LUISS

Dipartimento di

Impresa e management

Cattedra Asset pricing

New approaches in the valuation of NPLs

Prof. Paolo Porchia

RELATORE

Prof. Marco Pirra

CORRELATORE

Matr.729171

CANDIDATO

Anno Accademico 2021/2022

Sommario

INTRODUCTION

Nowadays, one of the hot topics concerning the business world is that of non-performing loans. This topic became relevant, in particular, following the global economic crisis of 2007-2008 and now, with the outbreak of the COVID-19 pandemic and the decline in many global financial markets, it has become even more relevant.

Banks and consultancy firms therefore found themselves having to deal with a large amount of non-performing loans and to understand which might be the best methods to enable them to recover as many loans as possible.

In recent years, Machine Learning models are widely used in the calculation of non-performing loans, as they are able to estimate values from a large amount of data quickly and accurately. It is important to highlight the fact that very often banks and consulting firms are faced with a higher actual number of non-performing loans than originally assumed. This problem leads to having to reschedule debt collection from scratch, which is very time consuming. In particular, debt collection is proceeding much slower than had been hoped for, which is why efforts must be made to reduce any waste of time in order to allow this market to function as efficiently as possible.

For the aforementioned reasons, the objective of this thesis is to create a programme that, by means of Machine Learning techniques, allows the calculation of the accuracy of any dataset of credits, already divided into performing and non-performing, in order to allow those who use it to understand what percentage of error there might be in the calculation of performing credit portfolios and thus predefine alternative recovery strategies, limiting time losses.

Specifically, Chapter 1 focuses on non-performing loans by tracing their history, how they are broken down, the Italian NPLs market, and finally the impact of NPLs on financial intermediaries.

Chapter 2 focuses on the various methods and parameters for valuing NPLs currently used by banks and consulting firms. In particular, the following methods are described: forecasting models based on judicial recovery; forecasting models based on repayment flows or on company value; statistical forecasting models. Once the models have been described, we proceed with the explanation of how the valuation of the appraised receivable is carried out, thus focusing on the difference between the value derived from the appraisal and, expressed as book value, and the value that buyers are willing to offer on the market. For this reason, the valuation criterion set out in IFRS 9 is then explained in an attempt to bridge this difference.

Finally, we turn to the description of securitisation transactions that follow the valuation phase.

Chapter 3 illustrates the results regarding the accuracy of Machine Leaning methods, focusing on the two methods most widely used today (Logistic Regression and Support Vector Machine), in calculating NPLs and in particular the Phyton programme created to calculate the final values.

The final results derived from the analysis conducted lead one to realise, first of all, that there is no Machine Learning model that can be considered a priori more accurate than another, but one can certainly state that the greater the number of data elements in the analysed dataset, the better the final result will be.

Having used 7 datasets in the analysis, relating to the time period from 2008 to 2013, belonging to three different countries: Germany, Australia and Poland, it can be seen that the Logistic Regression method shows more accurate results in the case of both the 'German dataset' and the 'Australian dataset’ than the Support Vector Machine method. In particular, the German dataset shows an accuracy of 77% with the Logistic Regression method, while 76.6% with the Support Vector Machine method. With regard to the Australian dataset, the Logistic Regression method has an accuracy of 81.6%, while the Support Vector Machine method has an accuracy of 82.6%.

The 5 datasets relating to Poland present more satisfactory results in general than the 2 previously mentioned, this stems from the fact that they have a larger amount of data. The accuracy for all 5 years is always greater than 90%. In the following lines, I will only give two years as an example, for the dataset relating to 2008, the accuracy resulting from the Logistic Regression method is 95.5%, while with the Support Vector

Machine method it is 96.3%. For the 2010 dataset, the accuracy using the Logistic Regression method is 95.6%, and 95.7% using the Support Vector Machine method.

What is most interesting to highlight is the fact that the above results were all obtained after the use of Principal Component Analysis, which drastically reduced the number of features analysed for each dataset, while maintaining a very high accuracy, proving that feature reduction methods are a great way to reduce analysis time and allow banks and consultancy firms to more quickly analyse the immense number of NPLs they have to deal with.

Finally, we would like to emphasise the fact that the programme created in this paper can be used for a large number of other NPLs datasets, as indicated and discussed in more detail in Chapter 3.

CHAPTER 1

1.1 Some introductory remarks

The correct management of Non-Performing Loans (NPLs) has become increasingly important in Italy after the intense and prolonged recession that hit the Italian economy between 2008 and 2014, causing a deterioration in the assets of banks' balance sheets, with a related increase in the number of impaired loans.

Following the financial crisis of 2008, the so-called subprime mortgage crisis, but especially after the subsequent Italian sovereign debt crisis of 2011, there was a significant growth in the total value of impaired loans between 2009 and 2015, which, gross of the devaluations already recorded by the Italian banking system, tripled to 341 billion euros at the end of 2015, compared to 133 billion euros in 2009.[1]

The Bank of Italy defines Non-Performing Loans as "exposures to entities that, due to a deterioration in their economic and financial situation, are unable to meet all or part of their contractual obligations". Today, the world of impaired credit, known as Impaired Loans, is divided into three main categories:[2]

1) Bad Loans, in Italian defined “sofferenze”: they are credits against entities in a state of insolvency or substantially equivalent conditions; this is the most serious case where the debtor is no longer in a position to pay its debt.

2) Unlikely to Pay, in Italian defined “inadempienze probabili”: these are loans for which the likelihood that the borrower will be able to respect the contractual terms in full, or simply pay the loan in its entirety, is considered low by the creditor bank, even though the borrower is not yet insolvent. Banks often prefer to restructure a position rather than default, so that it can be considered "performing" and so that it does not become a real loss for the bank.

3) Past Due, in Italian defined “scaduti”: they are past due loans for which the counterparty has defaulted on payment for at least 90 days. The difference between Bad Loans and Past Due is based on the degree of riskiness of the individual claim. A credit is classified as a Bad Loan when the bank, due to a specific situation (e.g. bankruptcy

---

[1] C. Barbagallo, «“I crediti deteriorati delle banche italiane: problematiche e tendenze recenti”,» Roma, 6 giugno 2017

[2] Ivi

of a company or non-payment for an extremely long period of time), considers the credit to be excessively risky. If, on the other hand, a counterparty fails to pay for at least three months, not because of insolvency but because of special contingent situations, the loan is classified as past due and does not become non-performing.[3] The distressed market is an international industry whose development began in the United States in 1989, with the creation of the first NPL market focused on the sale of impaired loans to third parties. In 1989, the US Congress established a temporary federal agency, the US Resolution Trust Corporation (RTC), to manage the Savings & Loans Associations crisis that emerged in those years, which had seen the failure of several hundred financial institutions.[4] Over a five-year period (1989-1995), the RTC handled the closure of some 747 failed financial institutions by selling performing and non-performing loans amounting to approximately 400 billion USD. Subsequently, this market spread to Japan (late 1980s), Northern Europe (early 1990s), Asia (late 1990s) and finally Europe (2007-present); the main buyers, such as large Hedge Funds and Private Equity, taking into account the various financial crises and cycles, moved from America to operate in the geographical areas that offered the best return opportunities, such as the Asian or European markets. The European distressed market began to develop in 2007, when the subprime mortgage crisis, which began the previous year in the United States, infected European financial institutions, due to the interconnection between the European and American financial systems, causing a significant increase in impaired loans. As a result of the sovereign debt crisis, not only Italy, but also other southern European countries, such as Spain, Greece and, recently, Cyprus and Malta, have been the areas of greatest interest for these large operators.[5]

1.2 Definition of Non Performing Loans

The main pillar of any bank's activity is to collect savings from the public and to provide credit. Statistics have shown that the relation between these two activities has lately

[3] Xolani Dastile, Turgay Celik, Moshe Potsane, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

[4] Q. Francesca, in Rischio di credito e valutazione della Loss Given Default, Roma, Bancaria Editrice, 2007

[5] D. Cucinelli, «"The Impact of Non-performing Loans on Bank Lending Behavior: Evidence from the Italian Banking Sector",» Eurasian Journal of Business and Economics, 2015

been in favour of lending to families and businesses, especially in the last decade. In 2010 there was a very large difference in favour of borrowing. Since that year, the trend has shifted in favour of loans to individuals, families and businesses.[6] In April 2018, the amount of loans granted to customers exceeded the amount of deposits by more than 53 billion, so much so that it has become one of the main sources of income for many financial intermediaries and the immediate solution for many families and businesses. In the last decade we have been experiencing a recession that has affected the entire financial world. This phenomenon has also had a strong impact on the behaviour of the little saver, who has gone from having 'a lot of availability' to having 'a lot of need'. Need for what? To buy a house, buy a car or simply to open a new business. As shown also in the table below, we see that the funding/loans relationship, according to the monthly ABI reports, has always remained in favour of the latter.[7] The strong impact of businesses on the level of indebtedness of Italian banks should also not be underestimated. According to an international comparison made by the Bank of Italy, our country is the " leader". There has always been a clear unbalance between loans and deposits. The economic crisis has not only increased customers' financing needs but has also worsened their economic conditions. Italy's economic situation has been stagnant for almost twenty years and is moving progressively towards a slow growth. Italy remains one of the countries with the highest unemployment rate. All this contributes to emptying the pockets of families and businesses, which are no longer able to pay their debts. As a result, many banks are experiencing a deterioration in payment conditions and interest payments by borrowers.[8] Loans are starting to underperform, hence the term Non-Performing Loans. The term Non-Performing Loans in English is the equivalent of what we call impaired loans, i.e. those loans for which the bank finds it difficult to get its liquidity under control. The reasons for the non-performance of these loans are mostly the difficulty of the debtors in meeting their debt payments and the impact on interest. Often, however, these situations are transitory, since the debtor, thanks to the many collateral guarantees linked to the loan granted, can manage, even partially, to remedy his debt position. There may be a temporary unemployment, or

[6] P. Angelini, "I crediti deteriorati: mercato, regole e rafforzamento del sistema", Roma, 9 ottobre 2018

[7] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

[8] Ibidem

other similar event, in which the debtor is unable to pay his debts for a certain period of time; in other cases, it may be a final and definitive situation.

Particular attention should be paid to those items that have undergone a change:[9]

- Restructured exposures: these are those positions for which the bank decides to give "a second chance" to the debtors by renegotiating the conditions. The data were taken from the ABI monthly reports, in particular the annual reports from 2012-2018 initially established by the contract. Such an operation represents a cost for the bank, but it is the last hope of getting back a small part of the debt. The bank decides to extend the deadline for repayment and also to lower the interest on the debt. In spite of the bank's efforts, it cannot be excluded that the debtor will not be able to repay the debt within the newly agreed terms. For this reason, the previous legislation provided that if the debtor did not succeed in repaying the loan within thirty days of the renegotiation date, the loan would become non-performing.
- Substandard loans: these are loans for which the debtor has temporary, but not permanent, difficulty in making payments within the prescribed period. Substandard loans include all loans with a maturity of three years or less that are at least 150 days past due. This category also includes loans with an original maturity of more than three years that have been past due for 180 days. It is the sole responsibility of the lending bank to report the existence of such a condition to the Bank of Italy's Central Credit Register (CR) so that other lending institutions " will think twice" before providing a loan to the customer.[10]

[9] Xolani Dastile, Turgay Celik, Moshe Potsane, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

[10] Aggiornamento n.7 della Circolare di Banca d'Italia n.272 del 20 Gennaio 2015

1.2.1 Past due exposures

These are credits that must be clearly distinguished from the other two sub-categories, as they are credits that are 90 days past due and have exceeded a certain amount.[11] This is a condition of a continuing nature. Here the non-performance by the debtor is not only objective but also tends to persist over time. It is important to point out that of all the other categories, this is the "mildest" and does not lead to drastic changes in the bank-debtor relationship.

Past due and/or in overdue positions are defined as such if they do not fall within the cash exposures already defined as probable defaults or non-performing and they result in default, exceeding certain exposure thresholds (Relevance Thresholds), consecutively for more than 90 days. There are two types of thresholds: Absolute Threshold and Relative Threshold. The first compares past due exposure with a predetermined amount that varies according to the type of customer (a distinction is made between retail and non-retail). The second threshold is determined by comparing the borrower's overdrawn exposure with the total exposure to the bank for that counterparty.[12]

1.2.2 Sufferings

Here we find in front of the real NPLs, since we are dealing with those situations that are certain, and not probable, in which the subject is in a real state of insolvency or in situations similar to that state.

This is a non-transitory situation that obliges the bank to take "drastic" measures against the debtor, as the latter is unable to fulfil its contractual obligations.

[11] https://www.bancaditalia.it/media/views/2017/npl/index.html

[12] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

In accordance with Circular 272, this category includes cash and off-balance sheet exposures to a person who is insolvent (even if not judicially determined) or in substantially similar situations, regardless of any loss forecasts made by the bank.[13]

Therefore, the existence of any collateral (real or personal) protecting the exposures is not considered.

Exposures that are impaired due to country risk profiles are excluded. This also includes:[14]

• exposures to local authorities (municipalities and provinces) in a state of financial distress for the part subject to the relevant liquidation procedure;

• loans purchased from third parties whose principal debtors are non-performing borrowers, regardless of the accounting allocation portfolio.

A customer must therefore be classified as non-performing:

• in any case, if one of the following events has occurred:

- either declaration of bankruptcy or compulsory administrative liquidation;
- either initiation of legal actions by the Bank, in accordance with the procedure established in the current regulations;
- or when the number of overdue monthly rate exceeds the objective limits (12 overdue monthly rates for all technical forms) with regard to counterparties with instalment loans, except in the presence of out-of-court agreements and/or formalised repayment plans;

• after an in-depth assessment if the following events have occurred:[15]

- or admission to the extraordinary administration procedure, in the hypothesis that there are no concrete prospects of recovering the economic and financial equilibrium of the business activities;
- or judicial acts promoted by third parties;
- or cessation of business activities;
- or voluntary liquidation;
- or request/admission to the arrangement with creditors if it can be considered that the state of crisis coincides, in fact, with the state of insolvency.

[13] KAGAN J., Resolution Trust Corporation (RTC), Investopedia, 2019, consultabile nel sito https://www.investopedia.com/terms/r/resolution-trust-corporation.asp

[14] Ivi

[15] REFERENCE FOR BUSINESS, Resolution Trust Corporation (RTC), consultabile nel sito https://www.referenceforbusiness.com/encyclopedia/Res-Sec/Resolution-Trust-Corporation-RTC.html

As a general matter, in accordance with regulatory provisions (Central Office of Risks - Instructions for Intermediaries Participating in Supervisory Authorities), the listing as a bad debt implies an assessment by the intermediary of the customer's overall financial situation and cannot automatically result from a mere delay by the latter in paying the debt.

The dispute of the credit is not in itself a sufficient condition for a registration as a bad debt.

The recovery procedures become as severe as possible and end up jeopardising the position and credit reputation of the subject, as the report goes to the Central Office of Risks and could result in a judicial recovery which is likely to result in bankruptcy. "Sofferenza" in the Italian language is an impaired credit whose collection is almost impossible.

1.2.3 Probable Defaults: Forbone and Non-Forbone

They are less serious from the point of view of unrecoverability since they are those credit positions that the bank thinks it cannot collect, even partially, without ad hoc actions. The concept, rather than realistic, is probabilistic and therefore they must be distinguished from real bad debts.[16] We are facing situations linked to a temporary difficulty of the counterparty, which for various reasons, income or personal, is unable to pay. Although linked to probabilistic conditions, this category of anomalous items has however had its significant weight, and corresponds to approximately 40% of all NPLs categories in June last year.

Probable defaults are "all cash and off-balance sheet exposures of a borrower with respect to which the bank, in its judgment, believes it is unlikely that the borrower will meet its credit obligations in full (principal and/or interest) without recourse to actions such as foreclosure.[17]

This assessment is independent of the presence of any overdue and unpaid amounts (or instalments)." It is therefore not necessary that there is an explicit symptom of a credit anomaly such as non-payment, but an event that signals a probable default (e.g. a serious

[16] https://www.lavoce.info/archives/56291/inadempienze-probabili-il-nuovo-fronte-dellebanche

[17] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

crisis in the sector in which the debtor operates) may be sufficient. Probable defaults also include all exposures to borrowers that have failed to make on time payments (principal and/or interest) on listed debt obligations.

To this purpose, the grace period set out in the contract or, in its absence, acknowledged by the stock market listing is recognised. According to Circular 272, probable defaults include "all exposures to debtors who have submitted an application for composition with creditors (art. 161 of the Bankruptcy Law), which must be reported "from the date of submission of the application until the outcome of the application is known. However, it is understood that the exposures in question should be classified as non-performing if:[18]

• there are new objective elements that lead intermediaries, in their responsible autonomy, to classify the borrower in that category;

• the exposures were already non-performing at the time the application was submitted.

Forborne Exposures is the new category introduced by the EBA (European Banking Authority) that does not replace those listed above but is seen as an additional definition to indicate all the loans that are the subjects of concessions by the bank. Tolerance measures are included which tend to modify the initial contractual conditions.

These are tolerance measures towards those debtors who show temporary economic difficulties and who, for various reasons, are unable to meet their commitments with the bank.

The main purpose is to return exposures that are impaired to performing status and, conversely, to prevent so-called performing loans from moving into non-performing status. What makes this category very different from the others listed above is the fact that part of the Forborne can be both Performing Loans and Non-Performing Loans.[19]

In concrete terms, in order to establish the debtor's state of insolvency and to talk about an impaired loan, it is necessary that 90 days have passed since the debt was due; whereas, in order to have a forborne loan, it is sufficient that a position is past due even by only 30 days before the contractual renegotiation.

---

[18] Xolani Dastile, Turgay Celik, Moshe Potsane, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

[19] COSENTINO L.., Debt Restructuring & NPLs Management, Luiss Master in Corporate Finance & Banking- Major in Banking, 3 marzo 2020

The term "Forborne Performing Exposures" is therefore used in the case of customers who are somehow considered "healthy" but who are temporarily in a state of financial difficulty. The term "Non-Performing Exposures with forberance measures" refers to the three above-mentioned categories of NPLs, which are linked to debtors who are in a state of insolvency, i.e. their credit position is in deterioration.[20]

The main purpose of these forbearance measures is to reduce the interest rate or to facilitate the customer by rescheduling the plan. It is a common mistake to confuse the concept of forborne loans with that of restructured exposures. Restructured exposures are those exposures for which the credit institution decides to modify the initial contractual terms due to a deterioration in the economic conditions of the counterparty.[21]

Such changes may consist of a lengthening of the loan agreement, thereby generating a loss for the bank, as a longer term implies a lower interest income. Forborne Credits do not necessarily generate a loss for the bank. Therefore, while in the first case we are talking about a certain "loss", in the second case the prevailing concept is the "concession", i.e. the right that the assigning institution gives to its debtors in order to help them face situations of presumable economic difficulty.

A novelty that has greatly satisfied the expectations of small and medium-sized enterprises in the process of corporate restructuring, avoiding the weight of a negative judgement by rating agencies.[22]

A rating that weighs heavily on company's finances and can influence the cost of subsequent financing. If used correctly, concession measures can significantly reduce the amount of NPLs. Proper use means above all being able to respect the basic concepts of "effectiveness" and "efficiency":

• - efficiency, understood in terms of prudence in the choice of the overall amounts of funding for which the bank offers "forbearance" measures. It is true that lending measures aim at facilitating clients in difficulty, but this should not be translated into an "act of charity". It is essential to pay attention to the amount of the measures to be taken in order not to consider debtors who will then not be able to meet their contractual

[20] BANCA CENTRALE EUROPEA, Vigilanza Bancaria: “Linee guida per le banche sui crediti deteriorati (NPL)”,Marzo 2017

[21] BANCA D’ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

[22] Xolani Dastile, Turgay Celik, Moshe Potsane, “*Statistical and machine learning models in credit scoring: A systematic literature survey*”, Applied Soft Computing, June 2020

obligations. Debtors have different financial capacities and therefore the ideal would be to grant a degree of indebtedness that corresponds to their financial and economic capacities. Unfortunately, banks often follow their "commercial instincts", based solely on selling too much too quickly, a belief that seems to be mistaken and leads to premature and subjective judgements on the quality of the debtor. The same care and caution must be taken with debt rescheduling. A real threat, according to the Bank of Italy, is to leave debtors ample opportunity to take advantage of excessive forbearance measures for the restructuring of Non Performing Loans.[23]

• effectiveness, understood as the positive or negative outcome resulting from the adoption of such measures. It is not necessarily the case that a concessionary measure will always give the desired result, since, as with any renegotiation, many other factors come into play, mainly related to the behaviour of the debtor and its ability to keep its "contractual promises". In order to ensure the effectiveness of such measures, a thorough and accurate assessment of the counterparty is necessary, not based on trust and common sense but on documented and verifiable information.

1.3 NPLs in Italy

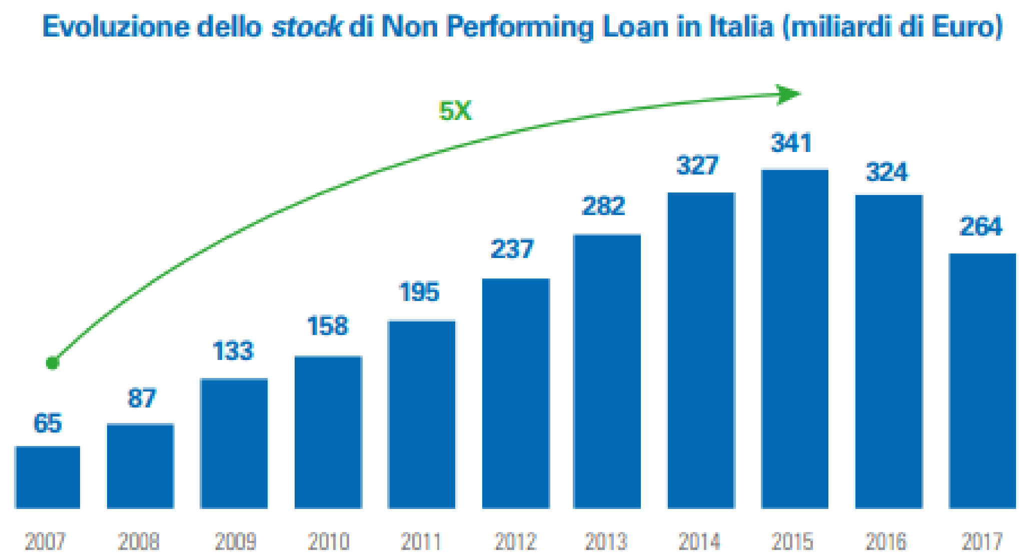
As mentioned above, Italy has been among the countries most severely affected by the financial crisis; as a result, its banking sector, which is the main lender to businesses and therefore exposed to the risk of corporate insolvency, also experienced a deterioration in asset quality. In fact, at the end of 2015, the amount of impaired loans of Italian banks, which in 2009 was 133 billion, had reached 341 billion gross of write-downs already accounted for (CAGR of 17%). The crisis was only one of the reasons why the amount of impaired loans in Italy increased so rapidly; other contributing factors were the sharp decrease in GDP and the poor efficiency of the Italian justice system.[24]

The issue of Non-Performing Loans is fundamental for the economy of the entire country: high levels of impaired loans have a negative impact on the general atmosphere

[23] COSENTINO L.., Debt Restructuring & NPLs Management, Luiss Master in Corporate Finance & Banking- Major in Banking, 3 marzo 2020

[24] COSENTINO L., Debt Restructuring & NPLs Management, Luiss Master in Corporate Finance & Banking- Major in Banking, 3 marzo 2020

of confidence in the banking system, causing an increase in the cost of funding and capital for institutions.[25] The stock of impaired loans in Italy, amounting to more than 264 billion Euros (of which about 60% represented by non-performing loans), is the highest in Europe. The peak was reached in 2015, with 341 billion Euros of Non Performing Loans in the portfolios of Italian banks, values more than five times higher than pre-crisis levels. The average NPL ratio of Italian banking groups (the ratio of gross impaired loans to total gross loans to customers) amounted to 13.5% at the end of 2017, a value close to that recorded in the weakest European economies and significantly higher than the European context, which presents an average NPL ratio of less than 5%. According to the Bank of Italy's estimates, three quarters of net non-performing loans in Italy are covered by real guarantees, only a small part of which are represented by industrial properties, which are more difficult to sell.[26]

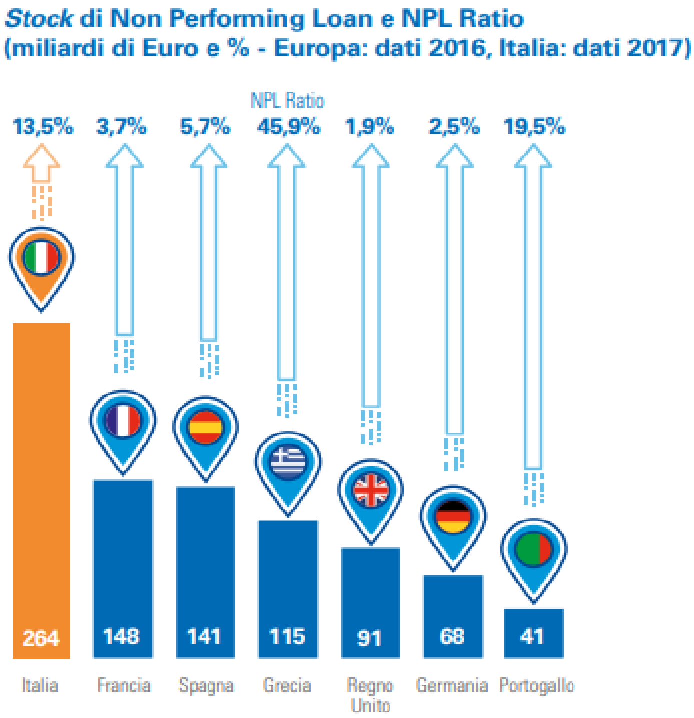


Evolution of the Non-performing loan stock in Italy (billions of Euro)

Source: KPMG Advisory Studies Department elaborations on Bank of Italy data

[25] Ivi

[26] KPMG, I Non Performing Loan in Italia, trend in atto e prospettive future, 2018, consultabile nel sito https://assets.kpmg/content/dam/kpmg/it/pdf/2018/07/I-Non-Performing-Loan-in-Italia-2018.pdf

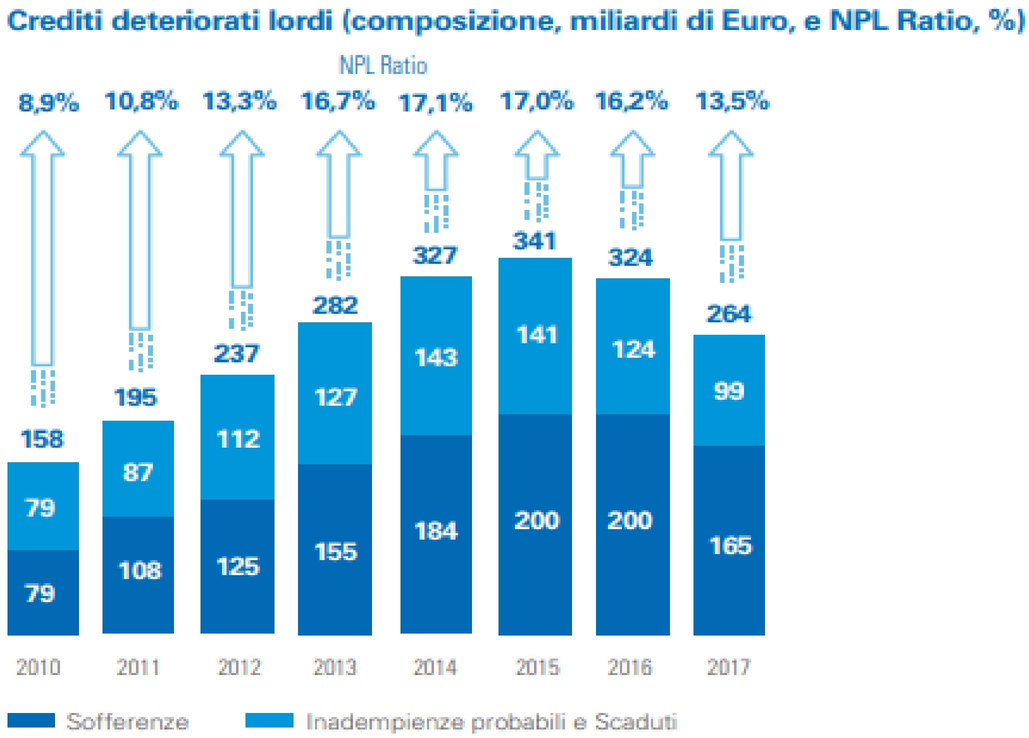


Non-performing loan stock and NPL Ratio

Source: KPMG Advisory Studies Department elaboration on European Parliament and Bank of Italy data

During 2016, Italian credit institutions reduced the stock of Non Performing Loans held in their portfolios for the first time in several years, a trend that continued throughout 2017. This trend was mainly supported by major deleveraging plans undertaken by major Italian banks to meet the requirements of supervisory authorities and investors' expectations.[27]

[27] EY, The Italian NPEs Market. From darkness to daylight, gennaio 2019, consultabile nel sito https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/emeia-financial-services/ey-the-italian-npes-market.pdf

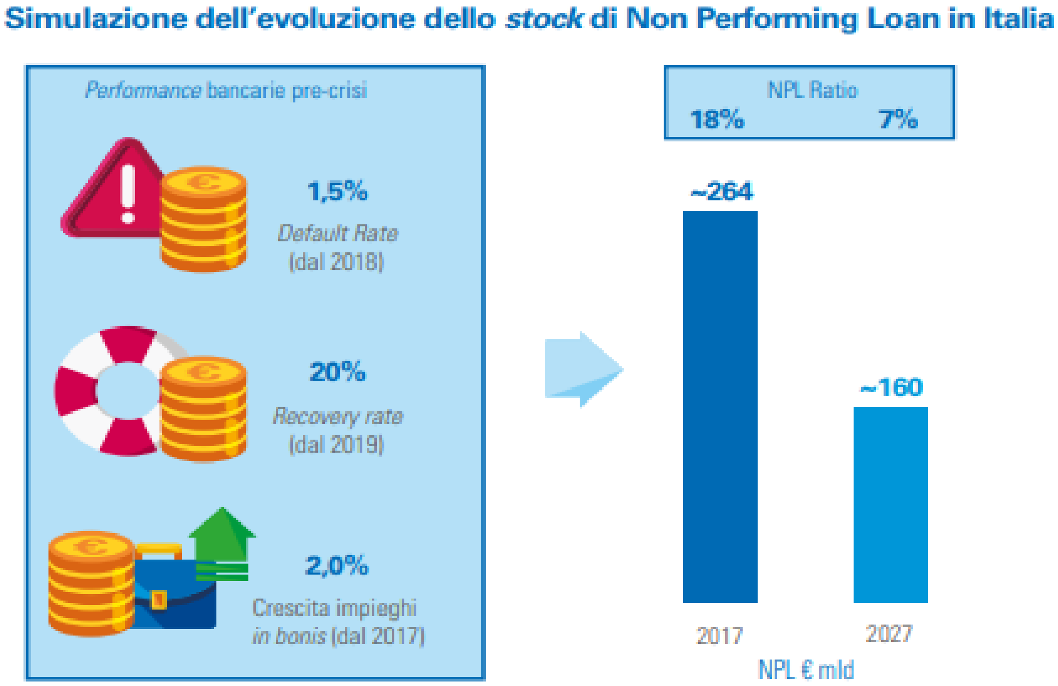


Gross impaired loans (composition, billion Euro, and NPL Ratio, %)

Source: elaboration of KPMG Advisory Studies Department on Bank of Italy data

The improvement in the macroeconomic context is leading to a 'normalisation' of new NPL flows (i.e., performing positions that 'slide' to non-performing) with a default rate (an indicator that relates NPL entries to the stock of performing loans at the beginning of the year) expected to be around 1.9% in 2017, significantly lower than the peaks recorded during the economic and financial crisis and close to pre-crisis levels (around 1.5% - 2.0%). On the recovery performance side, on the other hand, there was a continuous and profound reduction in the recovery rate (an indicator that relates collections and returns to performing loans to the stock of impaired loans at the beginning of the year), which stood at 11% in 2016 with an estimated slight increase in 2017 (13.4%). The reduction in the recovery rate is mainly due to three factors: on the one hand, the structure of the Italian judicial system, which is characterised by significantly longer procedural times than the average in other European countries; on the other hand, the operating capacity of Italian banks dedicated to the management of impaired loans, which is not structured to manage a phenomenon of such dimensions;

and thirdly, the difficult liquidity of collateral, mainly real estate, but also of a technical-industrial nature.[28]



Simulation of the evolution of the Non Performing Loan stock in Italy

Source: KPMG Advisory elaboration and estimate on balance sheet data of Italian banking groups

However, the regulator and the market are pushing for a 're-absorption' of the phenomenon in a shorter time horizon. On the one hand, the European Central Bank, through the publication of the 'Guidance to banks on NPLs' in March 2017, has pushed banks to define (and achieve) challenging targets for reducing the stock of impaired loans. Also for 2018, the ECB has the 'NPL node' among its supervisory priorities and further pressure from the regulator to reduce impaired loans is expected in the coming months.

On the other hand, the capital market assigns an increasingly marked correlation between the value of banks and the quality of their loan portfolio: banks with high NPL Ratios (generally above 15% - 20%) pay a significant discount in terms of capitalisation compared to their book value.

[28] EY, The Italian NPEs Market. From darkness to daylight, gennaio 2019, p. 5, consultabile nel sito https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/emeia-financial-services/ey-the-italian-npes-market.pdf

The regulatory environment in which banks are moving is complex and is creating significant pressure on them. Numerous changes have been introduced in recent years to address the crisis and to strengthen market confidence in financial institutions.[29] Many of these changes have directly or indirectly affected the issue of Non-Performing Loans, one of the topics of greatest attention and interest for supervisory authorities. Some of these regulatory changes have already entered into force, such as the Single Supervisory Mechanism (SSM), the new classification of impaired loans issued by the Bank of Italy and the European Central Bank's guidelines on Non-Performing Loans, while others are still being defined or awaiting implementation, such as IFRS 9, the guidelines for Less Significant Banks and Calendar Provisioning. Many other reforms have affected the banking system in the recent period, such as PSD2, and others will come into force in the coming years, such as the recent Basel III provisions on the calculation of Risk Weighted Assets, effective from the 1st January 2022.

The evolution of the regulatory framework on Non-Performing Loans implies significant impacts for banks:[30]

• Identification of strategies and governance models dedicated to the NPL segment;

• Increasing attention to the valuation and management of collateral;

• Strong focus on management processes and organisational structures;

• Importance of predictive management;

• Focus on monitoring and classification based on performance irregularities (e.g. overdue more than 30 days);

• Focus on financial/cash flow logic/methodologies;

• Detection/explication of implicit restructuring processes;

• Further structuring of NPL valuation and write-off logics and models.

[29] SPOTO D.M., TORINI D., KPMG, GACS Securitisation deals in the Italian NPL space, 27 aprile 2020, p.6, consultabile nel sito https://assets.kpmg/content/dam/kpmg/it/pdf/2020/04/GACS-Securitisation-deals-in-the-Italian-NPL-space.pdf

[30] KPMG, I Non Performing Loan in Italia, trend in atto e prospettive future, 2018, consultabile nel sito https://assets.kpmg/content/dam/kpmg/it/pdf/2018/07/I-Non-Performing-Loan-in-Italia-2018.pdf

1.4 Impact of NPLs on Financial Intermediaries

Just as the presence of loans in a bank is synonymous with profit (through interest rates), the presence of impaired loans is synonymous with loss. The resulting costs are multiple, including: the legal costs to be paid for recovery procedures, the cost of staff and resources involved in management, the loss of the payment itself, and the indirect costs deriving from the negative impact on image. The presence of high NPL stocks could also be read as an explicit index of inefficiency, both in terms of evaluation during the concession of loans and in terms of credit recovery. Appropriate funds have to be provided to cover the risk that NPLs represent, both by their own will and by the regulatory requirement to maintain a sufficiently high regulatory capital, which translates into reduced profits and missed reinvestment opportunities.[31] The sum of these aspects leads investors to mistrust the performance of the intermediary in question and to significantly raise the cost of financing sustained by him (or the risk premium demanded by the first parties). It has been proven that a high level of impaired loans modifies the behaviour of the intermediary who, in turn, will be more diffident in granting credit (credit crunch). Its risk aversion limits its possibility of expansion and its ability to take advantage of profitable investment opportunities. It is therefore intuitive to understand how it is a high priority for the bank to minimise the amount of its NPLs.[32]

Three main factors have undoubtedly caused the increase in NPLs in Italy over the past decade:

• the economic recession (due to the extension of the 2007/2008 financial crisis);

• the extreme slowness of credit recovery procedures;

• the absence of a secondary market for NPLs. In addition, imprudent lending policies, excessive tolerance of debtors (with sometimes dubious risk profiles), and conflicting or openly fraudulent payments were aggravating factors.

[31] Xolani Dastile, Turgay Celik, Moshe Potsane, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

[32] Ibidem

The capacity to process credit recovery procedures in Italy was half that of the rest of European countries. In other words, the time taken to collect debts was doubled (in turn a reflection of the generally long time taken by civil procedures in Italy), which meant that the NPL stock tended to assume significant dimensions when compared to other European countries.

Between 2014 and 2017, the average recovery time was 8.5 years. The problem did not arise only from 'out-of-bank' procedures, but also from internal procedures. The attention devoted by financial intermediaries to debt recovery was much lower when compared to the attention devoted to other operations.[33] The empirical reflection of this fact was visible in the scarce number of dedicated staff, in the scarce material resources, in the low level of informatization (files were mainly managed in paper format) and in the lack of appropriate databases that structured the information in an appropriate way. Operating "manually", a sudden increase in the number of files for an already reduced available staff, generated a stock of these files even within the banks. This manuality also had the aggravating factor of not making it easy to prioritise files. Algorithms applied in recent times have shown how it is possible to construct a list of files to be prioritised, thus enabling managers to deal first with those customers with significant exposures from which it is expected, on the basis of their characteristics, to obtain a higher percentage of the claimed credits. The focus is then on those counterparts for which there is a higher potential for recovery, avoiding wasting energy. For the manager working with paper formats, this type of prioritisation was not easily possible. The very fact that there are specialised debt recovery operators in the market with significant returns is at least partial evidence of the inefficiency of the banks' debt recovery procedures. The almost total absence of such operators in the Italian market was another reason for the high levels of NPL stock in Italy.[34] While this is not a direct cause of NPLs, the lack of a developed market for the purchase/sale of impaired loans was certainly an additional obstacle to their elimination, thus increasing the stock and widening the gap with other countries where such market existed. The main reason why there was not a sufficiently developed secondary market of NPLs in Italy is to be attributed to the so-called bid-ask spread. It refers to the difference between the price at

[33] COSENTINO L.., Debt Restructuring & NPLs Management, Luiss Master in Corporate Finance & Banking- Major in Banking, 3 marzo 2020

[34] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

which banks were willing to sell their impaired loans and the price at which the market operators specialised in credit recovery were willing to buy them. In turn, the reason for this difference in the two prices, supply and demand, stemmed from the different valuation criteria used by the banks to write the value on the balance sheet of these loans and the criteria used by potential investors. The difference in criteria can be summarised in the following two points:[35]

1. Since the loans had a high repayment risk, the rate of return demanded by the investors was much higher than the rate used by the banks for accounting purposes (in accordance with the IAS/IFRS accounting standards followed at the time), i.e. the original effective rate on these assets (much lower). By using a higher rate of return than the banks, the investors were faced with a much lower present value when assessing present value than what was calculated in the bank accounts;
2. The indirect costs of managing NPLs were taken into account by the banks, in accordance with accounting standards, in the exercise of their competence. Investors, on the other hand, deducted this amount immediately, further reducing the net value and thus the purchase price.

For obvious reasons, it is clear that the extended recovery time further worsens this situation. Direct and indirect management costs are higher if the recovery time is longer. Similarly, longer recovery times imply a higher probability of not recuperating the loans at all or a deterioration of the present value (even if only taking inflation into account), which is reflected in a higher rate of return (or risk) required by the investor.

The visible reduction in NPLs, especially after 2015, derives primarily from a strong legislative reaction in Italy. The various interventions include:[36]

• Regulatory interventions on the tax regime applied to banks' loan losses. The taxation of banks' loan losses has been revised, mitigating their procyclicality, encouraging the adoption of more prudent loan valuation policies and contributing to the transparency of banks' financial statements;

---

[35] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

[36] COSENTINO L.., Debt Restructuring & NPLs Management, Luiss Master in Corporate Finance & Banking- Major in Banking, 3 marzo 2020

• Reforms of the Bankruptcy Law and of the Code of Civil Procedure approved in 2015 and 2016, aimed at reducing the time and increasing the effectiveness of bankruptcy and enforcement procedures;

• Measure establishing a state guarantee mechanism (the "GACS") on securitisation transactions of impaired loans;

• Since 2016, the Bank of Italy has required banks to compile a statistical report containing very detailed data on individual non-performing positions.

Moreover, in March 2017, the European Central Bank published the "Guidelines for banks on impaired loans (NPLs)" with the aim of identifying best practices, constantly monitoring and supervising developments in NPL management, and promoting a higher timeliness of allocations and write-offs. Another substantial change for the Italian banking system is the adoption, from 1st January 2018, of the new international accounting standard IFRS9, which replaces IAS39. Among the many changes, there are mainly two that affected impaired loans:

• New classification and measurement of financial instruments;

• New credit devaluation model (impairment);

In the new classification it is necessary to consider in conjunction both the business model adopted by the bank in the management of financial instruments, trying to understand if the financial assets are held to collect and/or to sell, and the contractual characteristics of the cash flows of the individual assets using the SPPI test (Soley payment of principal and interest test). Assets can therefore be classified into the following categories:[37]

a) Financial assets held at fair value through profit or loss: these are instruments held primarily for trading purposes, instruments for which the fair value option has already been exercised and all other financial instruments that are not classified in the following categories;

b) Financial assets at fair value with impact on comprehensive income: this is measured at fair value and has an impact on the equity item "Valuation reserves". Financial activities include: those owned as part of a business model whose objective is achieved both by collecting cash flows and by selling the asset; those whose contractual terms originate,

[37] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

on specific dates, cash flows from nominal capital reimbursements and interests calculated in relation to the amount of the residual nominal value.

c) Financial activities valued at amortised cost: these are assets that must respect the same conditions as the previous letter with the difference that the accounting valuation is not made at fair value but at amortised cost.

The new loan impairment model requires a new calculation system for financial instruments to define the adjustments to the value of loans (impairment), in relation to the relative deterioration of their credit quality. Since the previous model, known as the incurred loss model, did not properly capture losses, the new model, known as the three buckets model, involves the recognition, and therefore the identification of provisions, of expected losses based on the degree of deterioration of the credit risk of financial instruments.[38] It is not therefore necessary that there is an explicit event or signal communicating the credit deterioration but it is sufficient to "suspect" it, given the information available. With respect to the assessment of expected losses, the new model is prospective (forward looking) as the estimate of expected losses must be made using verified and available information that takes into account not only historical and current data, but also prospective data. The three categories for the assessment are:[39]

1. Performing (Stage 1): Positions with low credit risk. The estimate of the expected loss is considered with reference to the relevant collective portfolios for a period of one year;
2. Under Performing (Stage 2): Intermediate level credit risk. These are positions that, while still non-critical, have suffered a deterioration in their rating or have shown clear economic difficulties and credit defaults. The expected losses (forward looking) are assessed over a period of time equal to the remaining contractual duration (lifetime);
3. Non Performing (Stage 3): Positions with a high credit risk for which the loss has already occurred or is almost certain to occur. The calculation of expected loss is performed analytically in relation to individual impaired positions, in proportion to the residual life of the individual exposure (lifetime).

[38] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html
[39] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

CHAPTER 2:
Methods and Parameters for the Valuation of NPLs

2.1 Forecast models based on judicial recovery

When the credit under evaluation is against a debtor that does not have the ability to repay it, whether the creditor has formally declared default or in substantially comparable cases, the valuer uses forecasting models based on judicial recovery using the following elements:[40]

a) analysis of the nominal amount of the credit (also known as Gross Book Value or GBV) to determine how the capital and interest have been formed over time in order to cover the guarantees;
b) analysis of the existing guarantees, to determine any risks of revocation, especially for judicial mortgages, the degree in relation to other possible secured creditors, the possibility of activating them;
c) analysis of potential guarantees, such as the possibility of enforcing judicial mortgages on free or capacious real properties or of distraining movable property owned by the debtor or by third-party guarantors not yet in the procedure;
d) valuation of assets subject to actual or potential collateralisation;
e) analysis of other debts contributing to the same assets in order to determine the pertinent credit collection fee;
f) analysis of the judicial procedures - executive or insolvency - in progress in order to estimate the time of collection, the percentage of abatement with respect to the valuation of the asset, the costs of the procedure;

By way of example, the same credit of 100, in a forecasting model based on judicial recovery, may foresee a collection from zero to more than 100 depending on the elements illustrated above and, in particular, due to the effect of the valuation of the assets covered by the guarantee. Indeed, in borderline cases where the collateralized

[40] XOLANI DASTILE, TURGAY CELIK, MOSHE POTSANE, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

asset greatly exceeds the amount of the credit, even taking into account all negative elements, it is theoretically possible to expect a collection equal to the nominal credit at the time of valuation plus appropriately calculated interest until collection.

For defaulted credits, however, it is rare for valuation models to predict collections equal to or greater than 100% of the nominal credit and, especially due to the lower value of assets that a judicial sale process presents compared to a sale on the open market, generally the estimated collection is a percentage of Gross Book Value.[41]

Let us consider, by way of example, a mortgage loan with a nominal value of 100,000. The first thing to do is to check how much of the loan consists of booked and unpaid interest.

Excluding cases of usury and anatocism for which the interest is null and void, even for correctly calculated interest there are legal limits so that part of the 100,000, let us assume 10,000, can be downgraded to unsecured. Let us then assume that the collateralised property has a reference value of 80,000. Compared to this value, judicial sales generally lead to discounted values on the basis of the mechanism of abatements following unsuccessful auctions. In real estate executions, these discounts differ depending on the type of property and its location, e.g. for finished and vacant residential properties located in large cities with liquid, real estate market average discounts of 0-20% are considered; for building plots located in peripheral areas discounts of 60-80% can be achieved. The value of real estate used in the forecasting model can have different sources:[42]

a) the valuation carried out by the bank at the time of disbursement is a piece of information to be taken into account, but it generally has a very low informative quality for the purposes of the collection forecast;

b) the judicial valuation carried out by the expert appointed by the tax assessor (CTU), even in cases where it is not drawn up according to the best professional criteria, is in any case a relevant piece of information both because the expert has access to the official documentation and to the asset, and because the value thus determined represents the basis for the start of the auction process;

[41] Ivi

[42] ZHANG YU, GUAN YONGSHENG, YU GANG & LU HAIXIA, "*Recognizing and Predicting the Non-Performing Loans of Commercial Banks*", International Journal of Signal Processing, Vol.9, 2016

c) the last value of a deserted auction represents an unavoidable reference point because it is the value that, even if in a negative key, has been the subject of a market comparison; however, it must be considered that sometimes after a deserted auction, at the next auction the price may rise above the base value of the previous auction;
d) the value judicially fixed for the next auction still to be held;
e) the valuation carried out by an expert appointed by the valuator as part of the process of developing the flow forecast model; valuations are generally carried out on a purely documentary basis (so-called "desk valuation") or an external visit to the property (so-called "drive-by valuation").

The valuer, in determining the expected value from the sale of the real estate subject to the guarantee, will take into account all the elements mentioned above and the abatement statistics available for the courts where the recovery procedures have been or will be started, also taking into account the type, location, and state of maintenance and rental of the property.

The statistics of the courts are then relevant for recovery time estimation. In fact, the differences in average recovery time differ significantly between courts.[43]
It is not unusual, even in a forecasting model based on judicial recovery, for the valuer to make the assumption that some of the credits will be recovered by negotiation. If supported by statistical evidence from the creditor bank or, more often, from the servicer that will manage the recovery of the credits, the valuer may assume that certain credits, after the initiation of the judicial recovery process, will be the object of transactional agreements with the debtor or with third parties aimed at the interruption of the compulsory recovery procedure and the reimbursement of a part of the credit or the assignment of the credit discount.
Finally, the forecasting model must take into account the recovery costs and the costs of maintaining the collateralised asset. Recovery costs in turn are divided into legal fees, sales costs and portfolio management costs.[44]
The length of debt collection procedures results in significant costs for banks for several reasons. First, it is one of the main factors slowing down the process of NPLs disposal,

[43] JUSTIN A. SIRIGNANO, APAAR SADHWANI, KAY GIESECKE, "*Deep Learning for Mortgage Risk*", March 8, 2018
[44] PROVENZANO A.R., D. TRIFIRO', A. DATTEO, L. GIADA, N. JEAN, A. RICIPUTI, G. LE PERA, M. SPADACCINO, L. MASSARON & C. NORDIO, "*Machine Learning approach for Credit Scoring*", August 5, 2020

and thus the accumulation of the stock. Our estimates indicate that, other things being equal, a reduction in the recovery time from 5 to 2 years would reduce the incidence of non-performing loans on bank balance sheets by about half; the peak of non-performing loans reached in 2015, amounting to 210 billion before adjustments, would have been about 100 billion. Given the difficulties in correctly valuing NPLs, a high stock translates into greater balance sheet opacity, and thus higher funding costs (both for risk capital and debt). In addition, the main instrument to reduce NPLs quickly is divestment, but since market prices are well below book values, banks have had to make significant additional impairments in the last four years to divest NPLs, with negative effects on the income statement.

Secondly, there is strong evidence that the recovery rate of non-performing loans decreases as the duration of the procedure increases. In 2017, on non-performing loans "closed" after one year, there was an average recovery rate of 61 per cent, compared to 34 per cent for positions closed after 6 or more years.[45]

The Bank of Italy, while adopting the harmonised classification of NPLs developed by the EBA, has kept alive in parallel the distinction between non-performing and probable defaults, categories similar to those of "going concern" and "gone concern" debtors used at the international level. As the harmonised European supervisory reporting defined by the EBA does not include this distinction, the supervisory pressure to reduce NPLs is exerted in the same way on loans to all troubled firms, without distinguishing between those still in operation and those in liquidation. This risks creating incentives for banks to withdraw support from debtors who, if properly serviced, would still have a chance of returning to viability.

The risk is mitigated or disappears altogether in countries where the timeframe for handling business crises is short. In these countries the distinction between non-performing and probable defaults therefore becomes less important. This helps to explain why our country struggled to build consensus when reporting this issue to the European supervisor and regulator.[46]

In summary, recent banking regulation of NPLs has been primarily aimed at reducing credit risks in banks' portfolios, without paying particular attention to possible side effects. Moreover, such effects are potentially relevant only in countries where civil

[45] Ibidem

[46] COTUGNO M., "*Gestione e Valutazione dei Non Performing Loans*", Agosto 2018

justice times are particularly long, including Italy. In these countries, loans to companies with less solid credit worthiness, or to more 'opaque' companies (e.g. start-ups), might be provided more cautiously, at higher prices, or not at all.

The management of corporate distress could also be affected if banks, having to dispose of NPLs, decide to reduce their support to businesses and rely on guarantees. This issue is particularly relevant given that, partly as a result of the massive disposals of non-performing loans in recent years, about half of the NPLs on banks' balance sheets are now represented by probable defaults (43 per cent of total gross NPLs, 55 per cent net of adjustments).[47]

The industry that deals with these loans is completely different from that of non-performing loans, and requires a parallel development of the operators and skills necessary for the proper management of corporate restructuring. An analysis of the data over the last four years reveals a much more modest trend in lending to smaller and riskier companies of all sizes. Over the same period, banks have reported a progressive decrease in the degree of loosening of supply conditions; signs of increasing difficulty in accessing credit have come from businesses, especially the smallest ones.

Overall, these trends are consistent with the hypothesis that the revision of the prudential treatment of NPLs may have contributed to reducing the supply of credit to small enterprises and those with low credit worthiness, but it is not possible to identify the effect of the reforms without careful microeconomic analysis.[48]

However, it is necessary to reiterate that these negative effects, if any, depend on structural weaknesses that our country has been carrying for years, rather than on the new NPLs regulation. It is therefore necessary to mitigate any undesirable effects of the regulation in the short and medium term, but above all to bring civil justice times up to the standards of other major European countries, and to improve the functioning of corporate restructuring procedures.[49]

---

[47] CESARONI T., Prociclicità dei sistemi di rating del credito: come gestirla, Settembre 2015. Al sito: https://www.bancaditalia.it/pubblicazioni/temidiscussione/2015/2015-1034/index.html?com.dotmarketing.htmlpage.language=102

[48] Ivi

[49] BANCA d'ITALIA, Note di stabilità finanziaria e vigilanza, Aprile 2016

2.2 Forecasting models based on repayment flows or on company value

In cases where the debtor, despite being in default, continues to have the capacity to generate income and cash flows, the valuer estimates such flows over a reasonable time horizon. In particular, if the debtor is a corporation, the same economic and financial data projection and valuation methodologies used in corporate practice and doctrine are used; if the debtor is a natural person, simplified methodologies based on available income and financial information are used.[50]

In the case of a credit to a corporate entity, a forecast is made of the cash flow available to repay the debts for a congruent time horizon to which is added a terminal value indicative of the flow obtainable from the sale of the going concern or, alternatively, from the potential capacity to refinance the debt.

With respect to forecasting methodologies and valuations of performing companies, it is necessary to take into account a series of peculiarities, including:[51]

- the quality, timeliness and completeness of information is limited by specific circumstances;
- the assessment of the risk of realisation of plan forecasts, generally higher than a performing company, and assuming the most prudent forecasting scenario with respect to the various assumptions;
- the legal framework within which the plan is implemented;
- the positioning of the assessed credit with respect to other liabilities within the payment priorities agreed with the debtor or otherwise presumable from the context.

Banks employ a formalised strategy to optimise the management of NPLs by maximising the present value of recoveries. The strategy should be defined on the

[50] SIDDHARTH BHATORE, LALIT MOHAN, RAGHU REDDY, "*Machine learning techniques for credit risk evaluation: a systematic literature review*", Institute for Development and Research in Banking Technology 2020

[51] COTUGNO M., "*Gestione e Valutazione dei Non Performing Loans*", Agosto 2018

basis of an analysis of the bank's own management capabilities, the external environment, and the characteristics of the impaired portfolios, and should identify the optimal combination of different possible recovery actions: internal management or outsourcing to intermediaries specialised in credit recovery; restructuring and granting of forbearance; acquisition of guarantees; legal or extrajudicial procedures; and disposals (including securitisation transactions) with accounting and prudential derecognition of the assets sold.

In particular, banks prepare operational plans for the management of NPLs in the short (indicatively one year) and medium/long (indicatively three to five years) term, in which the objectives of closing out positions (including the return to performing loans) and the actions to be taken (e.g. entrusting positions or portfolios to specialised external managers, market sales, etc.) to achieve them are defined.[52]

Targets must be set at least in terms of the level of NPLs before and after value adjustments, in absolute terms and as a percentage of total credit exposures to customers; by way of example, we provide a model that could be used to represent and monitor the targets. For more complex LSIs, it may also be appropriate to segment portfolios in order to facilitate appropriate treatments for different NPL sub-portfolios (e.g. by customer type, technical form, size of position). In a period of economic improvement, such as the one underway at the date of publication of these Guidelines, the supervisory authorities expect strategies to include targets for significant reductions in NPLs that are realistic in their quantification and supported by consistent implementation programmes.[53]

An indiscriminate and immediate reduction is not required. Banks are expected to adopt a conscious and active policy of managing impaired loans.

Therefore, the strategies adopted need to be supported by solid quantitative analyses, which, taking into account the business context, highlight the advantages and disadvantages among possible alternatives. In order to assess the advantage of different strategies, it is important to take into account the indirect costs associated with maintaining a high level of NPLs in the portfolio and to have quantitative

[52] COSENTINO L.., Debt Restructuring & NPLs Management, Luiss Master in Corporate Finance & Banking- Major in Banking, 3 marzo 2020

[53] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

evidence on management costs as well as past cash flows, recovery rates and recovery times.[54]

The NPL management strategy should be fully integrated into the strategic and management processes of the company, such as the definition of the business plan/budget, the RAF, the ICAAP, the recovery plans and the remuneration and incentive policies of the intermediary. At least in the phase of preparation of the first plan, eventual internal reorganisation processes should be defined, if necessary, with related investments in IT and human resources, as well as the resources considered necessary to achieve the objectives set. The Supervisory Authority reserves the right to require banks with high levels of NPLs to submit the above business plans and annual updates, and will assess the adequacy of such plans with appropriate consideration of the corresponding levels of total impairments. Where banks consider that the identified strategy may not lead to an effective reduction in NPLs in the medium-long term, the supervisory authorities expect this to be reflected in adequate levels of coverage. Consistently, the supervisor may consider the application of second-pillar measures to ensure that risks arising from relatively high levels of NPLs are covered and to provide incentives to achieve reduction targets.

It is essential that banks have an adequate IT database in which to record and manage relevant NPL data, which is indispensable for making informed and timely management decisions and for monitoring and optimising management performance. Therefore, it is necessary that what follows is available in an organised manner:[55]

a) information allowing to verify the progress of management/recovery procedures and the related interventions carried out, starting from the first signs of anomaly to the possible executive procedures attributed to external lawyers and/or specialised companies, with possible impacts on the valuations of collateral assets (e.g. the sequence of judicial auctions and the related variation of auction base values, etc.);

[54] Ivi

[55] PROVENZANO A.R., D. TRIFIRO', A. DATTEO, L. GIADA, N. JEAN, A. RICIPUTI, G. LE PERA, M. SPADACCINO, L. MASSARON & C. NORDIO, "*Machine Learning approach for Credit Scoring*", August 5, 2020

b) detailed information on guarantees supporting the positions and enabling an assessment of the expected cash flows and, therefore, the recoverability of the exposures.

A database with the above characteristics is also useful for possible analysis by investors interested in purchasing a portion of the NPL portfolio. The granular reporting on non-performing positions and the archive of historical losses on defaulted positions, provided by the Bank of Italy, are a useful starting point for the definition of the database.

2.3 Statistical forecasting models

In many cases, the large number of credits under assessment and the time available do not allow the development of a detailed forecasting model of the expected flows from all credits and it is therefore necessary to use a statistical approach.

Such an approach may relate to a portion of a portfolio or to the entire portfolio under assessment. In the case of credit portfolios that can all be valued using an analytical approach, e.g. a portfolio of mortgage loans, but that are composed of such a number of loans that the exercise is impossible within given time and cost limits, the valuer uses a sampling exercise.[56]

By means of example, if the portfolio consists of 1,000 positions and, on the basis of the restraints, it is only possible to value 150 of them, the valuer will segment the portfolio by amount classes and select 100% of the positions in the highest amount class, 50% of the second, and so on. At the end of the flow forecasting and valuation exercise, he will reverse the results of the analysis carried out on the entire portfolio.

The valuation of portfolios of granular unsecured loans to individuals is another case for which a statistical approach to cash flow forecasting is almost always applied. Since it is often impossible, as well as insignificant, to perform an analytical valuation of

[56] XOLANI DASTILE, TURGAY CELIK, MOSHE POTSANE, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

individual credits, assumptions based on the analysis of time series of credits with similar characteristics are projected.[57]

The starting point for the analysis is a database provided by the assignor regarding the collection history of credits of the same type and, once the reliability of the database has been verified, two approaches can be used.

The main methodologies can be summarised as follows:[58]

- approach based on recovery curves;
- comparative approach.

In order to draw up the recovery curves, it is necessary to collect data on write-offs and collections over a sufficiently long period, at least 7 or 10 years. The method is based on the assumption that each year of default generates a recovery flow over time according to a recurring curve. If this is the case within a certain confidence interval, then the curve so determined is used to forecast future flows.

Thus, for example, if the last 10 years of non-performing loans have generated average collections of 3.11% of face value in the first year, 6.28% in the second, and so on up to an average of 32.69%, it is assumed that all non-performing years have similar behaviour. Therefore, loans with one year's seniority being assessed, which have already generated collections equal to 3.08% of the face value, are expected to collect a total of 29.61 (32.69-3.08) in nine years, those two years old 23.05% (32.69-2.98-6.66) in eight years, and so on.

[57] COTUGNO M., "*Gestione e Valutazione dei Non Performing Loans*", Agosto 2018
[58] BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 giugno 2017, consultabile nel sito: https://www.bancaditalia.it/media/views/2017/npl/index.html

Table 1: Example of recovery curves

	Anno defult	Defult+1	Defult+2	Defult+3	Defult+4	Defult+5	Defult+6	Defult+7	Defult+8	Defult+9	Defult+10	Totale
Annata t-10	3,22%	6,54%	7,44%	3,99%	1,54%	4,78%	2,76%	1,88%	0,30%	0,10%	0,00%	32,55%
Annata t-9	4,01%	6,00%	7,88%	4,09%	1,09%	5,23%	2,60%	2,10%	0,25%	0,05%		33,30%
Annata t-8	2,99%	6,08%	8,21%	4,32%	1,00%	3,99%	2,54%	1,75%	1,00%			31,88%
Annata t-7	3,15%	7,00%	8,30%	3,78%	1,30%	4,21%	3,01%	1,90%				32,65%
Annata t-6	3,00%	6,43%	8,23%	3,22%	1,43%	5,35%	2,98%					30,64%
Annata t-5	2,91%	5,43%	7,99%	4,67%	1,22%	4,89%						27,11%
Annata t-4	3,09%	5,20%	7,50%	4,02%	1,87%							21,68%
Annata t-3	3,06%	6,09%	6,98%	3,85%								19,98%
Annata t-2	2,67%	7,32%	9,00%									18,99%
Annata t-1	2,98%	6,66%										9,64%
Annata t0	3,08%											3,08%
Totale	**3,11%**	**6,28%**	**7,95%**	**3,99%**	**1,35%**	**4,74%**	**2,78%**	**1,91%**	**0,52%**	**0,08%**	**0,00%**	32,69%
Cumulata	3,11%	9,38%	17,33%	21,32%	22,67%	27,41%	30,19%	32,10%	32,61%	32,69%	32,69%	

By comparative approach, we mean the method assuming that loans with similar characteristics have similar probability of being collected. Thus, for example, if loans to people aged 45-50, resident in Lombardy, with an income at the time of disbursement of more than €30,000 and a high school education, are collected at an average rate of 35% three years after being placed on non-performing status, I will apply this assumption to all loans with such characteristics.

2.4 Credit valuation

The market valuation of loans is mainly carried out through the method of discounting financial flows according to the classic formula:[59]

$$V = \Sigma(FC_{t=0},..FC_{t=n})*1/(1+i)^t$$

As we have seen in the preceding paragraphs, the determination of the flows to be applied is the result of forecasting models based on the credit category.

Once the expected net flows have been determined, the discount rate remains to be determined in order to arrive at the loan valuations.

[59] COTUGNO M., "*Gestione e Valutazione dei Non Performing Loans*", Agosto 2018

The determination of this parameter, which is decisive for the results of the valuation exercise, is not carried out according to commonly accepted methodologies in line with other fields of application of the discounted cash flow method. In particular, the Capital Asset Pricing Model is not applicable to the impaired loans market as it lacks a liquid and transparent market on which to measure a beta coefficient of variability. As a result, in practice, the discount rate for the market valuation of impaired loans depends on the investors' target return and varies with:[60]

a) perceived country risk at the time of valuation;
b) type of loans;
c) perceived uncertainty in cash flows.

The specific risk component of Italy in the discount rates used in impaired loan market moves in line with the trend of the spread between Italian and German government bonds.

As regards the type of loans, it is not possible to quantify exactly the premium required for certain categories of loans with respect to others, but it is possible to establish a hierarchy of riskiness in relation to:[61]

a) type of collateral (secured / unsecured);
b) type of debtor (private/corporate);
c) type of asset under guarantee (residential/commercial/industrial);
d) asset status (finished/under construction/construction land);
e) debtor location and guarantees (north/centre/south);
f) valuation methodology (analytical/statistical).

With the necessary exceptions, therefore, an unsecured loan to a company located in Sicily will be valued using a much higher discount rate than a mortgage loan to a natural person secured by a mortgage on a flat located in Milan.

In most cases, the portfolios of loans being valued are mixed and the valuer has two options: to apply a single discount rate that takes into account the portfolio mix; or to apply a model with multiple rates.

Finally, the valuer, in determining the overall risk premium included in the discount rate, will take into account the degree of aggressiveness used in the development of the

[60] JUSTIN A. SIRIGNANO, APAAR SADHWANI, KAY GIESECKE, "*Deep Learning for Mortgage Risk*", March 8, 2018
[61] Ibidem

flow forecasting model and the intensity of the originator's processing of the portfolio prior to sale.

Finally, a number of circumstances related to the transaction's legal and contractual structure may affect the price paid. To mention the most relevant factors:[62]

a) contractual provisions in terms of the seller's representations and warranties have a significant impact on the risk perceived and therefore on the discount rate;
b) the manner and timing of payment of the price;
c) the legal and tax structure of the transfer of credits and guarantees;
d) any accessory conditions to the transfer that generate an economic advantage or disadvantage for either party;
e) servicing agreements between buyer and seller (in both senses) and related remuneration;
f) option or pre-emption rights on further portfolios;
g) access to particular forms of guarantee and/or financing (e.g. GACS).

The credit valuation process is very complex, uses a series of subjective assumptions and is affected by strong information asymmetry between the parties. As a result of these circumstances, it is not unusual for different potential buyers to arrive at valuations that differ by more than 30% in an auction process for the same portfolio.

2.5 Initial valuation and market values

The book value at which impaired loans are booked is certainly higher than the price that market buyers are willing to offer. This difference is mostly due to the different valuation criteria applied by banks compared to those used by speculative investors. There are two factors that underlie these differences and hinder the formation of an Italian secondary market for NPLs. The first factor is related to the interest rate required by speculative investors compared to the rate used for the preparation of financial statements and established in the international accounting standards (IAS-IFRS) to discount the expected cash flows from NPLs. Banks use a much lower effective interest

[62] XOLANI DASTILE, TURGAY CELIK, MOSHE POTSANE, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

rate which, when used as a discount rate, results in a much higher present value/price than the market value.[63]

The second factor concerns the indirect costs of managing NPLs. Banks recognise them in their balance sheet, while speculators deduct them from the purchase price.

A direct consequence of the above is that the element included in the formula that is mostly removed from the discretion of the evaluator is time factor.

A shortening of the recovery procedures would significantly raise the current value of NPLs. This would certainly give an impetus to a larger supply on the market and a decrease in the total amount of impaired loans on banks' balance sheets. The formula for assessing the present value (PV) of loans gross of value adjustments is the following:[64]

$$VA = \sum_{t=1}^{n} \frac{F_t}{(1+i)^t}$$

- F are the expected cash flows arising from the receivables;
- the time (t) exponent in the denominator reflects the time value of money rule.

The formula is used to determine the net value of impaired loans and, if the borrower has problems repaying the loans, the following must be assessed:

a) the likelihood of not being able to recover the full amount within the agreed time frame;

b) the actual recoverable amount, which is often a function of the guarantees provided;

c) the time frame in which recovery will actually take place.

In a second step, the flows have to be re-estimated and the value adjusted downwards. In the revaluation, banks will also take into account the direct costs of managing NPLs. The net book value (NBV) of an NPL will be:

$$NBV = \sum_{t'=1}^{n'} \frac{F_{t'}}{(1+i)^{t'}}$$

The value adjustment is equal to the differences between VA and NBV

Value adjustment = VA-NBV

[63] COTUGNO M., "*Gestione e Valutazione dei Non Performing Loans*", Agosto 2018
[64] Ivi

In subsequent periods it is possible that loans will deteriorate further or that they will regain value. The negative or positive difference is the result of the adjustment sum.

2.6 Valuation criteria according to IFRS 9

As from 2018, IFRS 9 accounting standards have come into force in Europe, providing for stricter and more stringent rules for making adjustments to loans.[65] The valuation of provisions will no longer take place exclusively as a coverage of impaired loans (incurred loss), but on the basis of expected losses. The main criteria by which assets and liabilities are classified under IFRS 9 depend on two factors:[66]

1) the business model;

2) the contractual characteristics of the financial instrument. Unless reclassification is necessary, the designation of assets and liabilities takes place in the stage prior to their registration in the financial statements and is normally not modifiable. There are three business models for classification that are mentioned in IFRS 9.[67]

1) The first model concerns activities that one intends to hold until expiration and from which one intends to collect only and exclusively their cash flows.

2) In the second model there is a dual purpose; both to obtain cash flows and to generate profits from selling the asset before its natural expiration.

3) The third model encompasses those assets whose purchase and management have the exclusive purpose of profiting from price fluctuations in the financial markets. Business models are linked to two applicable valuation models.

In the first business model, the amortised cost criterion is used except for the way in which impairment is carried out. Portfolios valued at amortised cost provide greater stability in economic results and company margins. Assets classified in the second and third models are instead measured at fair value but with different accounting impacts depending on the business model.

In the second model, a change in the valuation of activities is counterbalanced by an equity reserve (fair value over comprehensive income FVOCI).

---

[65] IFRS 9 par. B.5.5.17, Al sito: http://www.revisorionline.it/IAS_IFRS/ifrs9- p1.htm

[66] PROVENZANO A.R., D. TRIFIRO', A. DATTEO, L. GIADA, N. JEAN, A. RICIPUTI, G. LE PERA, M. SPADACCINO, L. MASSARON & C. NORDIO, "*Machine Learning approach for Credit Scoring*", August 5, 2020

[67] Ivi

In the third model any change in fair value is recognised in the income statement. The second key aspect is the type of cash flows expected in the contract. There are two categories here: solely payment of principal and interest (SPPI) or non-SPI, including all exposures not belonging to the first category. Passing an SPPI test is a necessary condition for classifying an asset at amortised cost; the following requirements must be met simultaneously:[68]

1) the principal must correspond to the fair value at initial recognition;

2) the interest earned must represent the consideration for money's time value, remuneration of risk factors, costs and a profit margin.

It is clear that fair value is used in all other cases, i.e. when the test is not passed. Always keeping in mind the different perspectives between regulatory and accounting expected losses, we analyse the parameters of the impairment model for the calculation of expected loss (EL).

In particular, the expected loss is synthetically the difference between contractual and expected flows. Before proceeding to the calculation of provisions on individual positions, it is essential to establish the reference time horizon. The expected loss can be calculated on a 12-month annual basis (12-month expected credit loss) or on a multi-period basis (lifetime expected credit loss ECL).

The first method represents the loss portion referring to 12 months after the calculation date, while the multi-period method considers the credit's entire life. A further step before quantifying prudential provisions on exposure lines is to divide NPLs into three classes. Stage 1 includes exposures whose credit risk has not increased since the date of recognition. Stage 2 comprises exposures whose credit risk has increased significantly. Stage 3 includes impaired loans (past due, unlikely to pay and non-performing loans).[69]

During the assessment it is necessary to start from the original risk and to quantify the increases in credit risk in order to determine the stage transition of a credit line.

Such a transition leads to a significant increase in provisions and it is important to establish general operating rules. IFRS 9 contains methods and valuation rules to be

[68] XOLANI DASTILE, TURGAY CELIK, MOSHE POTSANE, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

[69] Ivi

applied in this delicate phase. When a debt is more than 30 days past due, there is an indicator of increased risk and the responsibility is on the bank to prove (using historical customer data) that the delay in payment does not indicate increased risk. The general formula for calculating the lifetime expected credit loss (LECL) is the following:

$$LECL = \left(\frac{\sum_{t=1}^{n} MPD_t \cdot LGD_t \cdot EAD}{(1+i)^t}\right)$$

In which:

MPD is the marginal probability of default in year t;

LGD is the loss given default in year t;

EAD is the exposure value;

The LECL quantifies the prudential provisions that must be made and is the sum of three components discounted at a rate i.

Each component must be included in the formula on a time reference basis. For exposures classified as stage 1, the probability of default and LGD for the first year must be considered. For stage 2 exposures, the entire life of the financial instrument including the effects of possible changes in the borrower's economic circumstances must be considered using an effective (risk-adjusted) discount rate. In particular, point-of-time (PIT) ratings are used to estimate PD.[70]

These are ratings associated with PD that reflect the current characteristics of the counterparty, with reference to the conditions of the sector in which it operates and economy as a whole.

In recessionary periods, the PIT approach allows the application of more truthful rates with the effect of migrating to rating classes that have a higher riskiness. The logic for estimating accounting LGD should not be confused with the logics for regulatory LGD calculation.

The logic remains the PIT logic in which a macroeconomic forecast is used and the assessment also takes into account the effect of all guarantees but not the indirect costs of recovery. Similar to the other parameters, the EAD should not represent the mere

[70] ZHANG YU, GUAN YONGSHENG, YU GANG & LU HAIXIA, "*Recognizing and Predicting the Non-Performing Loans of Commercial Banks*", International Journal of Signal Processing, Vol.9, 2016

book value but be representative of the macroeconomic scenario and incorporate future forecasts (forward looking).[71]

2.7 Securitisation

The Italian market of asset-backed securities transactions, ABS, mortgage-backed securities, MBS, has been developing at a moderate and steady pace since the 1990s.
At present in Italy, almost all banks and other financial intermediaries securitise leasing instalments, mortgages and consumer loans, which in fact account for most of the loans sold. In particular, however, reference is made to the cession of non-performing loans (NPLs).
In fact, financial intermediaries can sell either performing or non-performing loans to customers.
The term non-performing loans refers to those loans granted to borrowers that have already shown symptoms of anomaly in the past and have normally already been classified by the originator in one of the categories of doubtful loans. The first securitisation operations carried out in Italy concerned impaired loans, since the securitisation technique represented a valid instrument for financial intermediaries to manage problematic loans.[72]
Despite the presence of a significant process of innovation with respect to the type of assets sold and the structure of the transaction, banks and other financial intermediaries continue to represent the great majority of originators, and credits of various technical forms and nature, such as leasing instalments, mortgages, consumer loans and corporate bonds, the great majority of assets sold.
Particularly in cases involving the transfer of non-performing loans, the use of the transaction is subject to the presence of a precise regulatory framework that defines the benefits for the transferor and adequately protects the subscribers: the transfer of this class of assets to third parties at favourable conditions depends, in fact, on the availability of effective legal instruments for the management of insolvent debts.

[71] Ibidem
[72] SPOTO D.M., TORINI D., KPMG, GACS *Securitisation deals in the Italian NPL space*, 27 aprile 2020, consultabile nel sito https://assets.kpmg/content/dam/kpmg/it/pdf/2020/04/GACS-Securitisation-deals-inthe-Italian-NPL-space.pdf

In Italy, due to the costs and timing of credit recovery procedures, all other things being equal, the convenience of NPL securitisation is, in principle, lower than that of intermediaries operating in countries where the effectiveness of procedures is higher. Moreover, since the target of ABS emissions in Europe is represented by international institutional investors interested in subscribing bonds issued in relation to securitised assets mainly for portfolio diversification purposes, and not to increase the risk profile, in order to make the bonds attractive and obtain an adequate rating, the Italian originator has to obtain or directly lend higher guarantees.[73]

However, despite the fact that the risk is not fully transferred to the subscribers of the bonds, choosing an NPL securitisation may be more advantageous than following the route of a recourse sale; in the event that a liquid secondary market exists for the issued bonds, it is possible, in fact, to obtain a cheaper divestment price.
Since the pool of assets is made up of non-performing loans, the stability of the portfolio itself is strongly threatened by the greater likelihood of the occurrence of default events, which cause dangerous mismatch between the expected cash flows and the amounts owed to the subscribers of the bonds. Therefore, the credit enhancement techniques used and the ability to recover the credits sold are of primary importance for the success of the transaction.

In contrast to securitisation of performing loans, i.e. loans granted to performing customers for which regular repayment of the debt is expected on the contractual due date, securitisation of anomalous loans is a contingent technique whose main objective is to eliminate undesirable loans from the balance sheet. For performing loans, on the other hand, it is an active loan portfolio management practice adopted as part of a long-term strategy.[74]
Finally, the securitisation of NPLs may affect the originator's ability to engage in new operations and, more generally, may have negative effects on its reputation.

[73] PEZZUTO A., "*La garanzia dello Stato per le operazioni di cartolarizzazione di crediti in sofferenza*", in Diritto bancario, Diritto finanziario, 28 febbraio 2020, consultabile nel sito https://www.tidona.com/lagaranzia-dello-stato-per-le-operazioni-di-cartolarizzazione-di-crediti-in-sofferenza/;

[74] COTUGNO M., "*Gestione e Valutazione dei Non Performing Loans*", Agosto 2018

Although there are no particular limits concerning the quality of assets to be securitised for the above critical aspects related to securitisation of non-performing loans, it is possible to state that the optimal loans to be securitised are performing loans.

The implementation of a securitisation operation is always rather complex, even more when dealing with non-performing loans.

Setting up a securitisation transaction and creating an appropriate structure always presents problems that are not easy to solve, and these are even more difficult in the case of non-performing loans.

Before assessing the relative profitability of the transaction, as well as the advantages and disadvantages with respect to available alternatives, it is useful to review the various steps with respect to this particular type of underlying asset.

In particular, it is useful to focus on the following steps, which require the involvement of multiple parties:[75]

- identification of the portfolio involved in the securitisation process of the underlying risk and of related charges;
- definition of the operation's financial structure with particular reference to the characteristics of the titles to be issued and to the forms of support, credit enhancement;
- evaluation by rating agencies;
- operational and organisational aspects related to portfolio management and servicing and monitoring activities.

This articulation is valid in expositive terms, given that the final result of each transaction is the outcome of a complex process of interaction and comparison between the various counterparties involved, assignor, assignee, arranger, servicer, and raung agency, and the elements themselves are closely interconnected.

In fact, the quality of the credit portfolio sold and the quality of the service provided by the servicer can both be considered as external factors affecting the rating. Moreover, according to international practice, each level of rating corresponds to a minimum level

[75] SPOTO D.M., TORINI D., KPMG, GACS *Securitisation deals in the Italian NPL space*, 27 aprile 2020, consultabile nel sito https://assets.kpmg/content/dam/kpmg/it/pdf/2020/04/GACS-Securitisation-deals-inthe-Italian-NPL-space.pdf

of credit enhancement, the actual level of credit enhancement can be manipulated to ensure that the transaction is rated as required and as necessary.

If we want to define the boundaries of credit enhancement, we can say that very often there is a need to make ABS more attractive with respect to the characteristics of the securitised activities.

For example, it is possible to limit the exposure of ABS subscribers to the specific risks implicit in the portfolio of securitised activities: the credit risk that manifests itself with the default of the main debtor, the liquidity risk due to the mismatch between inflows and outflows, the reinvestment risk with early repayment by the assigned debtors, the interest and exchange rate risk in case of mismatching between flows collected from debtors and flows paid to investors in terms of interest rate or currency.[76]

In order to reduce the risks for the investor, to increase the rating of the issue and to reduce the rate of return required by the market, the quality of the underlying portfolio can be enhanced by forms of internal guarantees, i.e. provided directly by the original creditor, or external guarantees, i.e. provided by third parties.

In the former case, the originator does not fully transfer the credit risk to the transferee, but bears part of it.

Examples of internal guarantees are:[77]

- overcollateralisation: guarantee of the emission, an amount of higher underlying activities is transferred to the market value of the issued bonds;
- the establishment of a guarantee deposit: the originator establishes a deposit with another financial institution that can be used if the issued bonds are not repaid in full due to the insolvency of the assigned debtors;
- the ercess soread: this is a sort of reserve formed over time by payment of a positive difference between the return on the securitised loans and the return paid on the ABS. The SPV can draw on this reserve in the event of the assigned debtors' insolvency or simple late payment;
- the emission of subordinated tranches: titles are issued in several classes or tranches, each of which is characterised by a different order of priority in the distribution of cash flows. For example, bonds of the so-called senior class,

---

[76] Ivi

[77] ZHANG YU, GUAN YONGSHENG, YU GANG & LU HAIXIA, "*Recognizing and Predicting the Non-Performing Loans of Commercial Banks*", International Journal of Signal Processing, Vol.9, 2016

which guarantee payment in priority, are reserved for the final investors, while bonds of the junior subordinated classes, which have a lower rating than the senior classes or are not subject to any valuation, are generally retained by the originator.

It should be noted that through these forms of guarantees, the originator pursues and achieves the objective of limiting the risk of the emission and, therefore, the remuneration of the ABS, but does not free itself of the credit and liquidity risks implicit in the securitised portfolio.
The definition of the portfolio, potentially subject to securitisation, is connected, on the one hand, to the delimitation between assets susceptible and not susceptible of forming object of the securitisation, and, on the other hand, to the possibility or necessity of identifying the plurality of credits as a whole, since even reference to a debtor does not constitute an anomaly: in fact, it is possible that the debtor is also only one, perhaps for many debts or for only one of a significant amount.

The transformation of non-performing loans into negotiable instruments facilitates their transfer to other intermediaries or end-investors as a result of the improved combination of risk and return offered by the securitised credit titles.
In order to be transferred and to generate the flows necessary to pay interest and repay principal on the bonds, the bank assets must be:[78]

- individually identifiable in terms of loans and related cash flows;
- characterised by somewhat predictable cash flows in order to develop, on the basis of historical experience and the technical forms of the contracts, a model for predicting actual income;
- legally isolable because the credit and the rights deriving from it must be exercisable by the buyer or assignee without any link to the situation of the assignor, in order to limit the exposure of subscribers to the sole credit risk of the assigned portfolio and not that of the assignor;

[78] XOLANI DASTILE, TURGAY CELIK, MOSHE POTSANE, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

- characterised by a high degree of homogeneity and standardisation in terms of contractual nature and structure, but at the same time ensuring satisfactory risk diversification.

It is not possible to imagine that a securitisation can receive an acceptable rating if it concerns a plurality of credits which are very different from each other in terms of nature, maturity and origin; therefore, in economic terms, the banking assets, object of the transfer, must be composed of homogeneous entities, that is, aggregated and considered as a unit.

In the case of non-performing loans, the identification criteria must be used flexibly to assign to the pool the credits that the entity intends to sell, which are at the same time coherent with the risk-return profile of the titles to be issued. In this case, in fact, the pool does not generate, also because of the relative scarce availability of data on the past performance of analogous pools, stable and easily predictable cash flows and the revenues are not easily distinguishable between capital and interest shares to be destined respectively to the amortisation and to the payment of the interest respectively to amortisation and coupon payments.[79]

From an operational point of view, the sale requires certain preliminary activities to be carried out in order to determine and quantify the assets size, their value in terms of expected collections and the forecast of cash flows over time.

The analysis and definition of the portfolio selection criteria must consider the type of debtors and their geographical and sectoral diversification, the financial characteristics of loans such as rate, duration, date of granting and amortisation plan, the behavioral statistics analysed, depending on the case, for homogeneous pools or for single credits, with static or dynamic methodologies.

A particularly important factor in obtaining a rating is an in-depth analysis of the procedures for granting, disbursing, and managing credit, as well as the collection and recovery procedures used by the transferor.[80]

The conclusion of the transaction requires parameters and management control tools that allow the arranger to make an accurate assessment of the assigned portfolio. For

[79] SPOTO D.M., TORINI D., KPMG, GACS *Securitisation deals in the Italian NPL space*, 27 aprile 2020, consultabile nel sito https://assets.kpmg/content/dam/kpmg/it/pdf/2020/04/GACS-Securitisation-deals-inthe-Italian-NPL-space.pdf

[80] COTUGNO M., "*Gestione e Valutazione dei Non Performing Loans*", Agosto 2018

the purposes of the credit classification identifiable in block, loans may present, alternatively or jointly, different characteristics of homogeneity: the territorial areas of stipulation, the nominal value higher than pre-established limits, the circumstance that the appropriate legal actions for recovery have been started and the presence, total or partial, of a voluntary or judicial mortgage.

The guarantees protecting the loans, since they affect the financial structure of the title and its attractiveness for the subscribers, are in fact essential and securitisation of PL totally lacking of such guarantees is not conceivable, their presence makes possible the partial planning of the return flows and the evaluation of the capacity to cover the economic risk of the title, the capital refund and the payment of interests, otherwise, the credit enhancement level required to make the transaction acceptable for the market and with an adequate rating would be as high to make the transaction unprofitable for the originator.

For this reason, an important segment of bank-originated securitisation is represented by mortgages, which are more attractive and therefore easier to sell. In fact, guarantees are automatically transferred to the securitisation company, without any additional requirements or costs, and loans can also be recovered using the personal guarantees issued.[81]

A high degree of diversification of the pool makes it possible to improve the characteristics of the title to be issued and reduce its risk. In the rating agencies' practice, diversification is measured with the so-called diversity score which, considering the degree of correlation between the different categories of debtors, allows to reduce the risk of default or downgrading, but this methodology is not applicable to NPL securitisation, since the negative event has already occurred and diversification rather relates to factors generically defined as environmental or sectoral.

Although the analysis of credit risk and risk enhancement must always be referred to the individual case, the small size of the pool of assigned receivables and of the originator bank increases the potential risk of concentration, since securitisation translates into a lower sector and/or geographical diversification of debtors and the titles

[81] JUSTIN A. SIRIGNANO, APAAR SADHWANI, KAY GIESECKE, "*Deep Learning for Mortgage Risk*", March 8, 2018

issued are therefore exposed to risks dependent on the local economy, which is assumed to be more volatile than that of the entire country.
Considering the technical structure and the type of transferable credits, the assignee proceeds to assess the overall risk of the entire portfolio and the characteristics of the underlying cash flows; in relation to the outcome of this assessment, the transfer can take place non-recourse with no guarantees required to the assignor or third parties, or with guarantees such as the purchase of subordinated shares, letters of credit or insurance provided by third parties or, finally, other assets to guarantee the repayment of the obligation.[82]

The definition of transfer price is affected by a multiplicity of elements which make it effectively impossible to define a specific pricing model: in practice, to the economic value determined through the implementation of recovery forecasts for the estimated period sometimes is applied a further lump sum reduction in the face of a further credit deterioration and, more generally, of its recoverability conditions.

[82] SPOTO D.M., TORINI D., KPMG, GACS *Securitisation deals in the Italian NPL space*, 27 aprile 2020, consultabile nel sito https://assets.kpmg/content/dam/kpmg/it/pdf/2020/04/GACS-Securitisation-deals-inthe-Italian-NPL-space.pdf;

CHAPTER 3

Analysis of the result deriving from the use of Credit Scoring Models

3.1 Purpose of the Thesis

In this section we are going to analyse the accuracy of two different Machine Learning models in predicting whether a credit should be considered as performing or non-performing. In fact, very often banks and consulting firms are faced with a higher actual number of non-performing loans than originally assumed. This problem leads to having to reschedule debt collection from scratch, which is very time consuming.
In particular, as analysed in Chapter 1, debt collection is proceeding much slower than had been hoped for, which is why efforts must be made to reduce any waste of time in order to allow this market to function as efficiently as possible.
For the aforementioned reasons, the purpose of this thesis is to create a programme that, by means of Machine Learning techniques, allows the calculation of the accuracy of any dataset of credits already divided into performing and non-performing, in order to allow those who use it to understand what percentage of error there might be in the calculation of performing credit portfolios and thus predefine alternative recovery strategies, limiting time losses.

Machine Learning is a branch of artificial intelligence that includes statistical techniques, by which machines are allowed to perform and improve specific tasks. In particular, in order to develop the Phyton programme the two models most used in the NPLs research field in the last years by the most important consulting firms and banks were chosen: Logistic Regression and Support Vector Machine.

The most important decision to make for the use of Machine Learning models regards which dataset take into consideration. In particular, the different variables taken into account must be carefully chosen, as it is essential to avoid redundancies and thus allow the model to be as fair as possible. For this reason, the following Phyton programme was developed calculating firstly the accuracy taking into account all the variables in the initial datasets, then using a feature reduction method called Principal Component

Analysis (PCA) to eliminate the less useful variables and analysing whether the final results were in this way almost the same.

In order to develop this programme 7 datasets were taken into consideration. The datasets considered records loans of three different countries: Germany, Australia and Poland made by UCI Machine Learning Repository.83

The 7 datasets used contain data for the years 2008 to 2013 for the above-mentioned countries and they are already divided in performing, reported in the last column of each dataset with the number 0, and non performing, indicated with the number 1.

Below there will be the illustration of the composition of each dataset by showing the values divided into instances, which specify the number of loans, and features, which express the variables utilised for each debtor:

Germany: 1000 instances and 25 features;

Australia: 690 instances and 15 features.

The Polish dataset, on the other hand, is divided year by year:

year 1 7027 instances and 65 features;

year 2 10173 instances and 65 features;

year 3 10503 instances and 65 features;

year 4 9792 instances and 65 features;

year 5 5910 instances and 65 features.

Of all these features, however, only a few have proven to be crucial for the application of Machine Learning models. In particular, below there are the most important variables divided into qualitative and numerical.

Qualitative:

Credit history: it indicates whether the debtor has repaid all its debts in the past and currently and highlights also the case of late payments.

Purpose: which is the reason for which the loan has been requested. Highlight whether it was requested for the purchase of a new or used car; to buy furniture/equipment; for education.

[83] *"UCI Repository is a collection of databases, domain theories and data generators used by the machine learning community for the empirical analysis of machine learning algorithms". Source:* UCI Machine Learning Repository: About

Personal status and sex: it indicates whether the subjects are single, married, divorced or separated and if they are male or female.
Other debtors/guarantors: it highlights if the debtor has other debtors or guarantors.
Property: indicates whether the debtors has other collateral, i.e. if it owns houses, offices and cars.
Other instalment plans: this variable analyses whether the debtor has other instalment plans with banks or stores.
Housing: it highlights whether the debtor owns the house he lives in or is renting it.
Job: it indicates if the debtor is unemployed, employed, self-employed.
Foreign worker: it highlights if the debtor works in a foreign country.

Numerical:
Status of existing checking account: it indicates how much money the debtor has in existing checking accounts.
Duration in month: how many months ago the debt was incurred.
Credit amount: it indicates the amount of the debt.
Present employment since (in years): how many years they have been employed in the present job.
Age in years: how old the debtor is.
Number of existing credits at this bank.
Number of people being liable to provide maintenance for.

After having listed all the variables taken into account in order to develop this Phyton programme, there will be the indication of all the features which were not used.
Instalment rate in percentage of disposable income: this variable was not included because there were missing values relating to too many debtors, thus making its use unhelpful.
Present residence since: was considered to be of little relevance for the calculation of accuracy as the values contained therein can already be derived from other variables, such as “Housing”.
Telephone: this variable indicates whether the subject has no telephone at all or whether he has one or more telephones registered in his name. It might have been considered an interesting variable in the early 2000s, when not everyone owned a telephone, but

nowadays it no longer proves to be a useful characteristic for the purposes of the analysis.

Using these datasets, the Machine Leaning models developed to calculate the accuracy of predictions for performing and non-performing loans were Logistic Regression and Support Vector Machine.

To do this, it was created a model which, based on the values assumed by the above-mentioned variables, returns the percentage accuracy of the dataset using both models. In particular, the Phyton programme not only analyses values for the 7 dataset mentioned before, but it can be used for any other dataset. In fact, the Phyton programme has been automated in order to proceed with the calculation of the accuracy. For this reason, it is sufficient to enter the data for the new dataset on an Excel file called up in the programme in order to obtain new results.

In conclusion, the analytical logic of this chapter will be as follows: it will begin with the initial data cleaning phase, in which it will be presented the explanation of the different steps followed in order to obtain the final datasets used in the code and it will then proceed with the presentation of the variables and steps implemented in the code.
After that, it will continue with the analysis of the two Machine Learning models used, Logistic Regression and Support Vector Machine.
The chapter will be concluded with a description and explanation of the results obtained.

3.2 Data cleaning and pre-processing

In order to be able to make the comparison between the two Machine Learning models used, it is essential to clean up the datasets of any values that might lead the programme to return inaccurate values.
The first step developed in order to carry out data cleaning was to eliminate redundant variables and those that did not contain relevant information. In particular, after a careful study of which variables were most important to use, as already indicated in the

section 'Purpose of the Thesis', the following features were eliminated: telephone, instalment rate in percentage of disposable income, present residence since.

On the other hand, with regard to the missing values for some of the variables analysed, it was used a computer tool called Weka.
Weka is a programme that allows you to replace missing values within datasets using different approaches, but in this particular case the programme was used replacing the missing data with a constant. The decision stemmed from the fact that performing several tests by running the programme, the greater final accuracy came from using the datasets in which the missing values were replaced with the constant.

Once the dataset was cleaned of all redundancies and the missing values were replaced with the constant, in order to reduce overfitting, the datasets were ready to be inserted into the Phyton programme and to be processed.

At this point, before commenting on the results derived from the implementation of the code, it is considered necessary to describe the steps involved in obtaining the Phyton programme that calculates the accuracy of the datasets.

The first step is to choose the Phyton libraries to be used in order to achieve the desired result. In this case, the Sklearn library was mainly used to carry out the analyses and calculations.

At this point, since the aim of this thesis is to create a programme that could analyse not only the 7 datasets entered and explained in the next paragraphs, but also any other dataset of credits divided into performing and non-performing, a number of steps had to be taken to enable the Phyton code to perform this operation.

First of all, a maximum number of datasets that the code can use was defined, the value chosen is 1000. This choice stems from wanting to limit the maximum execution time of the code.

At this point, when the programme has been trained to read the datasets, one can proceed with the division of these into a training part and a testing part. It is important to highlight that, the choice of the percentage of training and testing was derived from several tests carried out with different training and testing percentages, the combination used turned out to be the one that allowed the most accurate results to be achieved. In particular, the best combination is 70% of the data used for the training part and the remaining 30% for the testing part. In particular, in order to split the dataset, it was used the function of scikit learn “train_test_split”, as illustrated in the figure below.

```
    X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.3,
random_state=0)
```

In particular, the 'Randomised Search' method was used for the selection of training and testing samples.

When the choice of parameters is so vast that it is not easy to choose which ones you want to test, a good option is to use the ‘Randomized Search’ that chooses the combinations of parameters randomly. In this case then you do not set the combinations to be tested but only the number that has to be tried, the algorithm will then choose, from the possible combinations that have been specified, those to be tested. For this reason, it is not possible to define which data has been used precisely in the model for the training part of the programme and which for the testing part. In fact, the programme was automated in such a way that it provided this function itself. We can specify, however, by means of an example, that since the training sample was set equal to 70%, while the testing sample was set equal to 30%, on a dataset consisting of 10000 instances, all feature values of 7000 creditors are used to train the programme, while the remaining 3000 are used at the end to test the already trained model.

After this step, it was performed the standardisation step, which consists of regularising and adapting the training data to contain the intensity of the weights, eliminate noise in the data and limit overfitting. To enable the programme to carry out standardisation, it was used of the function “StandardScaler” of scikit learn.

Finally, after these steps and before proceeding with the illustration of the methods used, it must be illustrated that, in order to reduce the programme execution time while avoiding a decrease in the accuracy of the results, was implemented the feature reduction method called Principal Component Analysis. 84

In order to find the correct percentage reduction in the number of features, several tests were carried out on the code. The most satisfactory result, in terms of accuracy, was the one explaining 80% of the variance in the data.

3.3 Machine Learning Methods

The objective of the next two paragraphs is to present the Machine Learning models used in the code to test the accuracy of datasets concerning performing and non-performing loans. The two methods used are Logistic Regression and Support Vector Machine, as at the moment they are the most commonly used by banks and consultancy firms due to the simplicity of application as well as a good level of accuracy in the results.

3.3.1 Logistic Regression

The Logistic Regression model is a statistical technique for determining the probability of occurrence of an event given certain risk factors.

In order to implement it, it is necessary to identify a dependent variable, consisting of the outcome (in our case performing or non-performing credit), and one or more independent variables, the so-called exogenous factors or regressors.

In order to use the Logistic Regression model to calculate the accuracy of our dataset, it was used the function of the Scikit Library called 'LogisticRegression'.

[84] "Principal Component Analysis, better known as PCA, is one of the most widely used algorithms for reducing the size of a dataset. Before going into the mathematical details of how the algorithm works, an attempt is made to explain the idea.
Having a dataset composed of m attributes creating a hyperspace of D dimensions, PCA 'projects' these m attributes into another system (or hyperplane) that is constructed in such a way as to maximise the variance contained in the original system with fewer dimensions. The first dimension (first principal component) of the new system will be constructed as a linear combination of the attributes with the greatest variance in the original system, the second dimension (second principal component 63 and orthogonal to the first) will be constructed trying to capture as much residual variance as possible, so on with the third and further dimensions."

Using the PCA as feature reduction method there are different results in terms of accuracy respect to the original one.
Below there will be a presentation of the different values obtained before and after PCA for the datasets in which Logistic Regression gave the best results in terms of percentage of accuracy.

The first dataset to analyse is the Polish dataset relative to 2008, which consists of 7027 instances and 65 features. In this case, before the use of PCA the percentage of TRUE POSITIVE rate is equal to 99.5% for class 0, while for class 1 is equal to 8.9%. The overall value accuracy was equal to 96.0%.
The interesting result is that after the application of PCA the overall accuracy is equal to 95.5%, so the value by which it decreased is to be considered almost negligible; although the number of features considered before and after the use of PCA decreased by a considerable amount, the programme went from considering all 63 features of the initial dataset to considering only 12. These 12 variables were chosen by means of the 'train' and 'test' function of PCA between the qualitative and quantitative variables indicated in the section 'Purpose of the Thesis' at the beginning of this chapter. This result certainly indicates the usefulness of feature reduction techniques, because although it greatly reduces the execution time of the programme, the end result remains more or less the same.
In particular, in the case analysed now, the reduction in code execution time was only just over a minute, but in the case of large datasets, its use can save a lot of extra time.

A summary picture on the performances of the model before the application of the PCA. In particular the metrics relative to TRUE POSITIVE[85], FALSE POSITIVE[86], precision[87] and recall [88] are presented.

	TP Rate	FP Rate	Precision	Recall
	0,995	0,911	0,965	0,995
	0,089	0,005	0,436	0,089
Weighted Avg.	0,960	0,876	0,944	0,960

Picture 1: *Metrics related to Polish 2008 dataset*

It is really important to pay attention to the FALSE POSITIVE rate express in the picture above. The FALSE POSITIVE is also defined type I errors and it represents a loss of profit, so it is very important to take this into account and calculate it accurately.

After analysing the accuracy rates and any errors, the final values of performing and non-performing loans were considered and a more accurate analysis of some of the variables retained most important in determining the results after the PCA implementation was carried out.

---

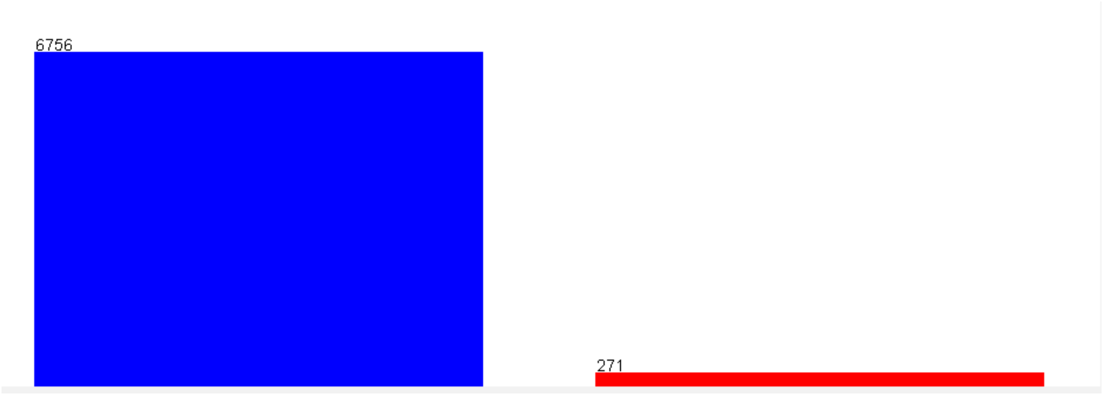
[85] True Positive Rate= $\frac{True\ Positive}{True\ Positive+False\ Negative}$

[86] False Positive Rate= $\frac{\text{Faòse Positive}}{False\ Positive+True\ Negative}$

[87] Precision= $\frac{True\ Positive}{True\ Positive+False\ Positive}$

[88]Recall= $\frac{True\ Positive}{True\ Positive+False\ Negative}$

With regard to the Polish dataset from 2008, out of a total of 7027 debtors, 6756 were considered to be performing and only 271 non-performing.



Picture 2: *number of performing and non-performing debtors Polish 2008 dataset*

Among the characteristics considered both before and after using Principal component analysis there are 'Savings account/bonds' and 'Credit amount'. The graphical presentation of these two characteristics, in the picture below, derives from the fact that they can be easily understood by everyone and make it clear what work was done by the programme in analysing the various attributes.

On the x-axis is expressed the variable ‘Savings account and bonds’, which we see as ranging from 0 to 70000 dollars and indicates the amount of savings account and bonds that the debtor has. On the y-axis, on the other hand, are presented the values for the variable 'Credit amount', this indicates how many years the debtor has been in the same job.

From the legend we can then deduce that the blue squares indicate the subjects considered performing, while the orange triangles indicate the subjects considered non-performing.

In fact, we can see that, in general, those with more savings accounts and bonds are less likely to be non-performing. Indeed, we can see that the number of orange squares is greater for borrowers with savings accounts and bonds between $10,000 and $20,000, and the number of blue triangles is greater for those with larger amounts of this variable. The latter are in fact more likely to repay even larger amounts of credits, as we can see from the last three blue squares at the top, which indicate borrowers with credit balances of $5,000 but savings accounts/bonds of $50,000 or $60,000, who are considered to be performing customers.

Picture 3: *graph representing on the x-axis 'Savings account/bonds' and on the y-axis 'Credit amount' related to Polish 2008 dataset*

Before continuing with the discussion of the Support Vector Machine, the data for the second dataset for which the results derived from the use of the Logistic Regression method proved to be more accurate are presented. Clearly, having already analysed the theoretical significance of the values illustrated for the previous dataset, we shall only proceed with a presentation of the results and graphs.

The dataset we want to analyse now is the Polish dataset for 2009. The choice of the latter stems from the fact that, through the use of PCA, the number of features fell from 63 to 13, being second only to the previously analysed dataset in the number of features removed.

The Polish dataset relating to 2009 consists of 10173 instances and 65 features. In this case, before the use of PCA the percentage of TRUE POSITIVE rate is equal to 99.6% for class 0, while for class 1 is equal to 10%. The overall value accuracy was equal to 95.8%.

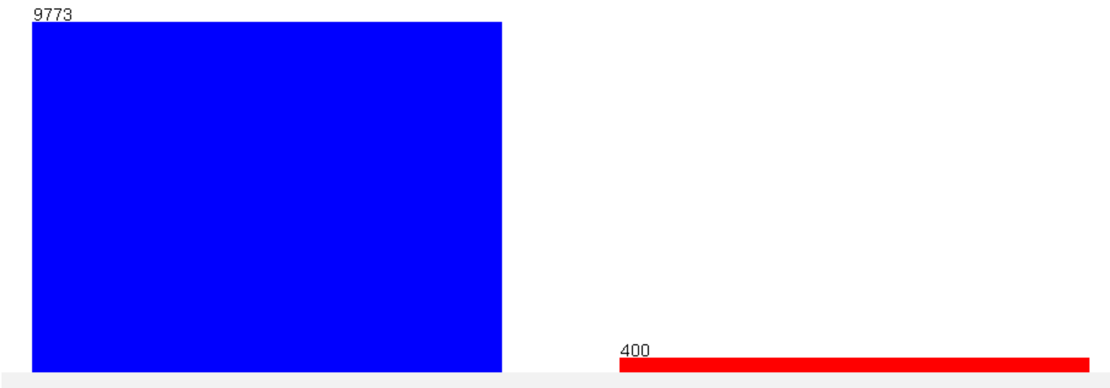
The interesting result is that after the application of PCA the overall accuracy is equal to 96.0%, so, also in this case, the value by which it decreased is to be considered almost negligible.

A summary picture on the performances of the model before the application of the PCA. In particular are presented the metrics relative to TRUE POSITIVE, FALSE POSITIVE, precision and recall.

	TP Rate	FP Rate	Precision	Recall
	0,996	0,990	0,961	0,996
	0,010	0,004	0,103	0,010
Weighted Avg.	0,958	0,951	0,927	0,958

Picture 4: *Metrics related to Polish 2009 dataset*

In this case, of the initial 10127 debtors, 9773 were considered to be performing and only 400 non-performing.



Picture 5: *number of performing and non-performing debtors Polish 2009 dataset*

In this graph, the two variables 'Savings account/bonds' and 'credit amount' are presented, as for the dataset illustrated above.

It is interesting to note, compared to the previous dataset, that here the correlation between higher value in terms of savings accounts/bonds and higher number of performing debtors is no longer so evident. We can see, however, that the final result in terms of performing and non-performing borrowers is in percentage terms similar to that of the previous dataset. This result in fact stems from the fact that, in this case, other variables proved to be more important in terms of defining performing and non-performing debtors. However, I have decided to present this graph and make the comparison with the previous one precisely to show how fundamental the use of Machine Learning models is, as by analysing together many more variables than one could otherwise do, and in a very short time, they make it possible to obtain very accurate results taking into account a huge number of factors.

Picture 6: *graph representing on the x-axis 'Savings account/bonds' and on the y-axis 'Credit amount' related to Polish 2009 dataset*

3.3.2 Support Vector Machine

The support vector machine is a model consisting of a set of examples for training, each labelled with a class from among two possible classes, and a training algorithm that builds a model that assigns new examples to one of the two classes, thus obtaining a non-probabilistic binary linear classifier.

In order to use the Support Vector Machine model to calculate the accuracy of our dataset, the function of the Scikit Library called 'svm.SVC' was used.

Using the PCA as feature reduction method there are different results in terms of accuracy respect to the original one.
Below there will be a presentation of the different values obtained before and after PCA. Although, even with the SVM method, the two datasets that present the most accurate results are those already presented in the Logistic Regression analysis (Polish datasets 2008 and 2009), we will proceed with the illustration of two other datasets in order to present other results obtained and, in particular, demonstrate how the code developed proved to be very accurate in all cases in calculating the final values.

The first dataset to analyse is the Polish dataset relative to 2010, which consists of 10503 instances and 65 features. In this case, before the use of PCA the percentage of TRUE POSITIVE rate is equal to 99.7% for class 0, while for class 1 is equal to 0.6%. The overall value accuracy was equal to 95.0%.
The interesting result is that after the application of PCA the overall accuracy is equal to 95.7%, so the value by which it decreased is to be considered almost negligible; although the number of features considered before and after the use of PCA decreased by a considerable amount, the programme went from considering all 63 features of the initial dataset to considering only 14.
Interestingly, the accuracy with the Logistic Regression method is 95.68%. For all datasets, the Support Vector Machine method was slightly more accurate. This fact could be attributed to the constitution of the dataset itself, which, as stated above, provides for a division that turns out to be quite clear between the variables used, with the final result being a value of 0 or 1, from which the model can go back and include the other variables.

A summary picture on the performances of the model before the application of the PCA. In particular the metrics relative to TRUE POSITIVE, FALSE POSITIVE, precision and recall are presented.

	TP Rate	FP Rate	Precision	Recall
	0,997	0,994	0,953	0,997
	0,006	0,003	0,091	0,006
Weighted Avg.	0,950	0,947	0,912	0,950

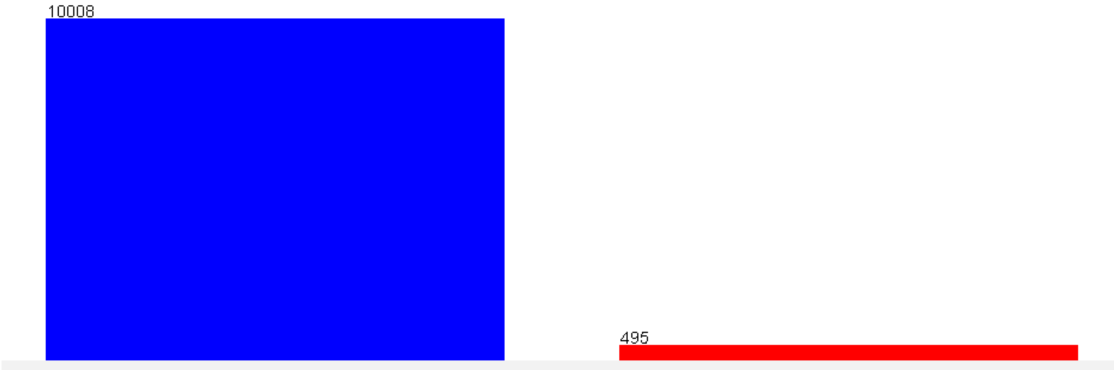
Picture 7: *Metrics related to Polish 2010 dataset*

A particularly interesting factor to be analysed in the case of this dataset is the value of the TRUE POSITIVE and FALSE POSITIVE rate compared to that of the other two datasets previously analysed.

In fact, TRUE POSITIVE RATE and FALSE POSITIVE RATE for class 1 are very close to 0 per cent, whereas in the two previous datasets they were closer to 10 per cent. This large variation is due to the fact that, in this dataset, there were a larger number of missing values for credits that were then placed within class 1. These, as specified in the introduction to this chapter, were replaced with a constant, leading to a decrease in accuracy. This variation, however, was considered negligible since the final average value of TRUE POSITIVE and TRUE NEGATIVE was greater than 90%.

After analysing the accuracy rates and any errors, the final values of performing and non-performing loans were considered and a more accurate analysis of some of the variables retained most important in determining the results after the PCA implementation was carried out.

With regard to the Polish dataset from 2010, out of a total of 10503 debtors, 10008 were considered to be performing and only 495 non-performing.



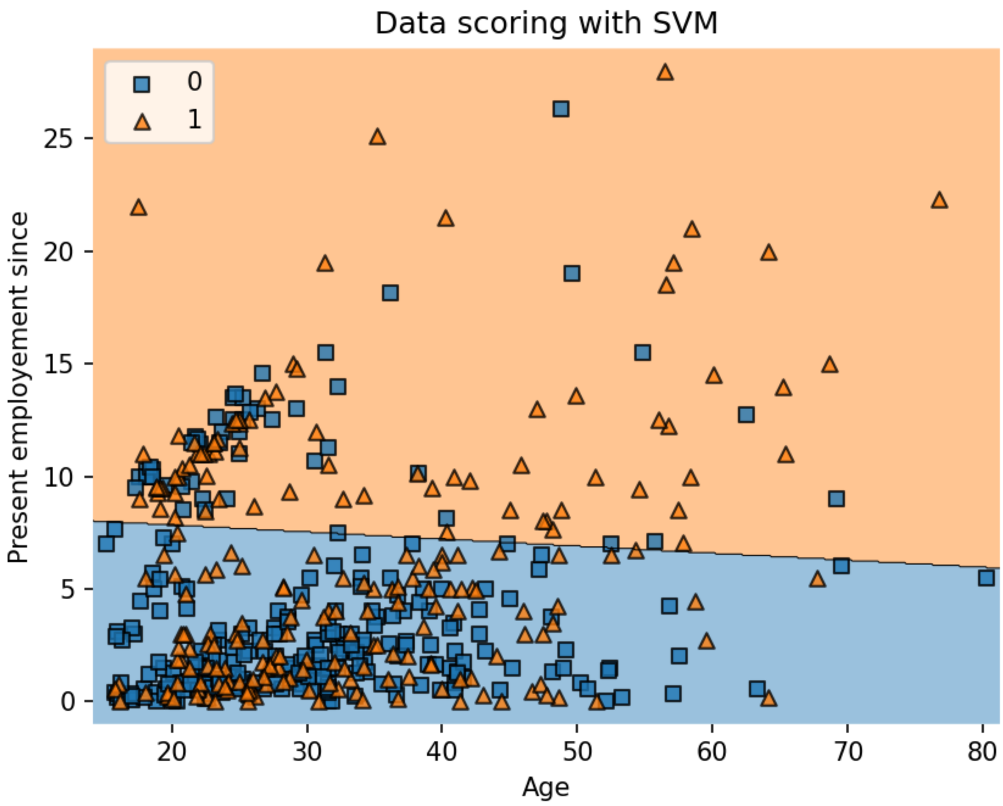
Picture 8: *number of performing and non-performing debtors Polish 2010 dataset*

Among the characteristics considered both before and after using Principal component analysis there are 'Present employment since' and 'Age'. The graphical presentation of these two characteristics derives from the fact that they can be easily understood by everyone and make it clear what work was done by the programme in analysing the various attributes.

On the x-axis is expressed the variable age, which we see as ranging from 0 to 80 years. On the y-axis, on the other hand, are presented the values for the variable 'Present employment since', this indicates how many years the debtor has been in the same job. This is the typical graph of the Support Vector Machine model, it is divided into two areas underneath, the lower part shows the area in which there is the presence of the greatest number of non-performing subjects, while the orange part at the top shows the area in which there is the greatest number of subjects considered to be performing.

From the legend we can then deduce that the blue squares indicate the subjects considered non-performing, while the orange triangles indicate the subjects considered performing.

Ultimately, we can see that subjects with a lower age and fewer years in the same job are more likely to not repay their debts.



Picture 9: *SVM graph representing on the x-axis 'Age' and on the y-axis 'Present employment since' related to Polish 2010 dataset*

Now, there will be the illustration of the second dataset with the method of Support Vector Machine. Clearly, having already analysed the theoretical significance of the values illustrated for the previous dataset, we shall only proceed with a presentation of the results and graphs.

The dataset we want to analyse now is the Polish dataset for 2011. The choice of the latter stems from the fact that, through the use of PCA, the number of features fell from 63 to 20.

The Polish dataset relating to 2011 consists of 9792 instances and 65 features. In this case, before the use of PCA the percentage of TRUE POSITIVE rate is equal to 99.6% for class 0, while for class 1 is equal to 2.3%. The overall value accuracy was equal to 94.5%.

The interesting result is that after the application of PCA the overall accuracy is equal to 94.6%, so, also in this case, the value by which it decreased is to be considered almost negligible.

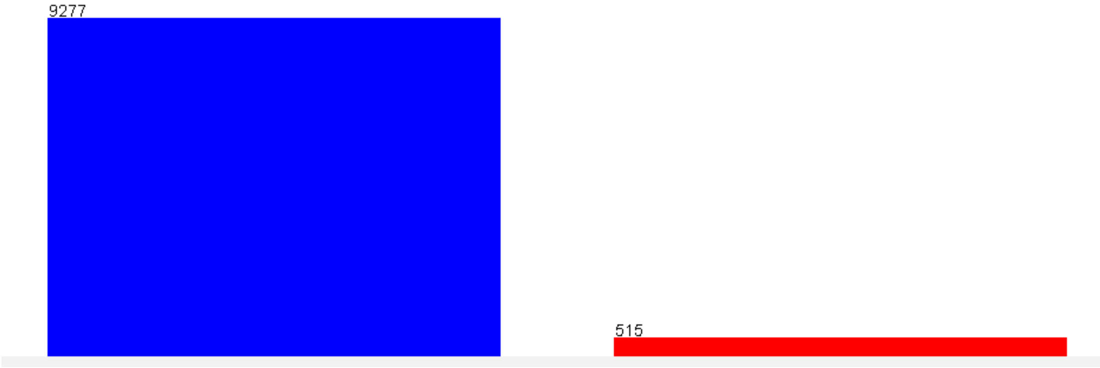
Interestingly, the accuracy with the Logistic Regression method is 94.14%. For all datasets, the Support Vector Machine method was slightly more accurate for the reasons already explained for the previous dataset.

A summary picture on the performances of the model before the application of the PCA. In particular are presented the metrics relative to TRUE POSITIVE, FALSE POSITIVE, precision and recall.

	TP Rate	FP Rate	Precision	Recall
	0,996	0,977	0,948	0,996
	0,023	0,004	0,255	0,023
Weighted Avg.	0,945	0,926	0,912	0,945

Picture 10: *Metrics related to Polish 2011 dataset*

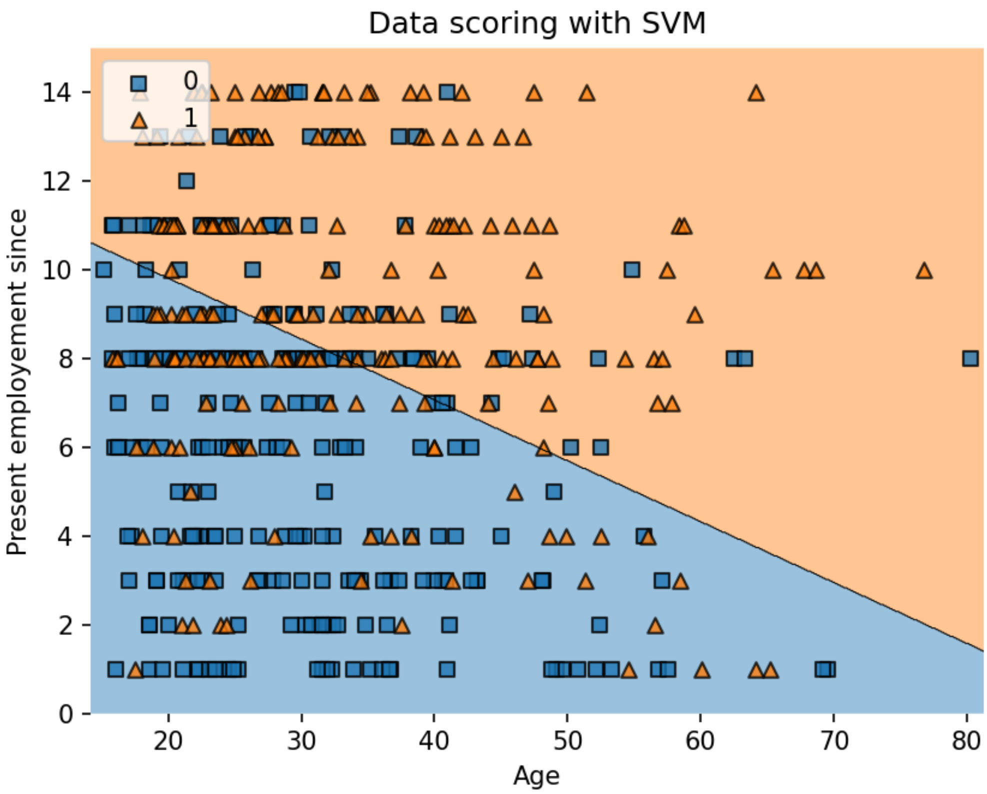
In this case, of the initial 9792 debtors, 9277 were considered to be performing and only 515 non-performing.



Picture 11: *number of performing and non-performing debtors Polish 2011 dataset*

In this graph, the two variables 'Present employement since' and 'age' are presented, as for the dataset illustrated above. It can be seen, however, that the values on the y-axis have changed; in fact, we no longer have a range from 0 to 30 years, but from 0 to 14; this change results from the different composition of the dataset.

However, as also highlighted for the previous dataset subjects with a lower age and fewer years in the same job are more likely to not repay their debts.



Picture 12: *SVM graph representing on the x-axis 'Age' and on the y-axis 'Present employment since' related to Polish 2011 dataset*

3.4 Final results

This section of the chapter briefly presents the final results for all 7 datasets used and compares the different accuracies obtained with the two methods considered (Logistic Regression and Support Vector Machine), before and after using PCA. This choice derives from the fact that, in the previous paragraphs, we preferred to concentrate on explaining the methods used and analysing only those datasets for which the results proved to be more accurate.

3.4.1 Comparison of machine learning models

The results obtained from the comparison between the Logistic Regression model and Support Vector Machine reflect the problem already highlighted in many academic papers, it is not possible to find one model that is more accurate than another in absolute terms. The model that turns out to be the most accurate for one dataset turns out to be less accurate than another Machine Learning Model for a different dataset, for this reason it is always very important to use at least two different methods in order to compare results.

In particular, in the 7 dataset used in this Phyton programme, the accuracy with the Logistic Regression model is higher than with the Support Vector Machine model in both the German and Australian datasets. In fact, with regard to the German dataset, the Logistic Regression model has an accuracy of 78%, whereas the Support Vector Machine model has an accuracy of 76%. With regard to the Australian dataset, the Logistic Regression model is 84% accurate, while the Support Vector Machine model is 83% accurate. The lower accuracy of the Support Vector Machine model for these Australian and German datasets stems from the fact that they have smaller datasets than the Polish one, both in terms of instances and in terms of features, and this leads the model to make more imprecise classifications between performing and non-performing loans than if the amount of data were larger. On the other hand, the Logistic Regression model succeeds in being more accurate as, having fewer characteristics to analyse, it has fewer regressors and thus the model succeeds in dividing between performing and non-performing more clearly.

On the other hand, with regard to the Polish data, not only it can be noticed that the Support Vector Machine model is more accurate than the Logistic Regression model for the reasons just stated, but we can also analyse the fact that as the number of data and features is greater, compared to the German and Australian datasets, the final accuracy is a much higher percentage. In fact, for all five years, the Polish datasets show an accuracy rate of over 90% with both methods. This factor allows us to emphasise, as stated in all academic writings, that Machine Learning models improve their predictions as the number of data presented increases, being able to count on a greater number of simulations not only in the training part but also in the testing part.

3.4.2 Comparison of machine learning models after PCA

After comparing the two machine learning models for the 7 datasets used, it was applied a feature reduction method using only the most relevant ones to see if and how much it would improve the percentage accuracy of my results. The feature reduction method adopted was Principal Component Analysis. PCA is the process of computing the principal components and using only them ignoring the rest.
In particular, after implementing several tests involving different numbers of principal components it was chosen those that explained 80 % of the variance in the data, as it was the most accurate.

Once PCA was applied, the new results are almost the same in accuracy percentages for all 7 datasets. It is very interesting to note that despite a substantial decrease in the number of features analysed, the programme results with an accuracy that is really similar to the previous ones, as it was explained more clearly in the paragraphs (‘Logistic Regression’ and ‘Support Vector Machine’) regarding the description of the 2 Machine Learning Models used. This shows how the use of feature reduction methods can be useful in allowing the programme to report very accurate results while taking into account far fewer features, thus decreasing the programme's execution time.

It is also interesting to note how much the number of features actually decreased with the use of PCA. In particular, for the German dataset the number of features fell from 23 to 14, for the Australian one from 13 to 9. The most interesting results, however, certainly concern the feature reduction for the five years of Polish datasets: for the first year the number of features analysed went from 63 to 12; for the second from 63 to 13; for the third from 63 to 14; for the fourth as well as the fifth from 63 to 20.

In conclusion, it is essential to use feature reduction methods because while they don’t cause a huge change in the percentage of accuracy, they permit to decrease execution time, one of the biggest problems faced by both consulting firms and banks in analysing performing and non-performing loans.

CONCLUSIONS

In conclusion, thanks to the results derived by the use of the Logistic Regression and Support Vector Machine Learning models, it is clear that the percentage of accuracy of the dataset of performing and non-performing loans depends fundamentally on the amount of data it consists of, usually more data corresponds to higher accuracy, but also and above all on the variables used. There are certain characteristics, such as, for instance, the age of the borrower or the amount and value of the assets that could be used as collateral, which prove to be crucial in allowing for highly accurate results.

When comparing the two models used, it can be seen that, in general, the results derived from using the Logistic Regression model turn out to be more accurate than those derived from using the Support Vector Machine method.

Another factor that is crucial to analyse concerns feature reduction methods. It has been shown how the use of these methods, due to the fact that some variables prove to be more fundamental than others, makes it possible to maintain an excellent accuracy of the dataset by reducing its size and thus allowing the algorithm to produce results more quickly and precisely. The feature reduction method used in the Phyton program described in this thesis is Principal Component Analysis (PCA).

In general, however, it must be concluded that both Machine Learning models used show that the accuracy of the analysed datasets is very high. This result allows banks and consultancy firms dealing with the recovery of non-performing loans to produce estimates that turn out to be very accurate on recoveries, allowing them to try to reduce the number of non-performing loans on the market with as little waste of time as possible. This factor proves to be fundamental, especially in the historical period we are currently facing, given the enormous financial crisis brought about by the Covid-19 pandemic, which has caused terrible uncertainty in the financial markets, leading to a drastic increase in non-performing loans.

Finally, as analysed repeatedly not only in this thesis but also in many academic papers in recent years, it is crucial to use machine learning methods to speed up the recovery process of non-performing loans, enabling the market to manage the huge amount of NPLs.

APPENDIX

Literature review

In advanced economies, the deterioration of a bank's credit quality is one of the main causes that can lead to financial crises or problems in the banking system. In recent years, an increase in the economic difficulties of borrowers has led to significant losses for banks. Indeed, the recent global financial crisis, and the subsequent recession in many developed countries, has led families and businesses to being unable to meet the costs of repaying their debts. Several studies and researches, carried out in developed countries, have shown a considerable link between impaired loans and economic growth. A slowdown in economic growth is likely to lead to an increase in the unemployment rate, as well as a reduction in available income and debtors will find it difficult to repay their debts. The stock of such debts should therefore be minimised so that there is no impediment to economic growth. In general, it has also been shown that certain macroeconomic variables affect credit risk. In this chapter we will learn about and analyse the main variables that influence the growth of non-performing loans, dividing them into country-specific (macroeconomic) variables and bank-specific (microeconomic) variables. Among the main macroeconomic variables that impact NPLs, we find the evolution of GDP (Gross Domestic Product), unemployment rate, inflation rate, interest rate on loans, exchange rate and public debt. Some scholars and researchers in their work have introduced the stock market index among the macroeconomic variables. Obviously, the influence of macroeconomic variables on NPLs changes depending on the type of loan to which the credit refers. For example, it seems that NPLs related to housing loans are less responsive to changes in country-specific variables than other types of loans. Over the last two decades, the determinants of non-performing loans have attracted the interest of many economists.

Analysing NPLs is very significant; among the main studies and researches concerning this topic we find:

1) Keeton and Morris (1987), their work was one of the first studies in this area, analysing a sample of about 2500 commercial banks in the United States, over a period from 1979 to 1985, to try to understand the main determinants of loan losses. They used simple linear regressions to demonstrate their expectations empirically and found that

a large proportion of loan losses reflected adverse local economic conditions. Specifically, these losses reflected the low performance of particular sectors such as agriculture and energy.

2) Gambera (2000), his analysis uses bivariate VAR systems and impulse response functions to study how economic development affects the credit quality of US banks. The VAR method allows all variables to be determined endogenously and it has the advantage of fully capturing the interactions between banks and macroeconomic variables. Gambera points out that a limited number of coutry-specific variables such as farm income, GDP and unemployment rate are good predictors of bank credit quality.

3) Bikker e Hu (2002), they provide an interesting perspective in assessing the procyclicality of banks' impaired loans on a sample of 26 countries belonging to the OECD (Organisation for Economic Co-operation and Development) between 1979 and 1999. The study finds out that GDP growth and inflation have a negative correlation when related to impaired loans, while the unemployment rate has a positive correlation. Thus, the results indicate that, although loan impairments decrease in good times, banks tend to reserve more in good times, that is to say, when profits are higher, and consequently banks are less procyclical than it would seem when looking only at their dependence on the business cycle.

4) Salas e Saurina (2002), they analysed the relationship between Non-Performing Loans and the economic cycle in Spain in the period from 1985 to 1997. Their analysis shows that, during the economic boom, banks tend to expand their lending activities and increase their market share. This is often achieved by lending to people with low creditworthiness. In fact, the study shows that non-performing loans increase during recession phases of the economic cycle and macroeconomic shocks are rapidly transmitted to banks' balance sheets.

5) Mario Quagliariello (2006), in his study "Bank's riskiness over the business cycle: A panel analysis on Italian intermediaries", he analyses the effects of the business cycle on borrowers' default rates relating the Italian banking system from 1985 to 2002. In particular, through the estimation of static and dynamic models, it tries to understand whether loan adjustments and Non Performing Loans have a cyclical trend. The estimates of these relationships can then be used to perform stress tests to assess the effects of macroeconomic shocks on banks' balance sheets. Indeed, after the peak of the

cyclical recovery, customers' profitability worsens and the creditworthiness of the borrowers deteriorates, increasing the stock of impaired loans and causing large losses on banks' balance sheets. This may result from a fall in asset prices which, in turn, further affects the financial wealth of borrowers and reduces the value of collateral pledged. In addition, the possible increase in unemployment reduces families' disposable income and consequently also their ability to repay debt. Banks' risk exposure increases, leading to higher loan adjustments and greater demand for capital at a time when its cost is higher or even unavailable. Intermediaries may react by reducing lending, but worsening the effects of the economic crisis. Quagliariello's report analyses the behaviour of about 200 Italian banks and the econometric results confirm that loan adjustments and Non Performing Loans are influenced by the evolution of the economic cycle.

6) Bofondi e Ropele (2011), in their analysis, they investigate the macroeconomic determinants that influenced the quality of loans in Italy over the period 1990 to 2010, using simple linear regressions. The two authors analyse the variables that can affect the quality of loans granted to families and those that can affect the quality of loans granted to firms separately, as these variables may behave differently. According to their estimated model, the quality of loans granted to families and firms can be explained by a limited number of macroeconomic variables mainly related to the general state of the economy, the cost of money and the debt weight. Moreover, the change in macroeconomic conditions generally affects credit quality with a certain lag. To measure loan quality, the two authors used the relationship between new non-performing loans and the amount of performing loans (NBL ratio). Using the results of their estimations, they quantitatively assessed how changes in macroeconomic determinants affect NBL ratios. According to the model that analysed the quality of loans granted to families, an increase in the unemployment rate of 100 basis points, all things being equal, will lead to an increase in the NBL ratio of 4 basis points, while an increase in Gross Domestic Product (GDP) of 1 percentage point will decrease the NBL ratio of 6 basis points after 4 trimesters. A positive change in the 3-month Euribor rate of one percentage point will increase the NBL ratio of about 12 basis points after three quarters, indicating that the cost of debt servicing is a significant determinant of Non-Performing Loans developments. On the other hand, a one percentage point positive change in the annual growth rate of the house price index will reduce the NBL ratio of

approximately 2.5 basis points after two trimesters. As regards the NBL ratio for business loans, an increase in the unemployment rate of 1 percentage point would be reflected in an increase in the NBL ratio of 27 basis points, while an increase in the annual growth rate of durable goods consumption of 1 percentage point would lower the NBL ratio of only 2 points after 3 quarters. In the specific case of corporations, general business cycle conditions and debt burden are both key to the evolution of New Bad Loans (NBL). In the conclusion of their analysis, the NBL ratio on household loans varies inversely when compared to the real growth rate of Gross Domestic Product (GDP) and house prices. It varies directly with the unemployment rate and the nominal short-term interest rate. Regarding loans to businesses, the NBL ratio increases with the unemployment rate and the ratio of interest expense to EBITDA, while it decreases as the consumption of durable goods increases. However, in contrast to their results concerning families, the NBL ratio of firms shows evidence of endogenous persistence. Indeed, the macroeconomic determinants described above influence the evolution of the NBL ratio with different time lags.

7) Messai e Jouini (2013), in their analysis "Micro and Macro Determinants of Non-Performing Loans", they tried to identify the determinants of non-performing loans in a sample of 85 banks in three different countries that suffered the most from impaired loans after the 2008 financial crisis, that is to say Italy, Spain and Greece. The reference period of their study starts in 2004 and goes up to 2008. Macroeconomic variables used by the two authors include the growth rate of Gross Domestic Product (GDP), the unemployment rate and the real interest rate. Through the application of panel data, their results show that Non-Performing Loans are negatively correlated with the growth rate of the Gross Domestic Product (GDP), the profitability of assets and positively correlated with the unemployment rate. Impaired loans should be important during economic recession periods. Commercial banks should also extend the range of macroeconomic surveillance to include prudential indicators such as GDP in order to assess the strength and stability of the banking system. However, like any other research, the analysis has some limitations; in fact, other macroeconomic variables such as the real exchange rate or inflation could have been used. In order to enrich the study, they could expand the sample of banks to European banks and use a combination of "macro stress testing" to measure the impact of macroeconomic shocks on non-performing loans and thus assess the resilience of banks in dealing with these shocks.

8) Beck et al. (2013), they analyse a very large panel of 75 countries between 2000 and 2010. Their contribution focuses not only on the classical determinants of Non-Performing Loans described above, but also on the possible existence of currency misalignments and the stock market performance of a given country. For countries with specific vulnerabilities, exchange rate depreciation may also lead to an increase in impaired loans, especially if countries hold a high share of foreign currency liabilities. When analysing equity prices, however, a negative correlation has emerged, especially in the case of advanced economies with more extensive financial markets.

9) Jakubíc e Reininger (2013), analysis similar to that of Beck et al. but focusing on a more limited time frame and sample of countries, 9 countries in the CESEE region from 2004 to 2012. Another distinguishing feature of their study is the use of additional variables such as the amount of real exports, real domestic demand, the VIX index and the global emerging market bond index (EMBIG). The results of the analysis confirm that GDP is a very influential variable on the amount of Non Performing Loans.

10) L. Abid, M. N. Ouertani e S. Zouari-Ghorbel (2013), in their study, "Macroeconomic and Bank-specific Determinants of Household's Non-performing Loans in Tunisia: A Dynamic Panel Data", through the use of a dynamic panel data method they attempted to examine the Macroeconomic and bank-specific determinants of Non-performing Loans referring to loans provided to families. The study considers a sample of 16 banks in Tunisia for a period ranging from 2003 to 2012. The volume of Non-Performing Loans, which relate to loans granted to families in Tunisia, has increased steadily in recent years. According to the authors, non-performing loans are the result of the growth in consumer credit. In fact, although the largest share of non-performing loans, about 80%, belongs to the productive sector during the period 2005-2012, unpaid consumer credit has increased more significantly, about 21%, than commercial credits, about 6%. The results of their analysis show how the real growth rate of Gross Domestic Product (GDP), the inflation rate and the real lending rate have an influence on impaired loans.

11) Nir Klein (2013), the author's paper analyses Non-Performing Loans in Central, Eastern and South-Eastern Europe (CESEE) in the period between 1998 and 2011. The observed results show that the rapid increase in impaired loans over the period and in the area considered not only increased banks' vulnerability to further shocks but it also

limited their lending operations with a broader impact on economic activity. This paper assesses these retrospective effects and identifies the main determinants of NPLs over time and across sixteen CESEE countries using a variety of panel estimation techniques. Among the main Macroeconomic variables, the results described in the paper show that Non-Performing Loans increase when the unemployment rate rises, when the exchange rate depreciates and when the inflation rate is very high. The examination of feedback effects between the banking system and economic activity largely confirms the strong macro-financial linkages in the CESEE region. While Non-Performing Loans appear to respond to macroeconomic conditions, the author also shows how an increase in impaired loans has a significant impact on credit, GDP growth, unemployment and inflation in subsequent periods, validating the idea that healthy and sustainable growth cannot be achieved without a strong and resilient banking system. In conclusion, Nir Klein's paper also finds some policy implications. Given the adverse effect of Non-Performing Loans on the economic picture, there are indeed grounds to strengthen supervision in order to avoid a severe increase in impaired loans in the future, also by ensuring that banks avoid excessive lending, maintaining high credit standards and limiting foreign currency lending to unhedged borrowers.

12) Makri, Tsaganos e Bellas (2014), in their analysis, "Determinants of Non-Performing Loans: the case of Eurozone", using an econometric model, they tried to identify the determinants of Non-Performing Loans in the Eurozone, focusing on the period before the financial crisis that hit European countries. Using panel data aggregated from a sample of 14 Eurozone countries, for a period ranging from 2000 to 2008 and applying different GMM estimates, they found a strong correlation between non-performing loans and several macroeconomic factors. It is worth mentioning that this is the first empirical study that explores the possible macroeconomic and banking determinants affecting NPLs ratios, using data at the aggregate Eurozone level. From a macroeconomic perspective, the results of the analysis conducted by the authors indicate that public debt, GDP growth rate and unemployment rate significantly influence the Non Performing Loans of European countries. This shows that the economic state, and in particular the economic cycle, is clearly linked to the quality of the loan portfolio. Specifically, there is a strong positive correlation between credit quality and unemployment, revealing that a lack of jobs weakens the borrower's ability to pay loan instalments. The GDP also seems to have a significant influence on impaired

loans in the Eurozone, revealing that, during periods of strong expansion, credit quality improves and vice versa.

13) Tanaskovic e Jandrić (2015), their study is one of the most recent in the literature and aims to analyse the macroeconomic and institutional determinants of Non Performing Loans growth. The research focuses on the CEEC and EEA countries (Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Hungary, Lithuania, Montenegro, Macedonia, Romania, Serbia and Slovenia) over the period from 2006 to 2013. Non-performing loans in Central Europe and the Baltics started to increase after the economic crisis had broke out in 2008. The growth rate of NPLs ratios was much higher than in the European Union as a whole. In addition, increases in impaired loan stock continue to be a significant pressure in many CESEE countries. The rising trend in Non-Performing Loans began after the outbreak of the financial crisis in 2008, but the deterioration in loan quality is irregular across the different countries. In fact, according to the analysis of the two authors, the average ratio of Non-Performing Loans for the sample analysed in 2006 were around 3.83%, rising to 18% in 2013. The highest growth of impaired loans since the beginning of the crisis was recorded by Lithuania, while the other countries recorded a steady growth, apart from Greece which recorded NPLs ratios around 30% after 2010. Through the use of statistical panel data, the results show a negative correlation between the increase in GDP and the increase in Non Performing Loans. Along with GDP, the ratio between foreign currency and loans and the level of the exchange rate are positively correlated with the increase in Non-Performing Loans. These results by the two authors confirm the expectation that countries where the domestic currency is not the main means of loan placement will have more problems in Non-Performing Loans, especially in periods of domestic currency depreciation. With regards to the inflation rate, in the model presented, it appears to be statistically insignificant for the countries in the analysed sample, despite the fact that a higher level of inflation could lead to a decrease in the debt value and therefore to an easier loan repayment. Institutional variables, on the other hand, include the quality of auditing, the development of financial markets and the validity of the banking system. In conclusion, the only variable that has a statistically more significant influence on the Non-Performing Loan trend seems to be the level of financial market development. In fact, a development of the market should lead to a reduction of Non-Performing Loans as there is the possibility to use additional sources of financing.

In conclusion, we could argue that banks' impaired loans (NPLs) are loans whose collection is considered at risk under different profiles. They are generally exposures of credit institutions to subjects who, due to a worsening of their own economic and financial situation, are not able to meet their obligations and, therefore, to repay their debts on time or in the amounts expected.

The Bank of Italy divides NPLs into three main categories that account for the process of gradual deterioration that a loan may undergo: impaired past due and/or exposures in arrears, probable defaults and non-performing loans.

Impaired past due and/or exposures in arrears are loans that are past due or in excess of credit limits for more than 90 days and above a certain relevance limit.

Probable defaults are the next step, where the bank considers that a debtor is unlikely to meet his contractual obligations in full, unless the bank takes action such as the enforcement of guarantees.

Finally, there are non-performing loans, i.e. loans to entities in a state of insolvency or in substantially similar situations.

The cases may therefore be very different and range from an only temporary inability of the debtor to meet his obligations to situations of evident insolvency. During the process of loan deterioration, the bank may also consider it more convenient to restructure the exposure. In such a case, the terms of the contract are usually modified, for example by rescheduling the debt or partially cancelling it.

It should also be noted that banks that experience credit deterioration act in support of their position with depreciations and provisions, thus reducing the value of the credit in the balance sheet (to bring it closer to that considered to be recoverable) or setting up their own resources to cover a growing risk exposure. Hence the distinction of non-performing loans into gross impaired loans and net impaired loans (i.e. excluding depreciations and provisions already made in the balance sheet).

The coverage level of an impaired exposure on a bank's balance sheet is called the coverage ratio (or simply coverage) and it is expressed as a percentage of the nominal exposure.

For example, if a bank has an (impaired) loan of €100 million, but for safety's sake it has set aside €60 million in its balance sheet in fear of the risks associated with recovery, the nominal value of the impaired loan is said to be €100 million and the coverage ratio is said to be 60 per cent. Clearly, the average coverage level of a probable default tends to be lower than that of a non-performing loan.

The economic and financial crisis that hit Italy between 2008 and 2014 also had a heavy impact on bank balance sheets, as the growing difficulties of families and businesses were reflected in their reduced ability to repay their debts. This led to an increase in impaired loans on bank balance sheets and to greater fragility of the banks themselves. For years, high levels of NPLs have been considered one of the major weaknesses of the Italian financial system and have prompted many banks to undertake ordinary and extraordinary consolidation operations. Leading Italian banks have securitised large portfolios of impaired loans, including ad hoc public guarantee schemes known as GACS, and simultaneously launched massive capital increases. National and international regulators, as well as market operators, have repeatedly requested a "clean-up" of bank balance sheets with the dual aim of strengthening the credit system and encouraging lending to the real economy.

The securitisation of an impaired loan is the main instrument of intervention on the financial items of a credit institution at risk. It consists of an originating bank, known as the originator, selling its portfolio of impaired loans to a special purpose vehicle (SPV) company. The SPV in turn issues securities, both rated and unrated, which are placed with professional investors and repaid with the amounts recovered from the loans. Clearly, impaired loans are generally sold at much less than their nominal value. Currently, 30% or more of the nominal value is considered acceptable or positive, but in the recent past credits have been sold at 18-20% of the nominal value. It is crucial for the bank to achieve the highest possible share of the credit's nominal value at the time of sale.
For example, if a bank has an NPL portfolio of 100 million with a coverage of 60%, and the loan is sold at 25%, the bank will immediately record a net loss of 15% of the

nominal amount sold (100%-65%-25%). For large portfolios, it has often been necessary for Italian institutions to make significant capital increases in order to cope with the losses from securitised loans.
In turn, securitised loans issued by an SPV can be divided into tranches of decreasing "quality" among senior, mezzanine and junior. On the highest quality securities (and therefore safer and with an investment grade credit rating) the Italian State has provided for possible public guarantees known as GACS and designated to facilitate their placement.

At the beginning of 2018, Banca Ifis estimated that in 2017 there were as many as EUR 72 billion of transactions related to impaired loans and that a further EUR 57 billion were expected for the following year.

In its July 2018 Monthly Report, the Italian Banking Association (ABI) calculated that net non-performing loans (i.e. net of depreciations and provisions) in May 2018 stood at €49.3 billion, with a decrease of €37.5 billion (-43%) on the December 2016 level (€86.8 billion). The highest level of net non-performing loans had been reached in November 2015 at 88.8 billion.

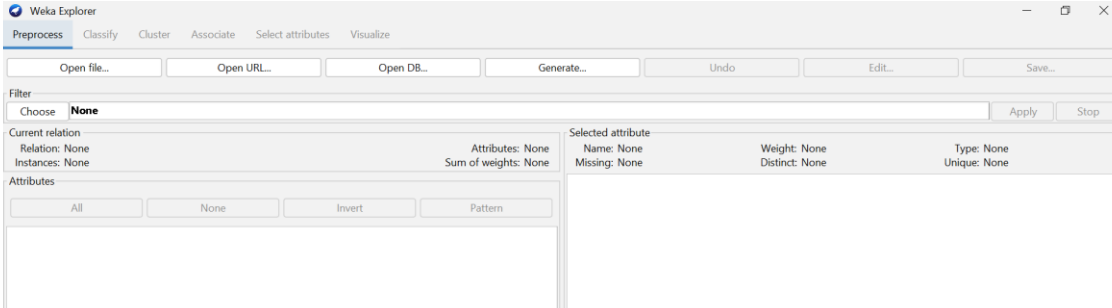
The results achieved, thanks to economic recovery and bank consolidation measures, had also brought the ratio of net non-performing loans to total loans (i.e. the weight of net non-performing loans on total bank loans) to 2.84% from 4.89% at the end of 2016.

The Bank of Italy's monthly bulletin Banks and Money of June 2018 (published at the beginning of the subsequent August) certifies that, at the end of the first half of the year, total gross non-performing loans of the Italian banking system amounted to approximately €131.7 billion and showed a decline of 26.1% year-on-year (-10% in May) thanks to several large securitisation transactions (for approximately €32 billion).

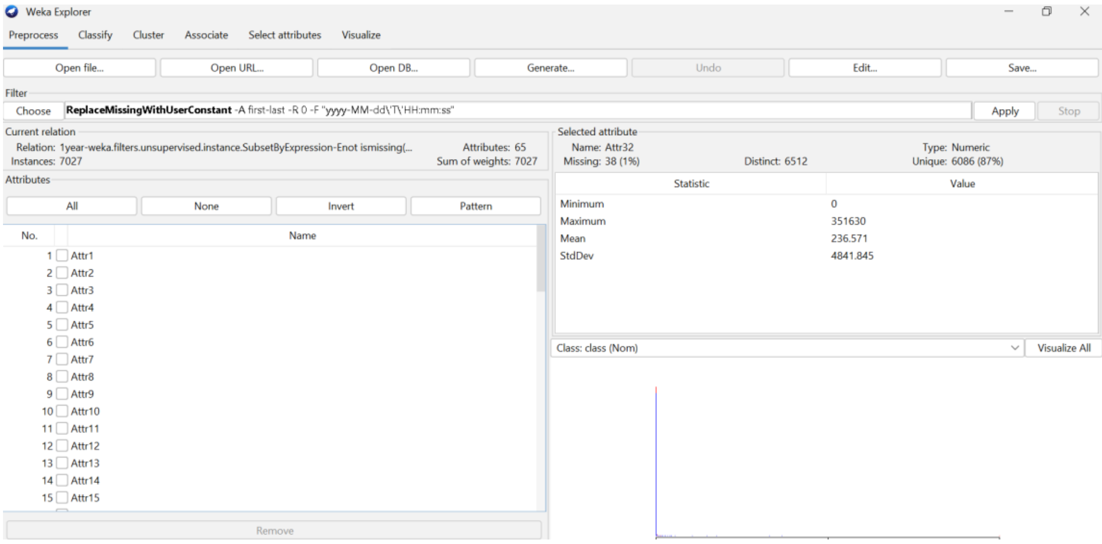
Guidelines for the use of WEKA for the code developed in this thesis

To get the final dataset and use it in the code, I had to make some adjustments, in fact I had to deal with a series of missing data. To overcome this problem I downloaded and used the latest version of WEKA program, version 3.8.6. Once the program has been downloaded, you have to open WEKA and click on Explorer.

Then click on OPEN FILE selecting the dataset that need to be edited from the Dekstop.



After that, click on CHOOSE>FILTER>UNSUPERVISED>ATTRIBUTE>REPLACE MISSING WITH USING CONSTANT

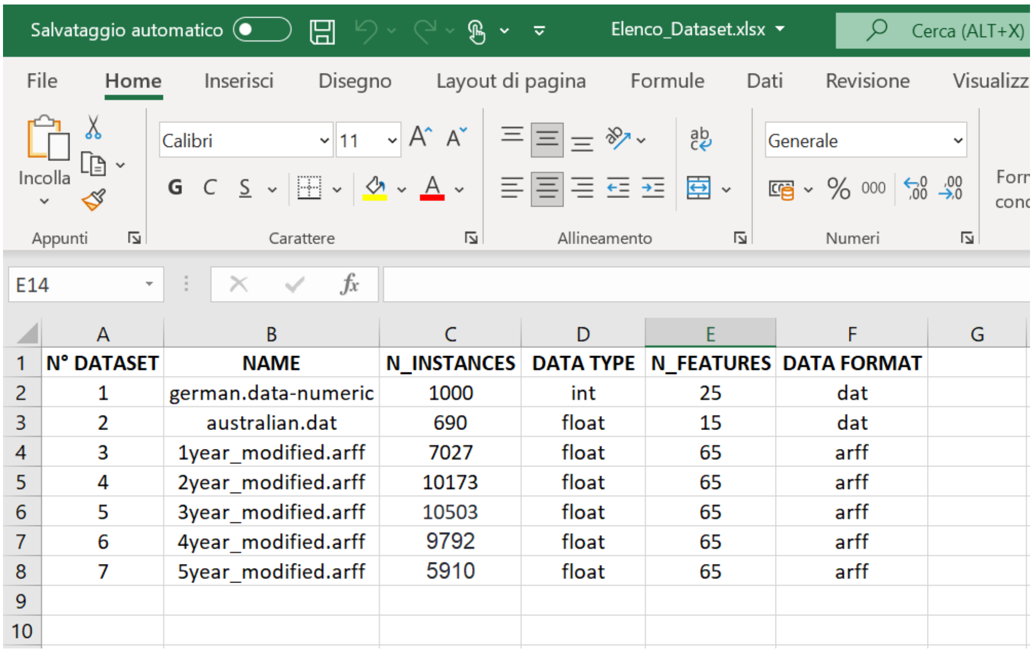


Now you can click on APPLY and finally SAVE.

In this way, the program has replaced all the missing attributes with the constant 0 and it has allowed me to create the final dataset.

The decision to replace the missing data with the constant derives from the fact that there were few missing values and most of them belonged to attributes whose main value was exactly this constant value.

Picture of the File Excel developed and used in the Phyton programme



	A	B	C	D	E	F	G
1	N° DATASET	NAME	N_INSTANCES	DATA TYPE	N_FEATURES	DATA FORMAT	
2	1	german.data-numeric	1000	int	25	dat	
3	2	australian.dat	690	float	15	dat	
4	3	1year_modified.arff	7027	float	65	arff	
5	4	2year_modified.arff	10173	float	65	arff	
6	5	3year_modified.arff	10503	float	65	arff	
7	6	4year_modified.arff	9792	float	65	arff	
8	7	5year_modified.arff	5910	float	65	arff	
9							
10							

This file contains all the datasets that are then analysed by Phyton. The Phyton programme can give the final results of 1000 different dataset, because of the file Excel above is called up by the ‘path’ command. For this reason, a few lines of code have been inserted to tell the programme how the Excel file is made up, indicating the various columns and the format of the data. In fact, the data formats for which the code has been trained are 'arff' and 'dat'. In particular, to enable Phyton to read the composition of the Excel file, a ‘for cycle’ was developed, using the function 'dataset.append'.

Code developed with PHYTON

```
import numpy as np

from sklearn import datasets

from sklearn.metrics import accuracy_score

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import StandardScaler

from sklearn.linear_model import LogisticRegression
```

```
from mlxtend.plotting import plot_decision_regions
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from scipy.io import arff
import pandas as pd
import openpyxl as xl
import warnings as w
from sklearn import svm



max_dataset = 1000
path = r"C:\Users\39349\Desktop"
w.filterwarnings('ignore')

def load_data(dataset):
    n_dataset = dataset[0]
    name = dataset[1]
    n_instances = dataset[2]
    data_type = dataset[3]
    n_columns = dataset[4]
    data_format = dataset[5]
    #print(dataset)

    if n_dataset == None:
```

```
        return False,[],0

    if data_format == "dat":
        data = np.fromfile(file=path+"\\"+name, sep=" ", dtype=data_type).reshape(n_instances, n_columns)
        return True,data, n_columns
    if data_format == "arff":
        data, meta = arff.loadarff(path+"\\"+name)
        data = np.asarray(data.tolist(), dtype=np.float64)
        return True,data, n_columns



file_excel = xl.load_workbook(path+"\Elenco_Dataset.xlsx")
sheet = file_excel["Foglio1"]

for i in range(2, max_dataset):

    name = (sheet["B"+str(i)].value)

    dataset = []
    dataset.append(sheet["A"+str(i)].value)
    dataset.append(sheet["B"+str(i)].value)
    dataset.append(sheet["C"+str(i)].value)
    dataset.append(sheet["D"+str(i)].value)
```

```
    dataset.append(sheet["E"+str(i)].value)
    dataset.append(sheet["F"+str(i)].value)

    state, data, n_columns = load_data(dataset)
    if not state:
        break

    y = data[:, [n_columns-1]]
    X = data[:, range(n_columns-2)]

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=0)

    # STANDARDIZZAZIONE
    #Nota. La regolarizzazione è l'adattamento dei dati di training per contenere
l'intensità dei pesi, eliminare
    #il rumore nei dati e limitare il fenomeno dell'overfitting.
    sc = StandardScaler()
    sc.fit(X_train)
    X_train_std = sc.transform(X_train)
    sc.fit(X_test)
    X_test_std = sc.transform(X_test)

    # ANALISI PCA
    #Riduzione del numero di feature evidenziando quelle più rilevanti
```

```
    pca = PCA(0.8)
    pca.fit(X_train_std)

    X_train_std = pca.transform(X_train_std)
    X_test_std = pca.transform(X_test_std)

    #lr = LogisticRegression(C=1.0, tol=1e-4, class_weight=None, solver="newton-cg", random_state=0)
    #lr = LogisticRegression(C=1.0, tol=1e-4, class_weight=None, solver="saga", random_state=0, multi_class="auto", warm_start=True)
    lr = LogisticRegression(C=1.0, tol=1e-1, class_weight=None, solver="lbfgs", random_state=0, multi_class="auto", warm_start=True)
    clf = svm.SVC(C=0.5, kernel='linear');
    #clf = svm.NuSVC()#gamma="auto")

    lr.fit(X_train_std, y_train)
    clf.fit(X_train_std, y_train)

    correct = 0
    for j in range(0,len(X_test_std)):
        z = np.array([X_test_std[j]])
        y_predict = lr.predict(z)
        if (y_test[j] == y_predict):
            correct+=1
```

```
print("Database: " + name)
print("Accuracy with Logistic Regression: ", lr.score(X_test_std, y_test))
print("Accuracy with SVM: ",clf.score(X_test_std, y_test))
print("Feature number before PCA: " + str(pca.n_features_))
print("Feature number after PCA: " + str(pca.n_components_)+"\n")
#print("Accuracy: " + str(correct/len(X_test_std))+"\n")




# visualizzazione grafica della classificazione
#features = range(3,4)
#features = [1, 2]
#lr.fit(X_train[:, features], y_train)
#clf.fit(X_train[:, features], y_train)
#print(len(y_test))
#y_print = np.reshape(y_train.astype(np.int), len(y_train));

#plot_decision_regions(X_train[:, features], y_print, clf=clf, legend=2)
#plt.title('Data scoring with SVM')
#plt.xlabel('Age')
#plt.ylabel('Present employement since')
#plt.legend(loc='upper left')
#plt.show()
```

quit()

Some comments on the development of the CODE

Logistic Regression

In order to develop the Logistic Regression model, a series of tests were carried out on the values to be attributed to the variables that make up this function, analysing those that allow the model to return the greatest accuracy. The variables of the Phyton function and their characteristics are shown below.

The first parameter to consider was the value of C to use; it represents the invers of the regularization strengths, smaller values specify stronger regularization. This value was imposed equal to 1.0 in all the tests. The second parameter to take into consideration was ‘tol’, it indicated the tolerance for stopping criteria. In the first phase of testing it was imposed equal to 1e-4, but then 1e-1 turned out to be the best value. Then, regarding the variable ‘class_weight’, it was imposed equal ‘None’: it is the default value and in this case all classes are supposed to have weight 1. The fourth variable to take into account is ‘solver’: it represents the algorithm to use in the optimization problem. In order to choose the best one you have to consider if your dataset is large or small and if you are considering a multiclass problems. In this case, it was used the ‘lbfgs’ solver. After that, it was imposed the default value both to the ‘random_state’ and ‘multi_class’ variable, so it is useful only to comment the last one ‘warm_state’, it was set to ‘True’ and in this case it reuse the solution of the previous call to fit the initialization, otherwise, just erase the previous solution.

Table: Potential combinations tested and best parameters.

Algorithm	Potential combination tested	Best combination
Logistic Regression	1. lr = LogisticRegression(C=1.0, tol=1e-4, class_weight=None, solver="newton-cg", random_state=0) 2. lr = LogisticRegression(C=1.0, tol=1e-4, class_weight=None, solver="saga", random_state=0, multi_class="auto", warm_start=True)	lr = LogisticRegression(C=1.0, tol=1e-1, class_weight=None, solver="lbfgs", random_state=0, multi_class="auto", warm_start=True)

Support Vector Machine

In order to develop the Support Vector Machine model, we proceeded with the analysis of the parameters to be used to enable the programme to return the highest accuracy.

The first parameter to consider was the value of C to use; it represents the invers of the regularization strengths, smaller values specify stronger regularization. This value was imposed equal to 0.5. The second parameter to take into consideration was kernel, which was imposed equal to 'linear'. Linear Kernel is used when the data is linearly separable using a single line. In this case, because of the dataset was already separated in performing and non-performing, the linear model was absolutely perfect, in fact the percentages of accuracy with this model are very high for all the datasets utilized.

Algorithm	**Potential combination tested**	**Best combination**
Support Vector Machine (SVM)	1. clf = svm.NuSVC()#gamma="auto") 2. clf = svm.SVC(C=0.5, kernel='linear')	clf = svm.SVC(C=0.5, kernel='linear')

Final part of the code

After the execution of the Logistic Regression and Support Vector Machine method, a 'for cycle' was inserted to analyse the results from the initial dataset. Basically, the inserted 'for cycle' allows the programme to recalculate, on the basis of the variables in the dataset, whether the debtor turns out to be a performer or not, evaluating whether at the end of the process its final results (i.e. the values 0 and 1) match those already in the dataset. In this case, the variable that proved to be crucial is 'predict', which placed within the 'for cycle' allows this to be achieved.

BIBLIOGRAPHY

BANCA D'ITALIA, I crediti deteriorati (Non-Performing Loans - NPLs) del sistema bancario italiano, 6 June 2017, https://www.bancaditalia.it/media/views/2017/npl/index.html

BANCA d'ITALIA, Note di stabilità finanziaria e vigilanza, April 2016

C. BARBAGALLO, «I crediti deteriorati delle banche italiane: problematiche e tendenze recenti» Rome, 6 June 2017

CESARONI T., Prociclicità dei sistemi di rating del credito: come gestirla, September, 2015. https://www.bancaditalia.it/pubblicazioni/temidiscussione/2015/2015-1034/index.html?com.dotmarketing.htmlpage.language=102

COSENTINO L., Debt Restructuring & NPLs Management, Luiss Master in Corporate Finance & Banking- Major in Banking, 3 March 2020

COSENTINO L., PREVITALI D., Modalità di gestione del credito deteriorato e sviluppo del mercato NPL in Italia, in Banche & Banchieri, 2016

COSENTINO L.., Debt Restructuring & NPLs Management, Luiss Master in Corporate Finance & Banking- Major in Banking, 3 marzo 2020

COTUGNO M., "*Gestione e Valutazione dei Non Performing Loans*", August 2018

D. CUCINELLI, «"The Impact of Non-performing Loans on Bank Lending Behavior: Evidence from the Italian Banking Sector",» Eurasian Journal of Business and Economics, 2015

EY, The Italian NPEs Market. From darkness to daylight, January 2019, https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/emeia-financial-services/ey-the-italian-npes-market.pdf

IFRS 9 par. B.5.5.17, Al sito: http://www.revisorionline.it/IAS_IFRS/ifrs9- p1.htm

JUSTIN A. SIRIGNANO, APAAR SADHWANI, KAY GIESECKE, "*Deep Learning for Mortgage Risk*", March 8, 2018

KAGAN J., Resolution Trust Corporation (RTC), Investopedia, 2019, https://www.investopedia.com/terms/r/resolution-trust-corporation.asp

KPMG, I Non Performing Loan in Italia, trend in atto e prospettive future, 2018, https://assets.kpmg/content/dam/kpmg/it/pdf/2018/07/I-Non-Performing-Loan-in-Italia-2018.pdf

P. ANGELINI, "I crediti deteriorati: mercato, regole e rafforzamento del sistema", Roma, 9 October 2018

PEZZUTO A., "*La garanzia dello Stato per le operazioni di cartolarizzazione di crediti in sofferenza*", in 'Diritto bancario, Diritto finanziario', 28 February 2020, https://www.tidona.com/lagaranzia-dello-stato-per-le-operazioni-di-cartolarizzazione-di-crediti-in-sofferenza/;

PROVENZANO A.R., D. TRIFIRO', A. DATTEO, L. GIADA, N. JEAN, A. RICIPUTI, G. LE PERA, M. SPADACCINO, L. MASSARON & C. NORDIO, "*Machine Learning approach for Credit Scoring*", August 5, 2020

Q. FRANCESCA, in Rischio di credito e valutazione della Loss Given Default, Roma, Bancaria Editrice, 2007

REFERENCE FOR BUSINESS, Resolution Trust Corporation (RTC), https://www.referenceforbusiness.com/encyclopedia/Res-Sec/Resolution-Trust-Corporation-RTC.html

SIDDHARTH BHATORE, LALIT MOHAN, RAGHU REDDY, "*Machine learning techniques for credit risk evaluation: a systematic literature review*", Institute for Development and Research in Banking Technology 2020

SPOTO D.M., TORINI D., KPMG, GACS Securitisation deals in the Italian NPL space, 27 april 2020, p.6, https://assets.kpmg/content/dam/kpmg/it/pdf/2020/04/GACS-Securitisation-deals-in-the-Italian-NPL-space.pdf

XOLANI DASTILE, TURGAY CELIK, MOSHE POTSANE, "*Statistical and machine learning models in credit scoring: A systematic literature survey*", Applied Soft Computing, June 2020

ZHANG YU, GUAN YONGSHENG, YU GANG & LU HAIXIA, "*Recognizing and Predicting the Non-Performing Loans of Commercial Banks*", International Journal of Signal Processing, Vol.9, 2016

UCI Repository, datasets related to Germany, Poland and Australia, https://github.com/JLZml/Credit-Scoring-Data-Sets