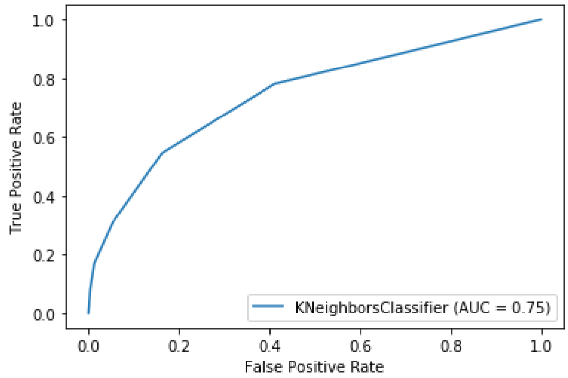
	precision	recall	f1-score	support
0	0.97	0.77	0.86	773
1	0.45	0.88	0.60	169
accuracy			0.79	942
macro avg	0.71	0.82	0.73	942
weighted avg	0.87	0.79	0.81	942

Table 29: Classification report KNN mean k=7

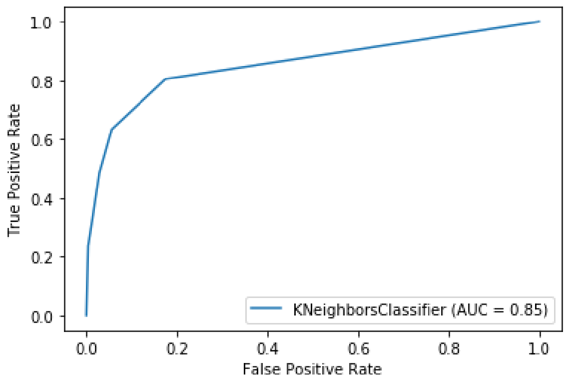
Comparing these with their correspondent three and four year-based models, the mean-based ones are superior over all points of view, thus confirming what stated above. To sum up, different values of k lead to different results in the labelling of instances, and in particular, for the 4-year based model k=7 appears to be the best choice, while for the other two specifications of the model k=5 is the winning choice. Therefore, I will now report here the AUC ROC curve of these three models only, in order to visually compare them.



Graph 17: AUC of KNN 4years



Graph 16: AUC of KNN 3years



Graph 17: AUC of KNN mean

The two graphs above show the ROC curve for the 4-years based and 3-years based model respectively, whereas the one on the right refers to the model executed with the mean values. The striking evidence is the increasing performance of the models when getting closer to the failure moment; indeed, the figures show an AUC equal to 0.70 for the 4-years model, that becomes 0.75 for the 3-years model, and reaches 0.85 for the mean model. As previously said, this is a foreseeable result if considered that the mean model also includes the years right before the default, when the situation was surely severe.

Random Forest (RF)

To recall the functioning of Random Forest, it suffices to say that this algorithm first creates a Bootstrap dataset by randomly selecting observations of the original one, and then creates a series of different trees based on different combination of the predictors. The new instance is thus classified into the group that gets the majority of votes, looking at the outcomes of all the trees created. In regard of the inputs chosen for the algorithms, the number of trees in the forest was set to 100, a very common choice and also the default one; and the criterion, or better the function to measure the quality of the split was chosen to be ‘entropy’. As for the previous models, I will now show the results of all the three model specifications, and it will be immediately clear the superiority of RF with respect to the previous ones, but this aspect will be examined in depth in the last section.

$$\begin{bmatrix} 676 & 97 \\ 59 & 110 \end{bmatrix}$$

	precision	recall	f1-score	support
0	0.92	0.87	0.90	773
1	0.53	0.65	0.59	169
accuracy			0.83	942
macro avg	0.73	0.76	0.74	942
weighted avg	0.85	0.83	0.84	942

Table 30: Confusion Matrix and Classification report of RF 4years

As can be seen above, Random Forest works very well even in the long period, it is able to label correctly 83% of the firms in the sample four years before the failure. Average precision and recall are also quite high if compared with the previous models. For what concerns the error rates, the total error is equal to 17%, while type 2 error is desirably low and equal to 13%, the only downturn for this model is the relatively high type 1 error, equal to 35%, that refers to the future failed firms that are here identified as solvent; we expect this value to be more encouraging in the next two outcomes.

$$\begin{bmatrix} 660 & 113 \\ 50 & 119 \end{bmatrix}$$

	precision	recall	f1-score	support
0	0.93	0.85	0.89	773
1	0.51	0.70	0.59	169
accuracy			0.83	942
macro avg	0.72	0.78	0.74	942
weighted avg	0.85	0.83	0.84	942

Table 31: Confusion Matrix and Classification report of RF 3years

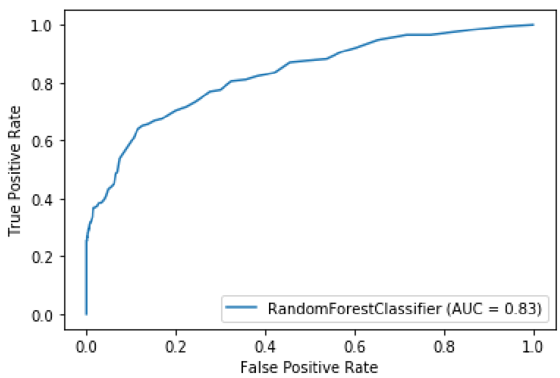
The prediction is confirmed; in fact, the recall is higher in this case and equal to 70%, however, being the specificity for the control group a bit lower, the overall accuracy turns out to be the same for the two model. Focusing on the other measures, average recall is higher for the reason just explained; on the contrary, average precision is slightly lower, and this comes from the default group. Overall, the model works well, with a total error of 17%, that is equal to 15% for the control group (type 2) and to 30% for the failed one; type one error is therefore considerably reduced.

$$\begin{bmatrix}730 & 43\\ 22 & 147\end{bmatrix}$$

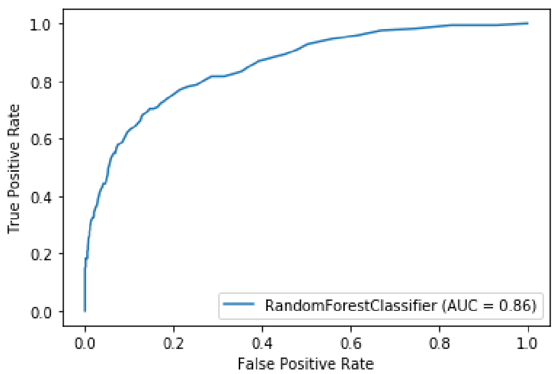
	precision	recall	f1-score	support
0	0.97	0.94	0.96	773
1	0.77	0.87	0.82	169
accuracy			0.93	942
macro avg	0.87	0.91	0.89	942
weighted avg	0.94	0.93	0.93	942

Table 32: Confusion Matrix and Classification report of RF mean

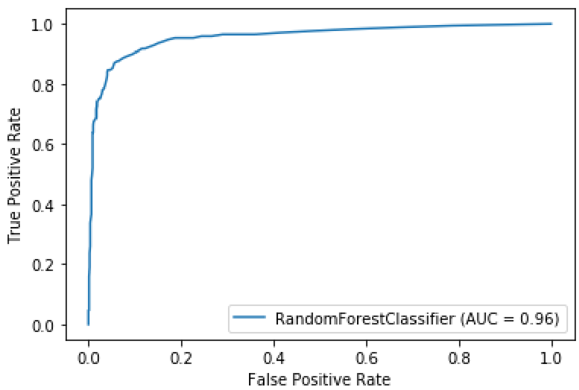
Even if an improvement on the performance was expected in the model concerning the mean values, the result is quite surprising: this model achieves the best result among all the specifications of all models executed, with an accuracy level of 93% which is considered to be excellent, and that leads to a total error rate of only 7%. This model well classifies both solvent and non-solvent entities, with a percentage rate of recall of 94% and 87% respectively; hence, the corresponding average recall is higher than 90%. Moreover, type 1 error turns out to be equal to 13%, whereas type 2 error is 6% only. The performance in terms of precision is higher as well, as it reaches 87%. All these parameters tell us that this model works very well, labelling correctly 93% of the companies in the sample.



Graph 18: AUC of RF 4years



Graph 19: AUC of RF 3years



Graph 20: AUC of RF mean

The evidence of the AUC ROC curve reflects the goodness of the analysis just completed on the report, displaying an AUC equal to 0.83, 0.86 and 0.96 for each models, in order: the 4-year-based, the 3-year-based and the mean-base one. Again, it must be emphasised the result of the mean model, that achieves an

0.96 AUC: this mean that this model works very well and better than the others considering all the possible threshold for the predictors involved in the analysis; recalling that the maximum value of the AUC is 1 since it expresses a probability, it is clear the excellent performance of this last model.

Neural Network- Multi-Layer Perceptron (NN-MLP)

As discussed in Chapter 2, a Neural Network is a system of layers composed of a pre-set number of nodes, all connected to each other. In this, as in the previous models, the variables used as predictors were standardized trhough the Standard Scalar. For the execution of the algortithm, the activation function chosen was the Rectified Linear Unit (ReLU), that returns the maximum among 0 and x: to clarify, the activation function is the one used to create that intermediate graph along the different connections of the algorithm, the rearrangement and sum of which gives the final squiggle that best adapts to data (more details can be found in Chapter 2). The hidden-layers size was set to be equal to 100, as it is the default number for the algorithm execution. Even if not pre-specified, it was noted after running all the three models, that the number of layers that the algorithm found optimal to use was equal to three, meaning that the hidden layers, in addition to the input and output, were three.

$$\begin{bmatrix} 630 & 143 \\ 58 & 111 \end{bmatrix}$$

	precision	recall	f1-score	support
0	0.92	0.82	0.86	773
1	0.44	0.66	0.52	169
accuracy			0.79	942
macro avg	0.68	0.74	0.69	942
weighted avg	0.83	0.79	0.80	942

Table 33: Confusion Matrix and Classification report of MLP 4years

This is the output of the 4-year-based model: it has an accuracy level of 79%, but we see that it classifies better solvent firms (82% correct) rather than the default ones (66% correct): this is the pattern we encountered also in other models, for example logistic regression. Hence, it seems that four years prior to failure, all kinds of models find some difficulties in identifying default firms, and this could mean that the situation for these entities was not so severe at the time to justify their labelling as failed. However, this is still a good result considering it is based on the longest horizon, e.g. four years.

$$\begin{bmatrix} 605 & 168 \\ 42 & 127 \end{bmatrix}$$

	precision	recall	f1-score	support
0	0.94	0.78	0.85	773
1	0.43	0.75	0.55	169
accuracy			0.78	942
macro avg	0.68	0.77	0.70	942
weighted avg	0.84	0.78	0.80	942

Table 34: Confusion Matrix and Classification report of MLP 3years

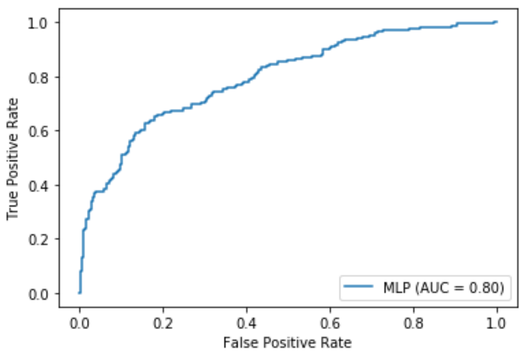
Looking at this table and comparing it with the previous one we see it is quite similar: overall accuracy is a bit higher in Table 33 and average precision is the same in both cases, with the within group precisions also being very similar. What is different are the recall and the specificity: while the 4-year-based model has an evidently better ability in classifying the solvent firms over the failed ones, this second model achieves results quite similar in the two groups; therefore, it can be said that it is more homogeneous. As a consequence, type 1 and 2 error are similar for this latter model, whereas for the former type 2 error is significantly lower than type 1.

$$\begin{bmatrix} 712 & 61 \\ 19 & 150 \end{bmatrix}$$

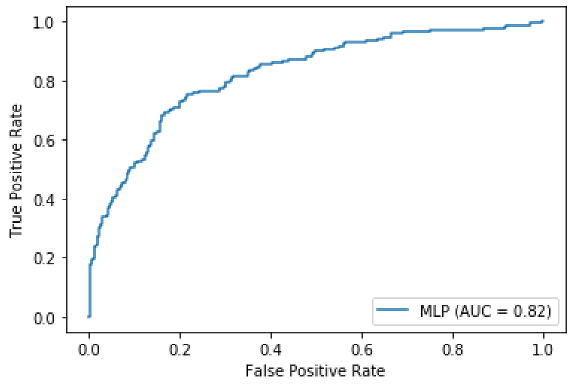
	precision	recall	f1-score	support
0	0.97	0.92	0.95	773
1	0.71	0.89	0.79	169
accuracy			0.92	942
macro avg	0.84	0.90	0.87	942
weighted avg	0.93	0.92	0.92	942

Table 35: Confusion Matrix and Classification report of MLP mean

As for Random Forest, the model based on the mean values achieves exceptional results: an overall accuracy of 92%, that is reflected on the same level of a specificity and on a very high recall (89%). Speaking of error rates, the numbers just mentioned imply a total error and type 2 error of only 8%, and a type 1 error of 11%. With respect to the previous two models, precision is higly improved and equal to 84%, with precision for the control group that reaches 97%. All this confrms the quality of this model, as was for the RF model based on the mean.

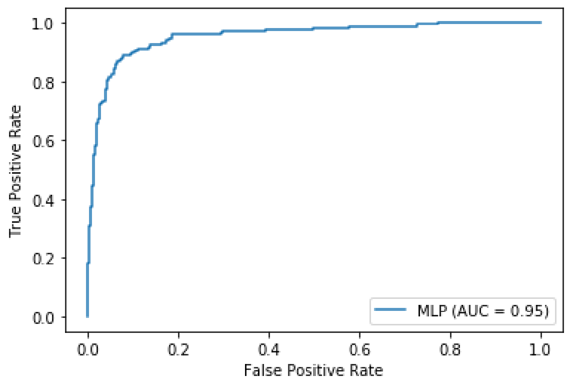


Graph 21: AUC of MLP 4years



Graph 22: AUC of MLP 3years

Observing the figures, the improving trend can be highlighted here as well, with AUC going from 0.80 for the 4years model, moving to 0.82 for the 3years model and reaching 0.95 for the mean-based model. Consequently, it can be concluded that also this final visual analysis confirms the goodness of this Neural Network model.



Graph 23: AUC of MLP mean

Algorithm Comparison

In this section, in order to make a more complete and relevant analysis, I compare the different algorithms, namely Logistic Regression, K-Nearest-Neighbour, Random Forest and Multi-Layer Perceptron, for the different datasets: this means that 3 differrent comparisons will be displayed, one for the four-year-based models, one for the three-year-based models and the last for the models based on the mean of the five years preceding bankruptcy. The comparison is based on a table summarizing the most important performance measures, and on the ROC curve graphs.

Four-year-based models

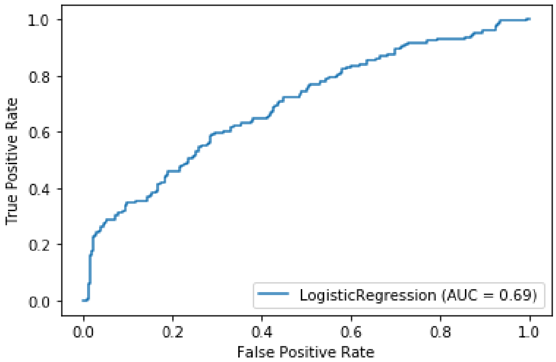
	Accuracy	**Recall**	**Specificity**	**Avg Prec.**	**Type 1**	**Type 2**
LR	0.69	0.59	0.71	0.60	0.41	0.29
7NN	0.73	0.52	0.77	0.61	0.48	0.23
RF	0.83	0.65	0.87	0.73	0.35	0.13
MLP	0.79	0.66	0.82	0.68	0.34	0.18

Table 36: Comparative table for the 4-year based models

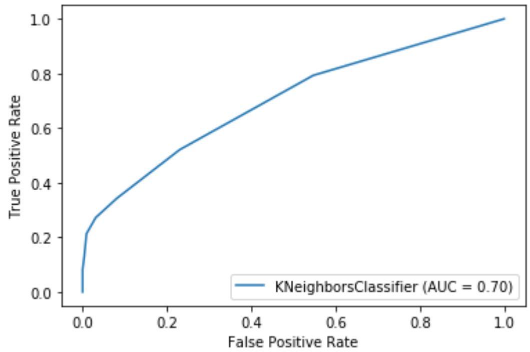
In this and in the following tables are reported the most important performance measures (namely Accuracy, Recall, Specificity, Average precision and the two types of errros), for all the four models executed, namely logistic regression (LR), K-Nearest-Neighbour (KNN), random forest (RF) and the neural network model indicated with MLP that stands for multi-layer perceptron, in order not to be confused with KNN. For what concerns the KNN, data reported only refer to what came out to be the best variation, among the two models with k equal to five and seven; therefore, in this case the 7NN is reported, while in the subsequent two tables the 5NN are reported.

It is quite evident that the algorithm that best applies to the dataset conctructed with data of the fourth year preceding failure is Random Forest: it has the higherst accuracy (83%), the highest precision(73%) and the lowest type 2 error among all, equal to 13% (thus, this means it has the highest specificity, in other words, it is the model that best classifies the solvent entities). It does not have the highest recall of all, that belongs to the MLP, but it gets very close in labelling failed firms and indeed is the second best algorithm according to this parameter (and to the type 1 error). Therefore, if all considered RF is the best algorithm, Multi-Layer Perceptron comes second, with the second best accuracy (79%) and specificity (82%) and the best recall, as already said, equal to 66%. For this reason, it ranks second in terms of type 2 error, and first in terms of type 1 error, still very high and equal to 34%. In regard of the other two models, 7NN outperforms LR in terms of accuracy, average precision and specificity, but LR has a higher recall, meaning that it classifies better the failed group at the expense of the other group.

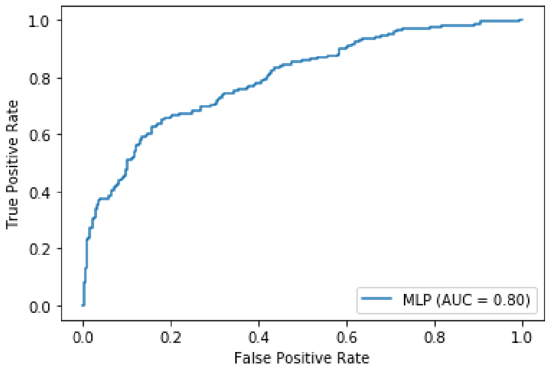
It is worth displaying again the corrispondent AUC of thr ROC curve of the models, to understand if the visual analysis confirms the ranking according to the performance measures or not.



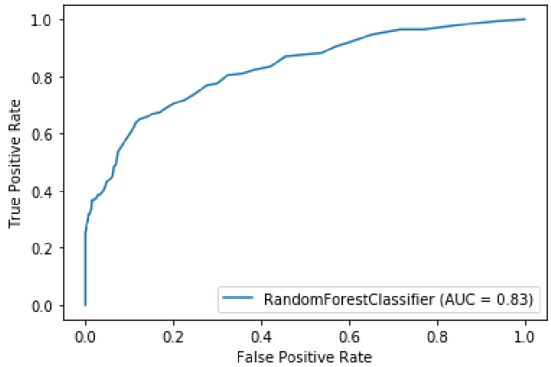
Graph 12: AUC of LR 4 years



Graph 15: AUC of KNN 4years



Graph 18: AUC of RF 4years



Graph 21: AUC of MLP 4years

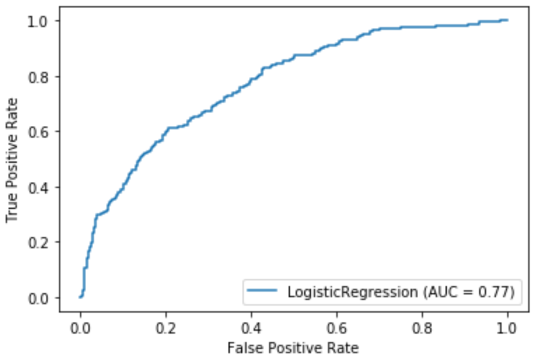
In line with the outcome of the table analysis, for the four-year-based models, the observation of the graphs ranks first the Random Forest, with AUC equal to 0.83, followed by the Neural Network algortihm (MLP), with an AUC of 0.80, to which follow a bit detached the 7-Nearest-Neighbour with AUC equal to 0.70 and Logistic Regression right after with an AUC of 0.69. I would like to underline the separation among models: RF and MPL perform well and are at certain level of performance that can be associated, while KNN and LR are at a completely lower level, and in my opinion, cannot be considered to have an high predictive ablity if compared with the former two, which are more reliable.

Three-year-based models

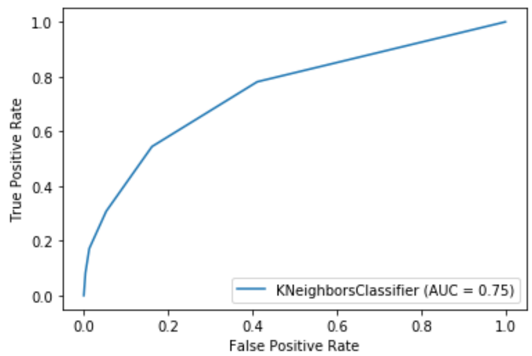
	Accuracy	Recall	Specificity	Avg Prec.	Type 1	Type 2
LR	0.76	0.61	0.79	0.65	0.39	0.21
5NN	0.79	0.54	0.84	0.66	0.46	0.16
RF	0.83	0.70	0.85	0.72	0.30	0.15
MLP	0.78	0.75	0.78	0.68	0.25	0.22

Table 37: Comparative table for the 3-year based models

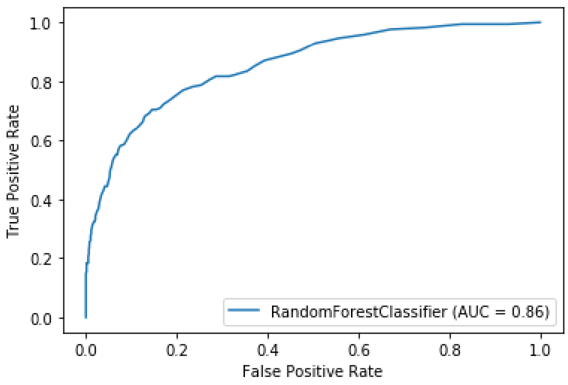
In this case the situation is a bit ambiguous; considering only the accuracies, the rank is the following: random forest, nearest neighbour, multi-layer perceptron and logistic regression. However, in my opinion, there is something clearly wrong in this order, since 5NN only predicts well 54% of the failed entities, and even if it gets very high results in terms of specificity, the performance of the two groups is not homogeneous, and this is not ideal in a predictive model. Moreover, all the other models have a higher recall, meaning that it is placed as the worst in this sense and as a consequence it has the highest type 1 error. Type 2 errors are similar and range between 0.15 of RF and 0.22 of MLP, but all are acceptable. Random forest also has the highest accuracy and highest average precision, hence, for these three reasons, I consider it to be the best. Next, I would say that MLP is ranked second, because it has the lowest type 1 error and an acceptable type 2 error (almost equally well classifies the two groups), the second best precision and a relatively high level of accuracy. So far, the ranking resembles the one obtained for the four-year-based models. For the last two algorithms, a trade off exists and I think the ranking of these depends on what the user wants to achieve when implementing the model, meaning that some would consider one thing to be more important than the other. The K-nearest-neighbour algorithm labels extremely well the control group, achieving the best type 2 errors, but performs poorly, as the worst, when clssifying the non-solvent group, resulting in the higher type 1 error (46%), but thanks to its high specificity, it has a higher overall accuracy with respect to LR. On the other hand there is logistic regression, that is more homogeneous in the groupings, meaning that it has recall and specificity values that are closer to each other, but still, the corrispondent error rates are among the highest in both cases. For the aforementioned logic, I do not believe that one method can be said to outperform the other.



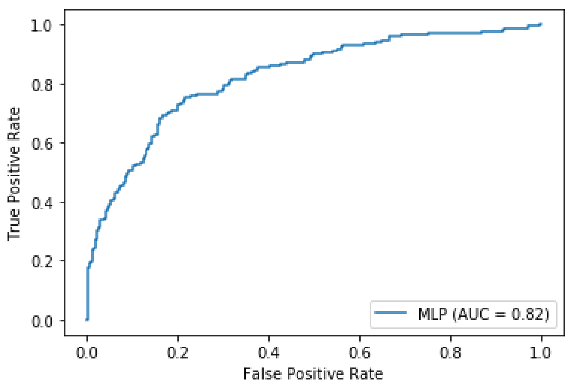
Graph 13: AUC of LR 3years



Graph 16: AUC of KNN 3years



Graph 19: AUC of RF 3years



Graph 22: AUC of MLP 3years

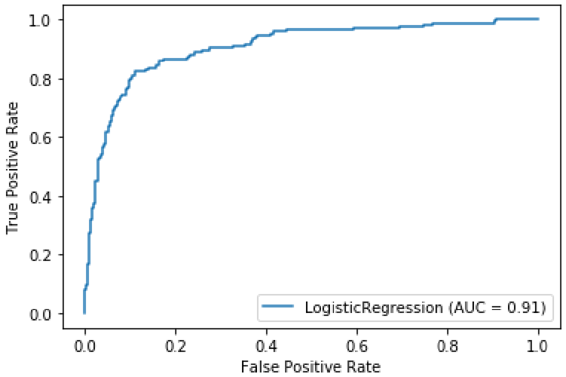
Observing the above figures, that represent sort of performance indicators of the model if all the possible thresholds are considered, random forest turns out to be the best with an AUC level of 0.86, followed by the neural network algorithm with an AUC of 0.82. The analysis of the ROC curves enables a classification among the two remanining models, and in particular, logistic regression is preferred to the nearest neighbour models, since it has an AUC of 0.77, higher than the KNN one, equl to 0.75.

Mean-based models

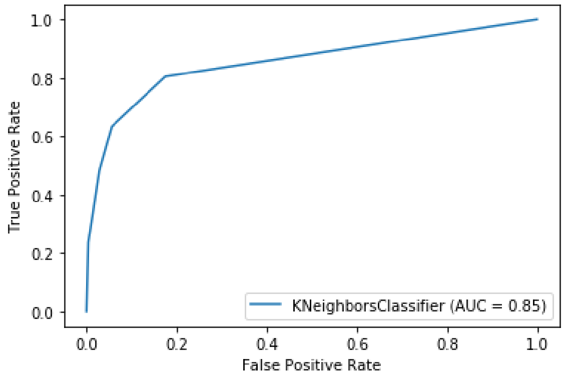
	Accuracy	**Recall**	**Specificity**	**Avg Prec.**	**Type 1**	**Type 2**
LR	0.88	0.82	0.89	0.79	0.18	0.11
5NN	0.82	0.80	0.83	0.73	0.20	0.17
RF	0.93	0.87	0.94	0.87	0.13	0.06
MLP	0.92	0.89	0.92	0.84	0.11	0.08

Table 38: Comparative table for the mean based models

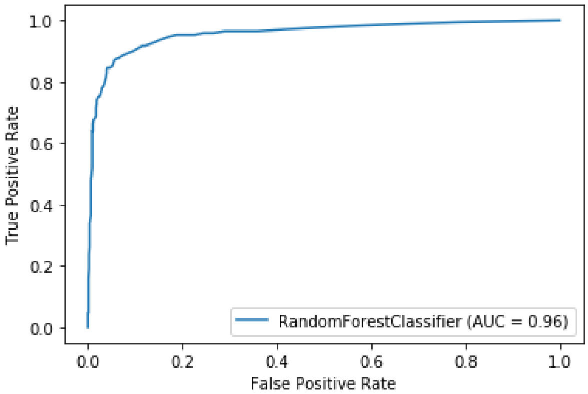
As already said in the preceding sections, the performances of the mean models with respect to the others are surely superior and, in some cases, excellent. With no doubts the two having the best results are the random forest and the neural network, with accuracies and specificities higher than 90% for both. This is also reflected in the type 2 error, that are in each case lower than 10%. The results of recalls are good as well, approaching the 90% level. Among the two, I would say that random forest again is preferable, beacause not only it has the highest accuracy and lowest type 2 error, but also the highest average precision. For what concern logistic regression and KNN, in this case the former outperforms the latter in every aspect, namely in terms of accuracy, error rates and average precision.



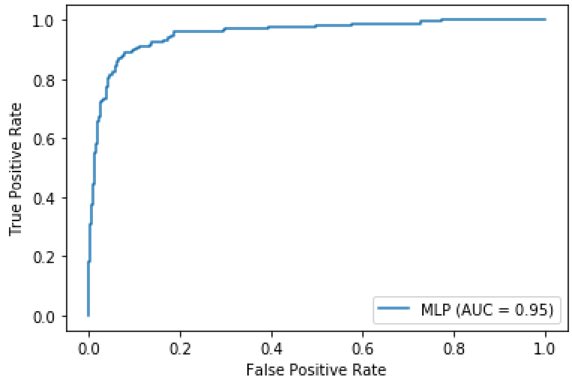
Graph 14: AUC of LR mean



Graph 17: AUC of KNN mean



Graph 20: AUC of RF mean



Graph 23: AUC of MLP mean

The analysis of the AUC of the ROC curves confirms what stated above, indeed the best performing is RF, with an AUC of 0.96, strictly followed by the neural network algorithm with 0.95. Then comes the logisitc regression with a good AUC of 0.91, and at last this time there is the KNN that stays well below with 0.85.

In conclusion, from all these analysis, it emerges that Random Forest is the best algorithm for these databases, in all the three cases analysed. This is not surprising since other researches report high predictive ability of RF (see Barboza et al. 2017). The second best performing is the Multi-Layer Perceptron, or the Neural Network approach, that in some cases shows results very close to those of RF. Regarding the other two models, the most basic ones, it can be said that they always show a performance that is lower with respect to the other two more complex algorithms; in addition, among the two is not possible to declare a winner, since in some cases KNN is more desirable than logistic regression (4-year-based models), while in some cases the opposite occurs (mean-based models). Last, I want to underline that the models working better, namely random forest and neural network, belong to the category of machine learning techniques, and this confirms the superiority of these over the classical or more simple statistical ones.

CONCLUSION

In this work I tried to predict the failure of a given sample of firms belonging to the Italian trading sector, through the use of both statistical and more complex approaches, such as machine learning techniques. On a preliminary phase, I controlled for outliers and I continued with the individuation of the best variables to be used as predictors, with the use of binning and univariate logistic regression with their performance measures used as criteria for ranking: the result was the choice of 25 variables. The next phase was the core of this research, namely the implementation of different models and algorithmic techniques for the prediction; in particular, I decided to use Logistic Regression as the main statistic approach; K-Nearest-Neighbour; Random Forest in order to include a Decision Tree technique; and a Neural Network model. I applied these four models to three different datasets: one considering data of the fourth year prior to failure, one that used data of the third year prior to failure and the last using the mean of the five years preceding bankruptcy.

The findings are explicit: the ability of prediction of all the models increases while increasing the proximity to the failure year; this means that it is possible to notice an improvement from the 4-year-based models to the 3-year-based models, and that the one having the higher predictive power is the mean-based variation (this was foreseeable since it includes data of years right before bankruptcy, when the situation is already compromised). For what concerns the comparison between algorithms, Random Forest undoubtedly reports the best results, in all the three variations; in addition, the Neural Network turns out to be a valid alternative to RF, since it returns good results as well. Among the other two models is not always possible to identify the winner, since the outcome depends on the input chosen; nonetheless, the superiority of RF and NN over the last two is enough to declare and prove the dominance of Machine Learning approaches over statistical or more basic ones. Indeed, the two are expected to give more precise results because of their complex nature with respect to the others. All of these findings are in line with the previous literature.

Even if the analysis performed in this work leads to very promising results, its shortcomings must be highlighted: it must be said that the analysis is based completely upon financial statements data, that by construction are backward oriented and referred to the period just passed. Moreover, I only used analytical information that could be translated into numbers, rearranging them into ratios and performance indicators; a good idea to increase the precision would be to include all those qualitative information that are often discarded in this kind of analysis, mainly because problems of comparability may arise. Luckily, this seems to be the trend in literature, as confirmed by the study of Lohmann and Ohliger (2020), that used soft information included in financial statements in order to have a more complete situation of the company. The ideal would be to collect soft and qualitative information from other sources too, to complete the framework.

This thesis could be continued and extended by testing the models retrieved by my own sample onto different datasets, ideally onto a sample of firms belonging to the trading sector of other countries, to verify the predictive ability in other contexts.

References

AIELLO, A., et al., 2020. *Il piano attestato di risanamento e l'accordo di ristrutturazione dei debiti della crisi di impresa*. Available at: <http://www.dirittobancario.it/approfondimenti/crisi-dimpresa/piano-attestato-di-risanamento-e-accordo-di-ristrutturazione-debiti-codice-crisi-impresa>[Access date: 20.08.2020]

ALTMAN, E.,1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance,* Vol. 23, No. 4 (Sep., 1968), pp. 589-609

ALTMAN, E., 2000. *Predicting Financial Distress of Companies: Revisiting the Z-score and ZETA models.* Available at: <http://pages.stern.nyu.edu/~ealtman/Zscores.pdf>[Access Date: 24.08.2020]

ANON, 2020. Top 10 largest economies in the world. Business Insider India. Available at: <https://www.businessinsider.in/top-10-largest-economies-in-the-world/articleshow/70547252.cms>[Acces date: 20.08.2020]

BAJPAI, P., 2020. *The 5 largest Economies in the World and Their Growth in 2020*. Available at: <https://www.nasdaq.com/articles/the-5-largest-economies-in-the-world-and-their-growth-in-2020-2020-01-22> [Access date: 20.08.2020]

BAGNULO, T., 2019. *Il procedimento unitario di accesso alle procedure nel Codice della Crisi d'Impresa e dell'Insolvenza*. Filodiritto. Available at: <https://www.filodiritto.com/il-procedimento-unitario-di-accesso-alle-procedure-nel-codice-della-crisi-dimpresa-e-dellinsolvenza-ccii>[Access date: 15.08.2020]

BARBOZA, F., KIMURA, H., ALTMAN, E., 2017. Machine Learning Models and Bankruptcy Prediction. *Science Direct-Elsevier,* Vol. 83, October 2017, pp 405-417

BEAVER, W., 1966. Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, Vol. 4, Empirical Research in Accounting: pp. 71-111

BELLOVARY, J., GIACOMINO, D. and AKERS, M., 2007. A Review of Bankruptcy Prediction Studies: 1930 to Present. *Journal of Financial Education*, Vol. 33 (WINTER 2007), pp. 1-42

CASTELLO, F., 2020. *Bancarotta e altri Reati Fallimentari alla luce del Nuovo Codice della Crisi dell'impresa e dell'Insolvenza*. Relazione finale, Università LUISS, Dipartimento di Giurisprudenza

CHEN, M., 2011. Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Science Direct-Elsevier*, Vol. 62, Issue 12, December 2011, Pages 4514-4524

CHEN, H., et al. (2011). A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbour method. *Science Direct-Elsevier,* Vol. 24, Issue 8, December 2011, pp. 1348-1359

FENG, J., 2020. *More than 240,000 Chinese companies declare bankruptcy in the first two months of 2020.* Available at: <https://supchina.com/2020/04/09/more-than-240000-chinese-companies-declare-bankruptcy-in-the-first-two-months-of-2020/>[Access date: 20.08.2020]

FERRARI, M., 2019. *Accordi di Ristrutturazione dei Debiti*. Altalex. Available at: <https://www.altalex.com/documents/altalexpedia/2019/04/02/accordi-di-ristrutturazione-dei-debiti>[Access date: 15.08.2020]

GIORDINI, N., 2014. A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. *Science Direct- Elsevier,* Vol. 41, Issue 14, 5 October 2014, Pages 6433-6445

Global Bankruptcy Report 2019, 2019. Dun & Bradstreet. Available at: <https://www.dnb.com/ie/perspectives/finance-credit-risk/global-bankruptcy-report.html> [Access date: 20.08.2020]

Global Restructuring & Insolvency Guide, 2020. BakerMcKenzie. Available at: <https://www.bakermckenzie.com/en/insight/publications/2020/05/global-restructuring-and-insolvency-guide-2020>[Access date: 20.08.2020]

IRIONDO, R., 2018. *Machine Learning (ML) vs. AI and their Important Differences*. Available on: <https://medium.com/towards-artificial-intelligence/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6 >[Access Date: 04.09.2020]

JAMES, G., WITTEN, D., HASTIE, T., TIBSHIRANI, R., 2017. *An Introduction to Statistical Learning*. 8th Ed. New York: Springer.

JOSHI, R., 2016. Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures. *Exilio Solutions.* Available on: < https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/#:~:text=Recall%20(Sensitivity)%20%2D%20Recall%20is,observations%20in%20actual%20class%20%2D%20yes.&text=F1%20score%20%2D%20F1%20Score%20is,and%20false%20negatives%20into%20account> [Access Date: 24.10.2020]

KENNEDY, P., 2008. *A guide to Econometrics*. 6th Edition. Oxford: Blackwell publishing. Pp 194-196

LIANG, D., et al., 2016. Financial ratios and corporate governance indicators is bankruptcy prediction: A comprehensive study. *Science Direct- Elsevier,* Vol.252, Issue 2, 16 July 2016, pp 561-572

LOHMANN, C. and OHLIGER, T., 2020. Bankruptcy Prediction and the Discriminatory Power of Annual Reports: empirical evidence from financially distressed German companies. *Journal of Business Economics*, Vol. 90, pp. 137-172 (2020)

MADONNA, S., and CESTARI, G., 2015. The Accuracy of Bankruptcy Prediction Models: a Comparative Analysis of Multivariate Discriminant Models in the Italian Context. *European Scientific Journal*. Available on: <https://iris.unica.it/retrieve/handle/11584/190111/363422/Madonna%20%26%20Cestari%20ESJ.pdf>[Access Date: 24.08.2020]

MARANI, S., 2019. *Bancarotta Fraudolenta.* Altalex. Available at: <https://www.altalex.com/documents/altalexpedia/2013/03/01/bancarotta> [Access date: 20.08.2020]

MARANI, S., 2019. *Bancarotta Semplice.* Altalex. Available at: <https://www.altalex.com/documents/altalexpedia/2019/07/16/bancarotta-semplice> [Access date: 20.08.2020]

MARR, B., 2016. *What Is the Difference Between Artificial Intelligence And Machine Learning?* Forbes. Available on: <https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/#6cdc677d2742 >[Access Date: 04.09.2020]

OHLSON, J. A., 1980. Financial Ratios and the Probabilistic prediction of Bankruptcy.

Journal of Accounting Research, Vol. 18, N. 1 (Spring, 1980), pp 109-131

PENMAN, S., 2013. *Financial Statements Analysis and Security Valuation*. 5th Edition. New York: McGraw Hill.

PING SHUNG, K., 2018. Accuracy, Precision, Recall or F1? *Towards Data Science*. Available on: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9 > [Access Date: 24.10.2020]

SERRA, L., 2019. *Concordato Preventivo*. Altalex. Available at: <https://www.altalex.com/documents/altalexpedia/2019/04/16/concordato-preventivo> [Access date 15.08.2020]

SERRA, L., 2019. *Liquidazione Giudiziale*. Altalex. Available at: <https://www.altalex.com/documents/altalexpedia/2019/05/07/liquidazione-giudiziale> [Access date 15.08.2020]

SERRA, L., 2019. *Piani Attestati di Risanamento*. Altalex. Available at: <https://www.altalex.com/documents/altalexpedia/2019/03/26/piano-attestato-di-risanamento> [Access date: 15.08.2020]

SHAN et al., 2015. *In*: BON-GANG, H., 2018. Performance and Improvement of Green Construction Projects. *Science Direct*, 2018, pp. 15-22

SHI, Y., and LI, X.,2019. A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms. *Science Direct- Heliyon*, Vol. 5, Issue 12.

TAM, K. Y. and KIANG, M. Y., 1992. Managerial Applications of Neural Networks: the Case of Bank Failure Prediction. *Management Science,* Vol. 38, No. 7 (Jul., 1992), pp. 926-947

ZHANG, W., 2017. Machine Learning Approaches to Predicting Company Bankruptcy. *Journal of Financial Risk Management,* 2017, Vol. 6, pp. 364-374

Sites

https://corporatefinanceinstitute.com/resources/knowledge

https://pocketsense.com

https://www.altalex.com/documents/codici-altalex/2019/02/15/codice-crisi-impresa

https://www.avvocatoflash.it/blog/diritto-fallimentare/procedure-alternative-fallimento

https://www.brocardi.it/codice-civile/libro-quinto

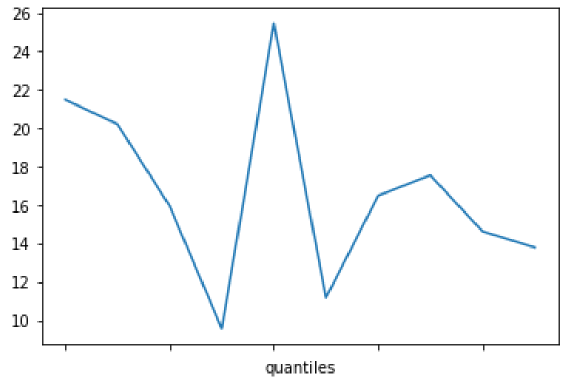
https://www.debt.org

https://www.investopedia.com

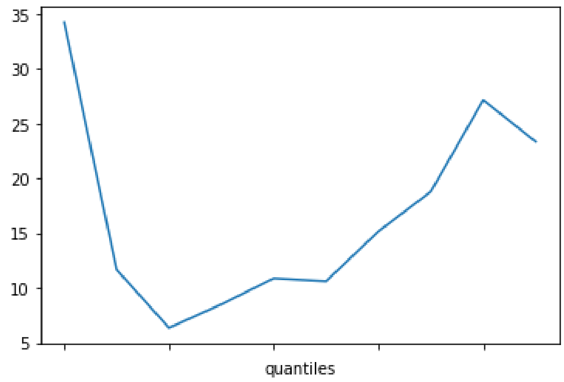
https://www.javatpoint.com

https://www.youtube.com/StatQuest

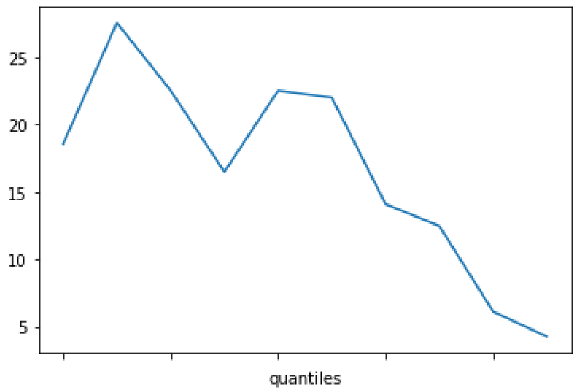
Appendix 1: Binning for the 30 potential predictors



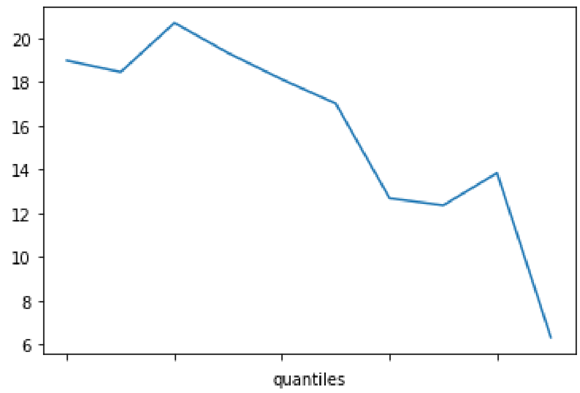
Graph 1: Net Sales to Cash from Sales



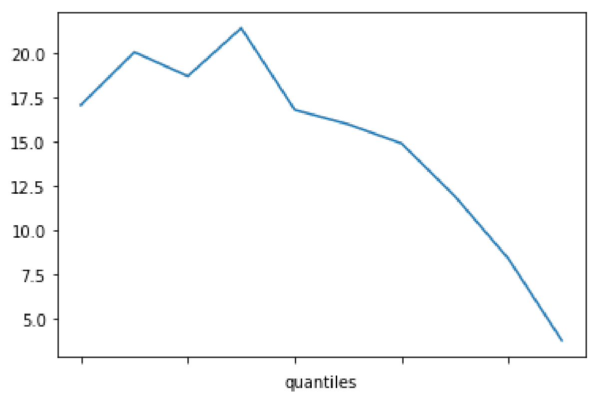
Graph 2: Cash Flow from Operations (CFO)



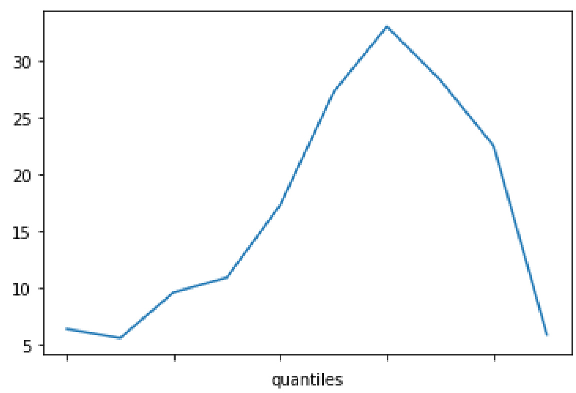
Graph 3: Ebit to Total Assets



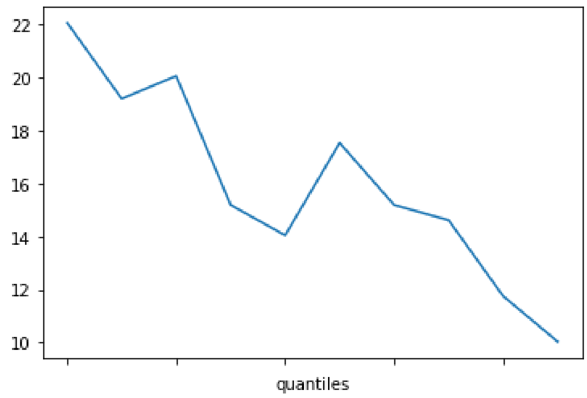
Graph 4: Turnover Payables



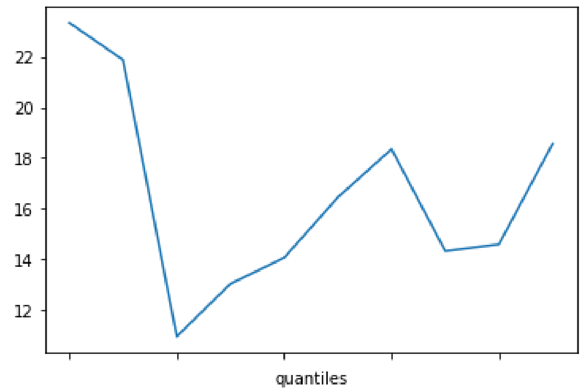
Graph 5: Acid Ratio



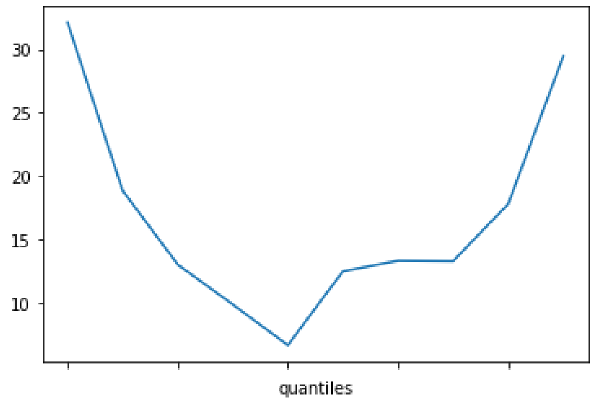
Graph 6: Interest Coverage



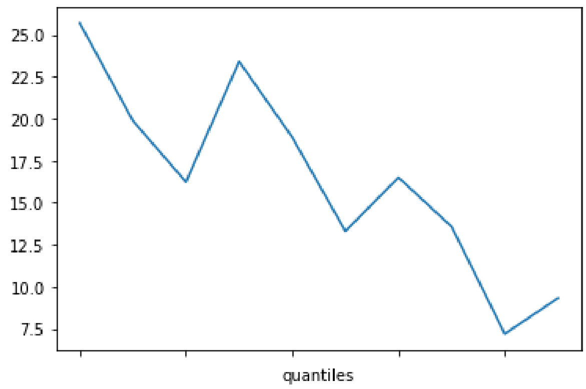
Graph 7: Net Sales to NAR



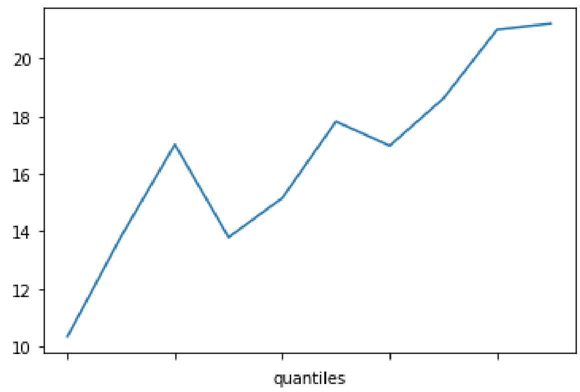
Graph 8: Ebitda to Ebit



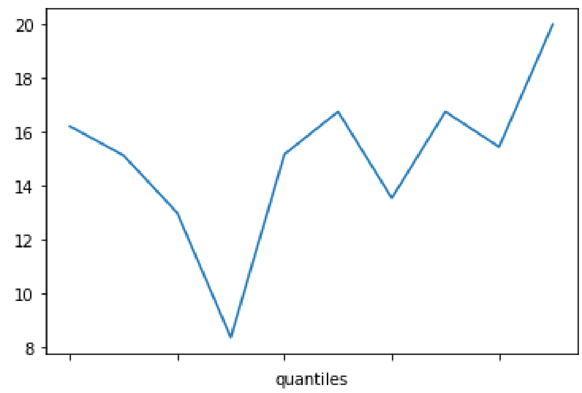
Graph 9: CFO to Ebit



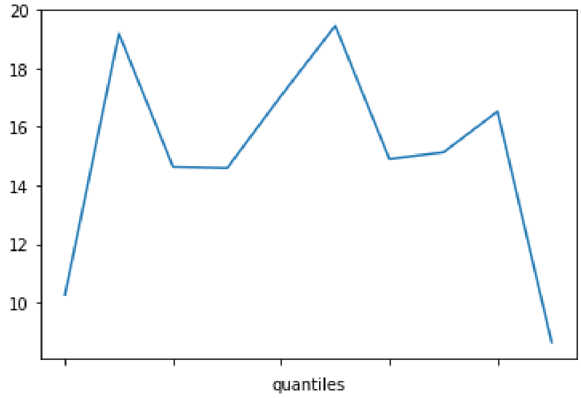
Graph 10: Tax to Ebit



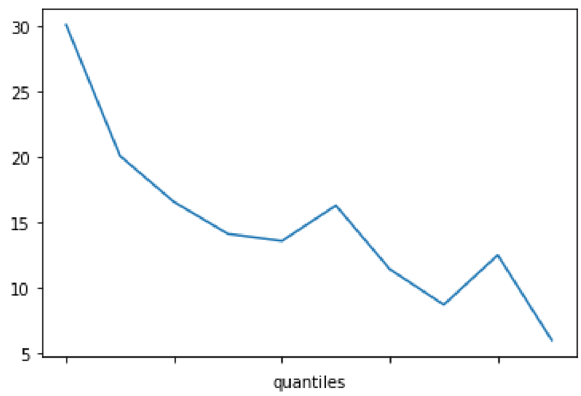
Graph 11: Other Revenues to Total Revenues



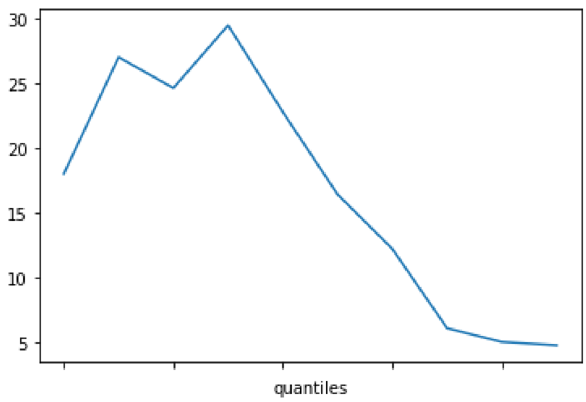
Graph 12: Fixed Charges Cash Coverage



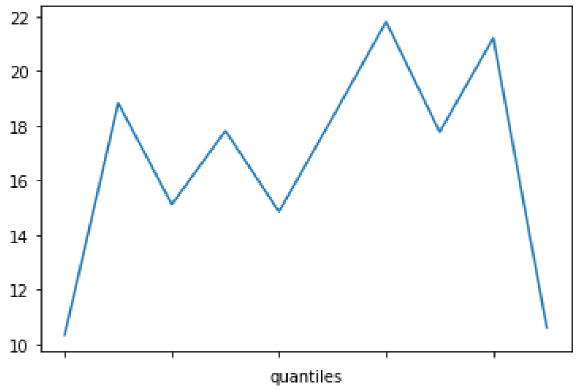
Graph 13: Fixed Charge Ebit Coverage



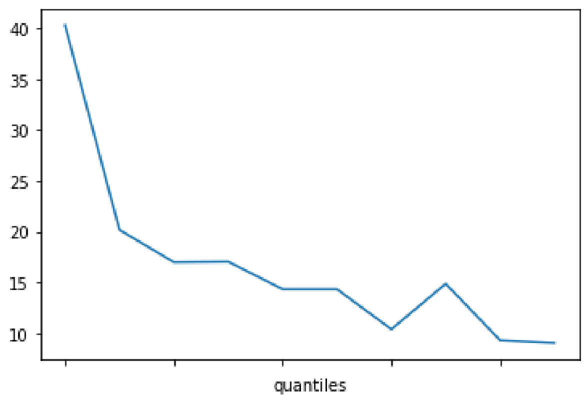
Graph 14: Current Ratio



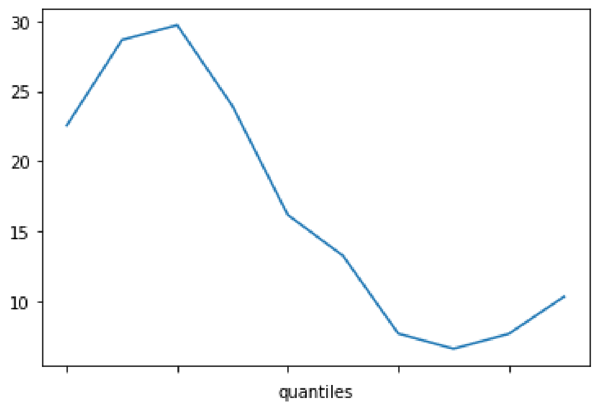
Graph 15: Net Income to Total Assets



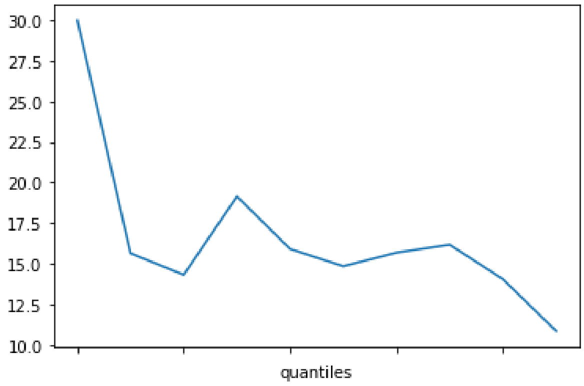
Graph 16: Total Debt to Total Assets



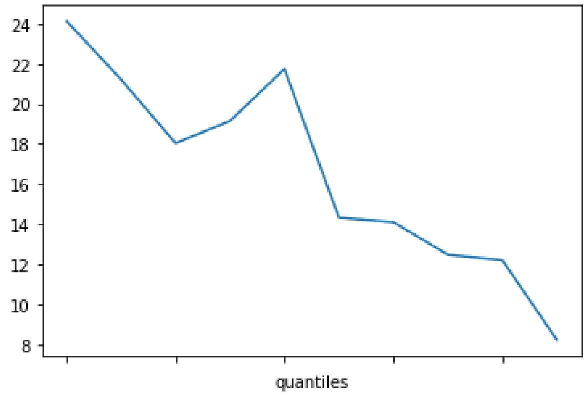
Graph 17: Current Assets to Total Assets



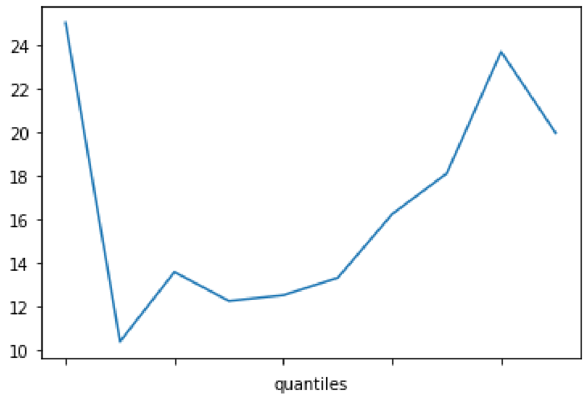
Graph 18: Net Income to Equity



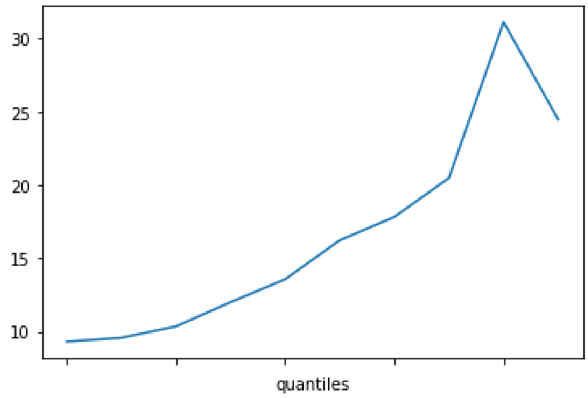
Graph 19: Current Liabilities to Total Assets



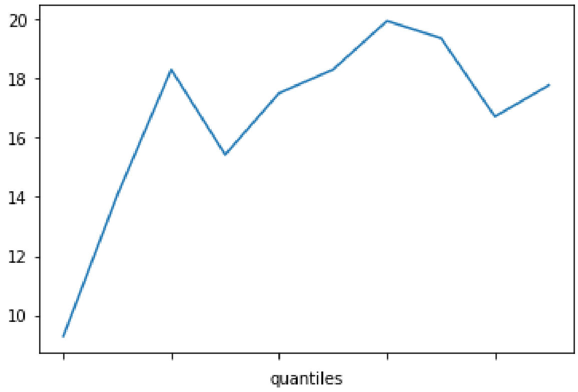
Graph 20: Quick Assets to Total Assets



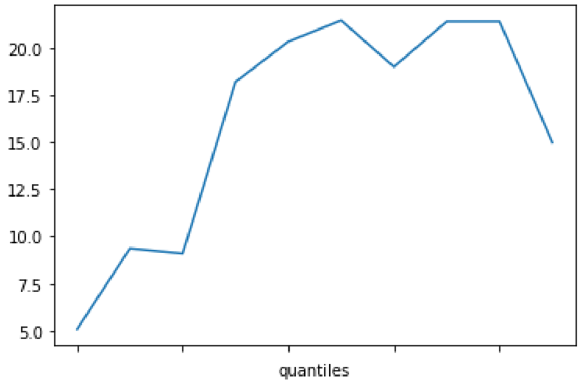
Graph 21: Current Assets to Sales



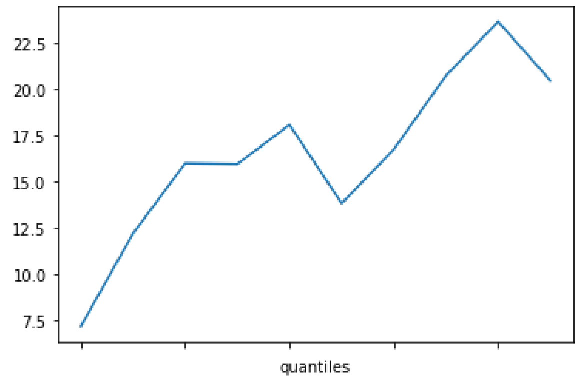
Graph 22: Inventory to Sales



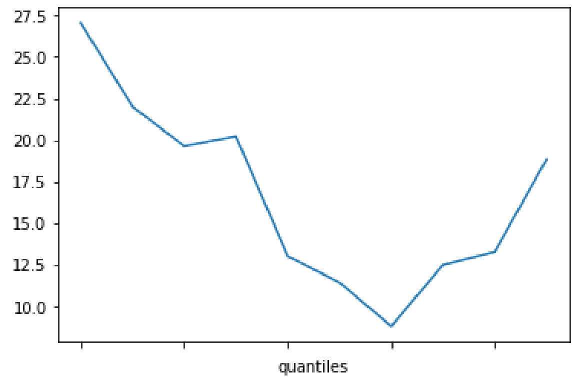
Graph 23: Total Liabilities to Equity



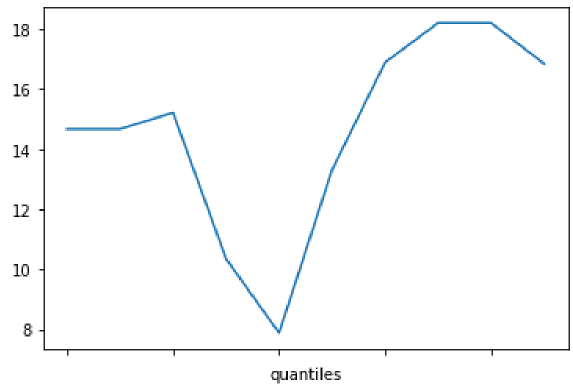
Graph 24: Logarithm of Total Assets



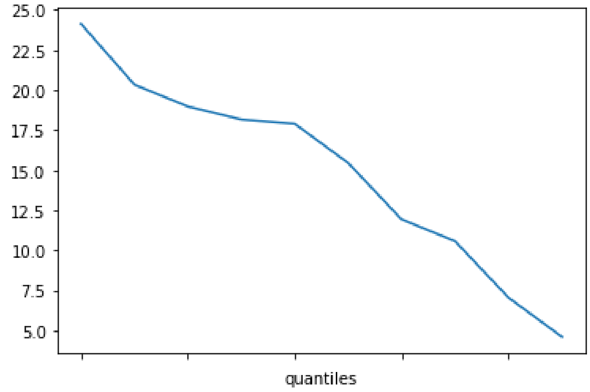
Graph 25: Operating Expenses to Sales



Graph 26: Net Working Capital to Equity



Graph 27: Def collection days - def payment days



Graph 28: Cash to Current Liabilities

Appendix 2: Confusion Matrices for the 30 potential predictors

$$\begin{bmatrix} 488 & 272 \\ 87 & 84 \end{bmatrix} \quad \begin{bmatrix} 380 & 380 \\ 51 & 120 \end{bmatrix} \quad \begin{bmatrix} 309 & 451 \\ 42 & 129 \end{bmatrix} \quad \begin{bmatrix} 353 & 407 \\ 32 & 139 \end{bmatrix} \quad \begin{bmatrix} 112 & 648 \\ 6 & 165 \end{bmatrix}$$

First row: Confusion Matrices of the logistic regression of Status over CFO, Ebit to Tot Assets, Turnover Payables, Acid Ratio and Interest Coverage respectively. As can be seen the output varies a lot, with some variables better predicting the control firms (CFO) and some others well classifying the default firms (Interest Coverage); both at the expense of the other group.

$$\begin{bmatrix} 382 & 378 \\ 74 & 97 \end{bmatrix} \quad \begin{bmatrix} 373 & 387 \\ 82 & 89 \end{bmatrix} \quad \begin{bmatrix} 670 & 90 \\ 135 & 36 \end{bmatrix} \quad \begin{bmatrix} 441 & 319 \\ 77 & 94 \end{bmatrix} \quad \begin{bmatrix} 428 & 332 \\ 70 & 101 \end{bmatrix}$$

Second row: Confusion Matrices of the logistic regression of Status over Net Sales to Cash from Sales, Net Sales to NAR, Ebitda to Ebit, CFO to Ebit and Tax to Ebit respectively. As above, it is evident the fact that single variables can either classify well the firms into one of the two groups (Ebitda to Ebit), or they predict moderately well both group, but with within group results that are worse with respect to the former case.

$$\begin{bmatrix} 540 & 220 \\ 103 & 68 \end{bmatrix} \quad \begin{bmatrix} 738 & 22 \\ 160 & 11 \end{bmatrix} \quad \begin{bmatrix} 151 & 609 \\ 24 & 147 \end{bmatrix} \quad \begin{bmatrix} 95 & 665 \\ 6 & 165 \end{bmatrix} \quad \begin{bmatrix} 360 & 400 \\ 39 & 132 \end{bmatrix}$$

Third row: Confusion Matrices of the logistic regression of Other Revenues to Total Revenues, Fixed Charges Cash Coverage, Fixed Charges Ebit Coverage, Retained Earnings to Tot Assets, Current Ratio respectively. The same pattern can be highlighted as in the previous two rows, with Retained Earnings to Total Assets performing extremely good for the default firms but very badly for the control group, meaning that the majority of solvent companies are seen by the model as non-solvent. The opposite happens with Fixed Charges Cash Coverage, which well identifies the control firms, but that sees default entities as solvent.

$$\begin{bmatrix} 417 & 433 \\ 50 & 121 \end{bmatrix} \quad \begin{bmatrix} 415 & 345 \\ 81 & 90 \end{bmatrix} \quad \begin{bmatrix} 463 & 297 \\ 69 & 102 \end{bmatrix} \quad \begin{bmatrix} 570 & 190 \\ 79 & 92 \end{bmatrix} \quad \begin{bmatrix} 668 & 92 \\ 126 & 45 \end{bmatrix}$$

Fourth row: Confusion Matrices of the logistic regression of Net Income to Tot Assets, Tot Debt to Tot Assets, Current Assets to Tot Assets, Net Income to Equity and Current Liabilities to Tot Assets respectively. The same trend is evident here as well, thus proving the necessity of combining ratios to achieve higher results.

$$\begin{bmatrix} 403 & 357 \\ 49 & 122 \end{bmatrix} \quad \begin{bmatrix} 760 & 0 \\ 155 & 16 \end{bmatrix} \quad \begin{bmatrix} 432 & 328 \\ 53 & 118 \end{bmatrix} \quad \begin{bmatrix} 277 & 483 \\ 23 & 148 \end{bmatrix} \quad \begin{bmatrix} 116 & 644 \\ 11 & 160 \end{bmatrix}$$

Fourth row: Confusion Matrices of the logistic regression of Quick Assets to Tot Assets, Current Assets to Sales, Inventory to Tot Assets, Long-term Debt to Tot Assets and Tot Liabilities to Equity respectively. It must be noted here that Current Assets to Sales identifies all the control companies correctly but lacks accuracy for what concerns the default group. Nevertheless, for the high specificity it achieves 83% accuracy and 92% average precision. Therefore, this ratio is the proof that accuracy alone is not enough, since looking at the average recall (0.54) we understand the poor predictive power within the default group.

$$\begin{bmatrix} 385 & 375 \\ 48 & 123 \end{bmatrix} \quad \begin{bmatrix} 12 & 748 \\ 2 & 169 \end{bmatrix} \quad \begin{bmatrix} 444 & 216 \\ 97 & 74 \end{bmatrix} \quad \begin{bmatrix} 347 & 413 \\ 56 & 115 \end{bmatrix} \quad \begin{bmatrix} 418 & 342 \\ 50 & 121 \end{bmatrix}$$

Fourth row: Confusion Matrices of the logistic regression of Logarithm of Tot Assets, Operating Expenses to Sales, NWC to Equity, Delta Def and Cash to Current Liabilities respectively. In this row it is worth mentioning operating expenses to sales and its high predictive power for the default entities; however, this ratio classifies almost all the control firms as default.