UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M. FANNO"

CORSO DI LAUREA IN ECONOMIA

PROVA FINALE

"Credit scoring models: Evolution from standard statistical methods to machine learning techniques"

RELATORE:

CH.MO PROF. WEBER GUGLIELMO

LAUREANDO: DONGHI FRANCESCO

MATRICOLA N. 2010152

ANNO ACCADEMICO 2022 – 2023

Dichiaro di aver preso visione del "Regolamento antiplagio" approvato dal Consiglio del Dipartimento di Scienze Economiche e Aziendali e, consapevole delle conseguenze derivanti da dichiarazioni mendaci, dichiaro che il presente lavoro non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Dichiaro inoltre che tutte le fonti utilizzate per la realizzazione del presente lavoro, inclusi i materiali digitali, sono state correttamente citate nel corpo del testo e nella sezione 'Riferimenti bibliografici'.

I hereby declare that I have read and understood the "Anti-plagiarism rules and regulations" approved by the Council of the Department of Economics and Management and f am aware of the consequences of making false statements. I declare that this piece of work has not been previously submitted — either fully or partially — for fulfilling the requirements of an academic degree, whether in Italy or abroad. Furthermore, I declare that the references used for this work including the digital materials have been appropriately cited and acknowledged in the text and in the section 'References' .

Firma (signature)

....

CONTENTS

ABSTRACT IN ITALIANO

Questo elaborato discute il modo con cui modelli di *credit scoring* si sono evoluti nel tempo a partire da tecniche statistiche più tradizionali fino a metodi avanzati di *machine learning* e quali miglioramenti sono stati introdotti dall'evoluzione di tali modelli. Più nel dettaglio, il lavoro tratta l'evoluzione sopracitata attraverso due differenti prospettive. In un primo caso, l'obbiettivo è quello di presentare in modo sintetico l'importanza ottenuta dai modelli di *credit scoring* nel corso del tempo e le loro principali aree di applicazione, con un *focus* particolare sul processo di creazione ed erogazione del credito. Questa presentazione si sviluppa attraverso una discussione delle pratiche che le banche, e in generale le istituzioni finanziarie, devono seguire per un corretto sviluppo e utilizzo del *credit scoring*, in modo da misurare il rischio di insolvenza associato ai vari debitori. Nel secondo caso l'elaborato analizza e presenta alcune tecniche statistiche utilizzate per il *credit scoring*, provando in questo modo ad evidenziare le principali caratteristiche di ogni modello e le principali differenze tra modelli standard e modelli avanzati. Questo dovrebbe servire per gettare le fondamenta per una comprensione dei progressi - in termini di performance, affidabilità e nuove aree di applicazione - che sono stati resi possibili dai modelli più sofisticati. Infine, per dare sostanza e concretezza alla discussione, questo elaborato esamina in dettaglio un modello di *deep learning* recentemente proposto dalla letteratura.

1. INTRODUCTION

This thesis investigates how credit scoring models have evolved over time from standard statistic techniques to advanced machine learning models and what advancements and challenges the evolution of these models has led to. More into details, it deals with the evolutions of these models with two different points of view. In one case, the work tries to briefly present the importance of credit scoring models over time and the main areas of application with the main focus on the lending origination process. This presentation is conducted by discussing the practice that banks, and in general financial institutions, have to follow for a proper usage of credit scoring to measure the default risk linked to borrowers. In the second case, the works analyses and discusses different statistical techniques used for credit scoring from a theoretical point of view, trying to highlight the main characteristics of each technique and the different features between standard and advanced models. This should lay the foundations to understand what improvements -in term of performances, reliability and new areas of applications- more sophisticated models have brought. Also limitations of these models are presented along with possible solutions. Finally, a deep learning model, recently proposed in the literature, is examined to give a practical example and substance to the discussion.

For this purpose, the research is divided into four different chapters. The first chapter is used to provide a brief introduction to the topic regarding the evolution and the spread of credit scoring over time. The second chapter is focused on the guidelines provided by regulators to ensure a proper development and usage of credit scoring for the creditworthiness assessment in the bank practise - lending to both retail customers and corporates is considered, however the main focus regards retail customers. The third chapter can be considered the most important one, since different techniques are presented and compared in order to identify and explain why machine learning methods show better performances than traditional models. In chapter number four, a presentation of some limitations related to credit scoring models is made and an analysis of the model proposed by Albanesi and Vamossy (2019) is carried out to show, from a practical point of view, how to solve possible limitations and what are the advantages brought by the application of machine learning and deep learning techniques to credit scoring.

Credit scoring refers to the set of statistical techniques used by lenders to measure the risk associated to a particular lending transaction. By feeding these models with specific customers' characteristics, lenders are able to measure their probability of default and thus, assess an estimate of the creditworthiness, the so-called score, that is useful to divide customers into homogeneous risk classes and compute the price of transactions. Credit scoring can be used for two main purposes: it can be used to decide whether to grant loans to potential borrowers or to decide how to manage relationship with existing borrowers, in the latter case we talk about behavioural scoring.

The usage of statistical techniques to distinguish between "good" and "bad" loans was first introduced by Durand in 1941. Since then, it is possible to identify different factors explaining the spread of credit scoring usage among financial institutions and consultancies over time. The increase in the level of competition and the number of applications for credit lines -i.e., due to the introduction of credit card in the '60s- required banks to adopt automatic lending decisions to speed up the assessment process. Moreover, banks needed to move from subjective lending decisions to objective ones in order to avoid discrimination among applicants based on factors such as race and gender. The adoption of automatic credit scoring models allowed banks to increase their ability to assess creditworthy customer with positive effects in term of reduction of default rates and losses.

Credit scoring models found an important application in 2007 after the implementation of Basel II Accords. Basel II requires banks to keep a minimum amount of regulatory capital to cover their credit exposures, with the final goal of ensure the banks' stability and the stability of all financial system. In this sense, banks are incentivised to adopt internal rating models for measuring the actual credit risk they face in order to determine the minimum amount of regulatory capital weighted for that risk. Thus, the role of scoring techniques assumed greater importance since a more accurate evaluation of the credit not merely discriminate between good and bad customer but increase the quality of the credit and ensure a higher stability and more confidence in the financial system.

Their importance has become even more evident during the high financial instability following the 2007-2008 crisis in which, the credit quality deteriorated significantly, the delinquencies rate increased and concerns about the actual ability to assess the risk linked to different financial instruments arose. In this context, even more attention has been paid to the evaluation and control of the quality of credit and to the assessment and management of credit risk, and this led regulatory authorities to introduce tighter regulations regarding capital adequacy and procedures in the lending process and banks to re-think their risk management principles and procedures. Again, the usage of credit scoring models capable of provide a sound analysis of default risk became almost fundamentals for banks to reduce losses and optimize profits. A practical example of the diffusion of credit scoring/ratings models in the crisis period can be searched in Del Prete et al. (2013) - occasional paper published by Bank of Italy. The diffusion and evolution of the usage of credit scoring was occurring even before the crisis in the Italian context and "the crisis seems to have speeded up the process". Regarding large banks, "the degree of diffusion of credit scoring [was] already almost complete for the major banks in 2006", while it increased in smaller banks in the period 2006-2009. The usage was higher for loans to households and SME than loans to large corporates, i.e., the diffusion in medium-large banks in 2009 was 97,1% 97,2% and 91,2% respectively. Furthermore, Italian banks' procedures regarding credit scoring changed during and after the crisis since "between 2006 and 2009, large banks moved the use of models from the granting phase to … pricing and monitoring", while the usage for small banks became more flexible - automatic evaluations combined with analyst evaluations- in all the phases of the lending process to face the uncertainty of that period.

Even the last couple of years have been characterized by uncertainty and economic and political instability and banks have reacted by tightening credit standards and terms and conditions on loans (ECB 2023), since their risk perceptions and concerns about consumer creditworthiness have increased and their risk tolerance has decreased. Also, in this context, it is possible to extrapolate the importance of credit scoring models in evaluating and managing the quality of credit, increasing the efficiency of the lending process and providing stability to the financial system especially during unstable period. Moreover, thanks to more complex, sophisticated and powerful models it is also possible to analyse determinants of defaults and get useful insights for design better regulations, management processes and prudential policies to reduce default rates.

2. LENDING ORIGINATION PROCESS

Even though the provision of financial services has become more popular among banks, the lending activity is the main activity carried out by banks. Lending activity involves banks to be exposed to various risks that can undermine their profitability and, in a worst scenario, its survival and thus, a proper management of the lending process is vital for their success.

The lending process can be divided into two phases: the origination and the managing of the credit. The first one starts with the request from a client, and it consists of gathering information about the client, using this information for assessing the associated creditworthiness through the usage of qualitative and quantitative analyses and making the final decision about the approval of the credit. The managing phase involves the monitoring of the credit exposure by checking for potential deterioration in the client position -reduction in the likelihood to receive the repayment and thus higher possibility to record a loss- and the managing of the relationship with the borrower to prevent or reduce potential losses. Only the origination phase will be covered in this chapter through a presentation of the development and usage of credit scoring models during the creditworthiness assessment process for retail customers and corporates. The topic is preceded by an overview of regulators guidelines regarding automated models used for the lending origination activity.

2.1 GUIDELINES FOR CREDIT SCORING MODELS: A GENERAL OVERVIEW

The credit-granting process and the conditions applied on loans by a bank depend on its measurement and management procedures related to the risk taken, in compliance with banking regulations and guidelines provided by different supervision authorities. Banking regulations and guidelines, considering in particular loan origination phase, ensure more sound and prudent standards for credit risk taking with an eye to increased consumer protection and higher stability and resilience for banks and the financial system -with positive effects for the real economy. Regulators set credit limits and rules within which banks have to develop inter alia their own credit risk culture that ensures the quality of the credit granted, their own risk appetite framework, an appropriate decision-making framework and robust credit risk procedures regarding, for instance, the collection of data for creditworthiness assessment and the approval of credit granting.

In this context, the usage of credit scoring models is subjected to a series of regulations and guidelines related to aspects such as data protection, model governance and fairness. These regulatory frameworks give banks a good amount of freedom in the development and usage of a proprietary model -in the financial system there is the tendency, especially for large banks and institutions, to develop proprietary models and keep them private- but also provide expectations for a correct, efficient and prudent design and usage.

With the main reference to the guidelines provided by the European Banking Authority, the model used must be understood by the user which is required to have necessary documentation regarding the development of the model, its functioning and the underlying assumptions -documentation should explain the theory, the assumptions and the statistical model used for the valuation. Moreover, policies and procedures that ensure the quality and adequacy of the inputs and model's usage are required as well as tests for the performance of the model prior to and during implementation in order to ensure the quality of outputs, the appropriate safeguards to provide confidentiality, integrity and availability of information and systems and the appropriate remediation measures in the case of detected issues -as stated by EBA (2020)-. Institutions should be able to understand the outputs of the models and reach a final lending decision that satisfies transparency and non-discrimination standards.

By providing these guidelines regulators, on the one hand, allow institutions to have models that suit their purpose -since some of them develop their own models in the way they prefer- and on the other hand, ensure a correct and appropriate development and usage of these models so that a good risk-based discrimination among customers can be done.

Important clarification: since there is not a specific international standardized framework, the main focus and reference of this paragraph was for European regulators -European Banking Authority and European Central Bank-, however the idea provided here regarding credit scoring usage can be considered as more general. Indeed, "an [effective] examiner's assessment of credit risk and credit risk management usually requires a thorough evaluation of the use and reliability of the models. … Regulatory reviews usually focus on the core components of the bank's governance practices by evaluating model oversight, examining model controls, and reviewing model validation" (FDIC 2007, p.1).

2.2 CREDITWORTHINESS ASSESSMENT

The main risk that arises from the lending activity is probably the credit risk, which is the risk of potential losses incurred by a lender due to the borrower inability or unwillingness to repay the debt -default risk- or due to the deterioration of the borrower creditworthiness that causes a reduction in the present value of the loan -migration risk. Through a creditworthiness assessment, it is possible to assess the potential borrower capability to repay the amount of debt requested by estimating the probability of default -the likelihood that the borrower will not repay the principal granted and the associated interest- as a measure of the credit risk. Credit scoring models are used for the estimation since they analyse consumer characteristics and provide a credit score as outcome that can be converted into the probability of default.

A good credit assessment is fundamental both from a lender perspective, by ensuring a higher quality of credit and a better pricing of the transaction that reflect the borrower risk -the calculation of the interest rate to be applied-, and from a consumers' perspective by protecting them from over-indebtedness and bankruptcy events and ensuring a fair transaction price that is linked with the individual risk profile. For a correct evaluation, institutions are required to collect reliable, accurate and up-to-date data and information regarding different aspects and factors that can influence the customer's ability to meet obligations. Data come from internal sources collected through the application form directly provided by the applicant or subsequent clarifications, and from third parties, i.e., credit information systems which provide reliable and up-to-date information about the credit history and the actual financial position of financial institutions customers.

The creditworthiness assessment process requires models that capture specific characteristics and information that describes the financial profile of the customer and the type of transaction. These characteristics and, more in general these models, differ depending on the type of customer involved in the transaction. For this reason, it is easier and appropriate to make a distinction between retail consumer, consumer lending, and enterprises, corporate lending -the use of credit scoring models concerns only small-medium enterprises and not large companies since the latter require a more detailed evaluation based quantitative analyses and qualitative considerations.

2.2.1 CUSTOMER LENDING

As already said, the process for assessing the customer creditworthiness suits the interest of the lender and it is also fundamental for the consumer protection, in fact retail consumer, without detailed information and specific knowledge of the subject, are the weaker party to be protected in the bank-borrower relationship. For this reason, lenders must assess and verify the borrower's source repayment capacity, compare the individual repayment ability and the personal actual financial position with the terms applied to the loan and consider possible factors or events that may undermine the future repayment capacity -thus preventing the customer from hardship and over-indebtedness.

Historically, the evaluation process was slow and inconsistent and carried out by managers whose final decision was subjective and based on personal experience and feelings. The development and introduction of credit scoring models has made the creditworthiness assessment faster, more reliable and systematic.

Automatic models are used extensively for consumer credit due to the lower exposure of each borrower that does not justify a deep and expensive qualitative analysis and due to the high number of applicants and transactions that, firstly, requires a greater speed in the assessment process and, secondly, provides enough empirical data to develop a sound and valid model. Banks may develop their own scoring model that are usually keep proprietary and private, and because of that there is a limited availability of specific information regarding the methods of the models' development and the variables used. Moreover, there are different ways to model and analyse same variables and data depending on the different type of credit -variables that explain the final outcome are weighted differently depending on lender's needs- and there are models with different levels of sophistication. For these reasons, it is possible to provide just a general presentation on how to build a credit scoring model. The idea behind credit scoring models is that past borrower data, thus past performances, are indicators for predicting future performance of similar borrowers.

The first step for developing a credit scoring model requires to define the population to which the model is applied, it is important that statistical units in the population share the same economic or financial characteristics. Then, it is crucial to select a representative sample -usually through random sampling to avoid systematic error- of the population that is used to build the model. The sample has to be made up of a sufficient number of past accounts or transactions that are identified as good or bad -bad accounts can be defined in different ways, a common definition is the delinquency of 90 days. Since the number of good accounts is likely to be much higher than the one of bad accounts, it can be necessary to carry out a sub-sampling in order to obtain a sample with similar numbers of good and bad units and so prevent imbalance in the dataset. The sample is also randomly divided into training sample, the subsample used for the development of the model, and the test sample, used for the model validation.

Once the sample is obtained, the choice of specific statistical techniques is required in order to design the model. For this purpose, the units that populate the training sample are analysed and explanatory variables -those that are considered to have a significant explanatory power in discriminating between good and bad accounts- are selected among units' observable features. The predictive -explanatory- variables can be selected using two different approaches: a priori identification of variables based on theoretical reasoning or an automatic selection of variables from a set of potential explanatory variables based on a selection criterion. In the latter case, each variable can be added, forward selection, or subtracted, backward selection, from the set of variables to obtain several models and check for the best one; it is also possible to use a hybrid stepwise procedure as a combination of the forward and backward selection.

At this point, the model assigns a specific point or weight to the attributes assumed by the selected variables and a table, the scorecard, with all the weights, can be built. In this way, a comparison between the attributes of a potential customer and the scorecard can be done and the final score can be obtained by summing up all the weights corresponding to the personal attributes. This procedure is automatic and by feeding the model with data, it provides the corresponding score -there are models that directly provide the probability of default instead of the score.

The selection of retail customers' data in the development and application of credit scoring models is a delicate process. Indeed, just some of the client information collected from different sources before the creditworthiness evaluation can be used to train credit scoring models. This limitation has as the objective to prevent models from conducting a discriminatory scoring that may exacerbate disparities in the credit access and inequalities; thus, potential factors of discrimination such as race, religion, gender and age are not used as well as the income that is not commonly used directly in the models -however information income-related is collected and used in the creditworthiness assessment along with the score-. The main factors that affect the score of a customer are related to the past credit and repayment performance, outstanding credit lines and variety and frequency of new products and inquiries. Moreover, the introduction of more sophisticated machine learning models allows the modelling and analysis of a greater amount of data coming from alternative sources i.e., utilities data, behavioural data and online transactions and digital data.

A public example of credit scoring model is the FICO score, developed by Fair Isaac Corporation in 1989 and probably the most used and known credit score. This score ranks borrowers in a scale from 300 to 850 by the probability of delinquency in the following 24 months. As a support of what written before, the key factors used by this model can be described as follow: “these are payment history and outstanding debt, which account for more than 60% of the variation in credit scores, followed by credit history, or the age of existing accounts, which explains 15-20% of the variation, followed by new accounts and types of credit used (10-5%)” (Albanesi, Vamossy 2019, p.68).

2.2.2 CORPORATE LENDING

The development and usage of credit scoring models for the evaluation of the riskiness associated with a potential borrower was also extended to small and medium enterprises. Indeed, the quite large volume of data regarding specific types of transactions and firms and the relative low exposure of a single transaction allow banks to exploit the speed and the cheapness of credit scoring models. However, the evaluation of the creditworthiness for SMEs requires a higher level of interpretation and judgement than the evaluation for retail customers: the repayment capacity is assessed considering the ability of the firm to generate present and sustainable future cash flows and income that is evaluated through a series of variables that require proper definition and handling. The most used variables are economic and financial ratios that come from a quantitative analysis of the financial statement of a firm and that are used to state the economic and financial position and stability of that firm - profitability, capital structure, liquidity, size and so on. Even though the analysis is quantitative, the calculation of these ratios, requires assumptions and choices regarding how to consider specific pieces of information in the scoring model -i.e., a yearly negative performance recorded in the income statement doesn’t necessary correspond to a poorly repayment capacity in the future since it might be the result of significative investments that can positively affect future performance and generation of cash flows.

Furthermore, other variables that can be considered in credit scoring models are related to credit and repayment history of the firm and the characteristics of the owner, since in SMEs it is a fundamental figure for the success of the business -sometimes the score for a small enterprise can be computed by just considering owner's features to reduce costs and needs for interpretation. By also focusing on finding new explanatory variables that explain the performance of firms, credit scoring literature found that "spatial risk factors as an indicator of local economy characteristics" (Onay and Öztürk 2018, p.390), daily transactions and behavioural data can be included in more modern models.

One example of credit scoring for SMEs is provided in the next chapter through the presentation of the Altman's Z-score.

3. CREDIT SCORING MODELS

3.1 OVERVIEW OF THE EVOLUTION OF CREDIT SCORING

Credit scoring can be seen as a classification problem in which the main focus is the prediction whether a customer will default in a specific period of time and the probability of that default -default forecasting. In this sense, there are different statistical techniques that are used to solve this type of classification problems. Historically, traditional statistical models such as linear discriminant analysis and logistic regression were the first and the most used techniques due to their simplicity and transparency. However, in the recent years, the vital role assumed by the usage of credit scoring along with the increase in computational power and a broader access to a large variety of data has triggered an evolution towards more sophisticated and complex machine learning and deep learning models, i.e. regression trees, ensemble methods like random forests, extreme gradient boosting and bagging, and neural networks. A lot of studies have developed complex models that are able to outperform and replace traditional linear models due to the ability to manage larger amount of data and recognized non-linear interactions among variables. Moreover unlike traditional models, machine learning algorithms can select automatically predictive variables that would otherwise be excluded since their relations with the default can be difficult to identify and interpret.

The sensibility and the prediction accuracy of these models has been tested and proven to be superior compared to standard techniques, especially the logistic regression that is the industry standards and the main benchmark -recent studied suggest the usage of decision trees (Dastile, Celik and Potsane 2020) and random forests (Lessman et. al 2015, see Gunnarsson et al. 2021) as better benchmarks-. In the literature, it is possible to find a good amount of research regarding the development of ensemble models that can be considered the best classifiers so far -especially extreme gradient boost techniques-, while the research about deep learning architectures is still limited but shows promising results depending on the features of the architecture and the dataset. Furthermore, an increasing number of papers has considered and studied hybrid models in the last decade since the combination of different statistical techniques may be able to obtain a better accuracy rate than the benchmark model and reduce the limitations of single models and algorithms.

Apart from the development and comparison of statistical techniques and the increase in the predictive accuracy, credit scoring literature has obtained important achievements in other research avenues. A significant number of studies has focused the attention on the identification of new explanatory variables from alternative data sources to boost and improve the creditworthiness assessment. For instance, the usage of machine learning models made possible the utilization of unstructured financial -granular transactional data- and non-financial data -digital footprints and social network data-, the so-called big data, for a better assessment process and for extending the credit granting to borrowers with a limited credit and financial history. In addition, profit scoring models has been suggested for estimating the expected profitability of a transaction rather than the probability of default in order to focus attention on the most profitable customers. Another topic that is becoming more relevant concerns the presence of distortions causing credit discrimination and unfairness, the relationship of alternative data sources on privacy violation and discrimination, and the development of "regulatory oversight for fairness and accuracy of [AI] scoring systems" (Onay and Öztürk 2018, p.391).

Despite the increase in accuracy –"improvement in default forecasts compared to traditional statistical models mostly ranging between 2 and 10 percentage points" (see Bonaccorsi di Patti et al. 2022, p.14)- and the introduction of new determinants, machine learning models present some flaws. In fact they are considered opaque, black boxes, because the complexity of their structure makes it difficult to explain and interpret the results and the relations between the features and final prediction and, because of that, the vital ability of financial institutions to justify outcomes and decisions is challenged. There are also some other issues related to sophisticated models -they will be discussed more into details in the next chapter- and financial institutions carefully consider the trade-off between accuracy and interpretability, and between benefits and limitations, in the decision regarding the adoption of machine learning models. Bonaccorsi di Patti et al. (2022) provide an example regarding the approach Italian financial intermediaries have towards these models: the usage of machine learning techniques, or more in general artificial intelligence -AI- methods, is spreading among financial intermediaries, mainly for the credit granting process, due to the higher optimism and confidence over the benefits entailed and the low perception of incremental risks compared to conventional models. A higher accuracy, efficiency and the possibility to use alternative sources of data have incentivized financial institutions to adopt AI models, in particular ensemble learning methods such as random forests and gradient boosting which guarantee a higher simplicity in the implementation and a better interpretability than deep learning models. Moreover, explainability techniques are used in order to ensure a greater level of transparency to the decision process without affecting the performance.

In the following part of the chapter, the several models mentioned above are presented through a synthetic discussion of the underlying theory and the logic.

3.2 LINEAR DISCRIMINANT ANALYSIS

Discriminant analysis was first introduced by Fisher in 1936. This analysis is a classification tool that divides a population into two -or more- classes, that differ in some features, and assigns a specific observation to the one class that is the closest in term of similarity between the features. This technique is useful for credit scoring -binary classification problem- since it allows to discriminate between solvent and insolvent borrowers through the estimation of a discriminant function as a combination of selected variables that best explain differences between the two groups -the function is such that the distance between the score means, centroids, of the two classes is maximized. The simplest approach is the linear discriminant analysis -LDA- in which the discriminant function is a linear combination of different variables.

Figure 1 provides a visual representation of the linear discriminant function and the logic

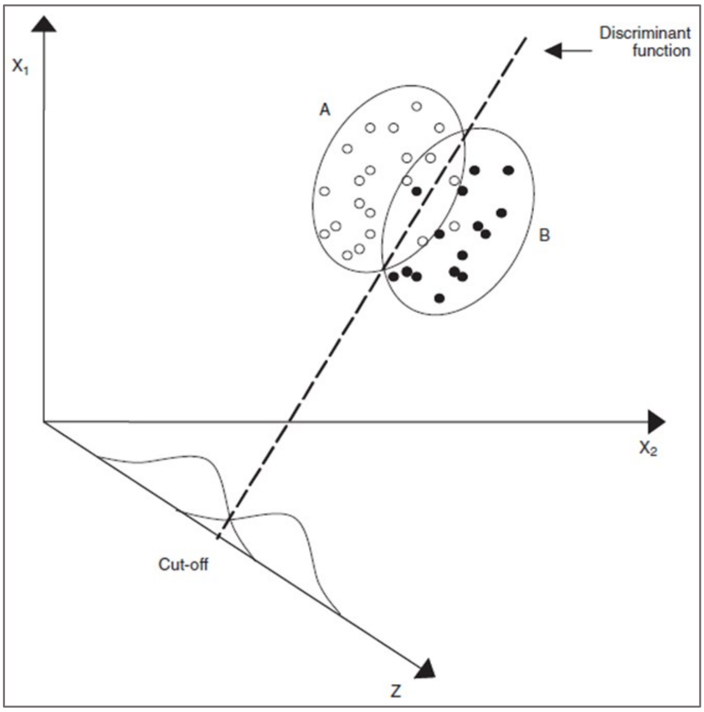


Figure 1: Graphic representation from Resti and Sironi 2007

underneath the analysis. The model described in the figure considers only two variables, x_1 and x_2, used to discriminate between solvent borrowers, set A, and insolvent borrowers, set B. The discriminant function combines the two variables to obtain, for each observation, a score that is shown on the z-axis. It is worth noting that the function represents the cut-off, the threshold score separating the two classes. Given n variables, the linear combination used to compute the score is described by the following formula:

$$z_j = \sum_{i=1}^{n} \gamma_i \, x_{i,j}$$

In this formula, z_j represents the score of the j-th observation while $x_{i,j}$ represents the i-th variables of the j-th observation. γ_i is the coefficient of the variable x_i and the vector γ of all coefficients has to respect the following condition in order to obtain the best partition: $\gamma = \Sigma^{-1}(x_a - x_b)$. x_a and x_b are vectors of the means of the n variables for set A and set B while Σ stands for the matrix of variances and covariances among the variables. Borrowers are assigned to either group, and thus rejected or not, depending on their score and the threshold selected. The selection of the cut-off may depend on the risk policies of the institutions or the model used; a possible cut-off can be the midpoint between the centroids.

The model is very simple and robust and allows to understand the relative importance of each variable used in the model so as to make the selection of only relevant variables and the drop of unimportant variables possible. However, the LDA relies on restrictive assumptions that are unlikely to hold in practice: independent variables that present a multivariate normal distribution and equal variances and covariances matrices for the two groups.

3.2.1 ALTMAN'S Z-SCORE MODEL

One of the most known and used application of the discriminant analysis for credit scoring is the Z-Score model proposed by Altman (1968). This model adopts a multiple discriminant statistical methodology for computing the Z-score that measures the risk of bankruptcy of a corporation -the model considers in fact small and medium enterprises. The variables considered in the model are financial ratios extracted from the balance sheet and the income statement of companies.

The final discrimination function proposed is the follow:

$$Z = 0{,}12\, X_1 + 0{,}14\, X_2 + 0{,}33\, X_3 + 0{,}06\, X_4 + 0{,}999\, X_5$$

Where: X_1 = Working capital/Total assets; X_2 = Retained Earnings/Total assets; X_3 = Earnings before interest and taxes/Total assets; X_4 = Market value equity/Book value of total debt; X_5 = Sales/Total assets.

Variables X_1 to X_4 are highly statistically significant while X_5 is not statistically significant but still is included in the model because of its contribution. The variable with the highest contribution is X_3 followed by the variable X_5 -these variables measures, respectively, the profitability and the overall efficiency. The overall significance of the model is shown by the F-Test (F = 20.7).

Altman proposed two different cut-offs: firms with a Z-score lower than 1.81 are considered high-risk and bankrupt while firms with a score higher than 2.99 are low-risk firms. Firms between 1.81 and 2.99 are in the so-called "zone of ignorance" where classification errors can occur and thus, each lender should decide how to handle these firms depending on its own risk policies.

3.3 LOGISTIC REGRESSION MODEL

Logistic regression is a technique that is used in classification problems where the dependent variable is dichotomous or binary, like credit scoring, since it is able to explain and estimate the likelihood that the binary dependent variable Y assumes value 1 or 0 by analysing a set of explanatory variables. For this reason and because it is robust, easy to develop and interpret, this technique is probably the most used for credit scoring -Y assumes value 1 in case of a healthy loan and value 0 in case of a bad loan.

The logistic regression is an exponential adaptation of linear regression since the usage of tradition linear regression in dichotomous problems -linear probability model- leads to a violation of classical assumptions of regression models and other some difficulties: errors are heteroscedastic and not normally distributed, different coding for Y will lead to different estimates of the model and the predicted Y can assume negative values and values greater than 1, thus making the interpretation of Y as a probability of default more difficult.

The exponential adaptation, thus the logistic model can be expressed by the following formula:

$$p_i = \frac{e^{\alpha+\beta X}}{1 + e^{\alpha+\beta X}}$$

Where p_i is the probability of default of the i-th borrower and can assume values between 0 and 1, α and β are the coefficients to be estimated and X is the vector of n explanatory variables. The formula can be rewrite in order to obtain a linear model in the following way:

$$log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta X$$

The right side of the equation, obtained as a linear combination of explanatory variables, is called log-odds and it expresses the likelihood that the default occurs as a proportion that the default does not occur. This rewrite allows to interpret the results and, in normal conditions, carrying out hypothesis tests on the model and the coefficients as in a standard linear model, providing an easier way to interpret and understand results. As said before, the model provides an estimate of the probability of default -in the credit scoring context- which measures the state of financial health of the borrower and, in this way, explanatory variables can be seen as indicators that provide insights about what can really influence the actual economic and financial situation of borrowers. This can be seen as an advantage over the discriminant analysis that only provides a score stating the borrowers' proximity to the group of defaulted or non-defaulted customers as a measure of the default risk.

3.4 INTRODUCTION TO OTHER MACHINE LEARNING MODELS

Machine learning is a subcategory of the wider field of artificial intelligence. The idea underlying machine learning is to develop an algorithm that is capable of acquiring knowledge and learning from experience by analysing available data and automatically optimizing with limited human intervention. The main approach of machine learning models is the inductive approach through which empirical data are used to train an algorithm to generalize rules and patterns found in these data -this approach is the one used for the classification problem in credit scoring. Machine learning is a broad category with different types of algorithms -for instance logistic regression- and the following part is focused on the presentation of the most used and discussed supervised learning techniques in credit scoring field: decision trees, ensemble methods and neural networks. In supervised learning, data are labelled, as dependent variable or explanatory variables, so that algorithms can explain and predict the variable of interest by considering other variables – in the case of credit scoring, the possible labels for the dependent variable are defaulted or non-defaulted.

3.4.1 DECISION TREES AND ENSEMBLE METHODS

Decision tree algorithms can be used for both regression and classification problems, including credit scoring. A classification tree algorithm can provide solid results through a recursive partition, with binary questions, of the training data into homogeneous and non-overlapping sub-sets, whose internal data appear to be similar in term of values of variables and patterns, and ultimately in term of default risk. The logic followed by Classification And Regression Trees, CART, can be described by the following formula:

$$\widehat{y_i} = \hat{f}(x_i) = \sum_{m=1}^{M} c_m I\{x_i \in R_m\}$$

“Where each observation x_i belongs to exactly one subset R_m. The identity function I returns 1 if x_i is in R_m and 0 otherwise. If x_i falls into R_l, the predicted outcome is $\widehat{y_i}$ = c$_l$, where c_l is the mean of all training observations in R_l” (Albanesi and Vamossy 2019, p.16).

The training data set is firstly split into two sub-sets by selecting one variable and choosing a value used to divide the sample; by doing that the algorithm obtains two nodes with more homogeneous data and the process is repeated until nodes, that will be called terminal nodes, satisfies a specific stopping rule. When the process is over, the results, the partitions and the nodes obtained can be easily presented as a tree structure as shown in figure 2. The variables and the threshold values used for the subdivisions are such as to maximize the deviance between different groups and minimize the deviance, or impurity, within groups. Usually, the structure tends to grow a lot and becomes so complex that the model overfits the data compromising the performance and thus, “the algorithm prunes the resulting tree after it has been fully grown by removing nodes that have resulted from noise in the training sample” (Gunnarsson et al. 2021, p.295).

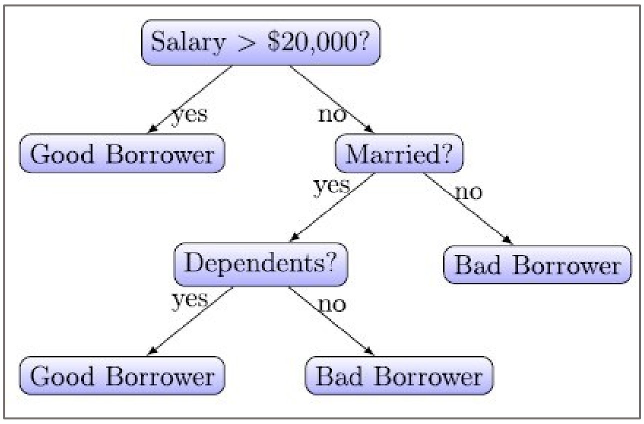


Figure 2: Basic example of a decision tree structure for credit scoring from Dastile, Celik and Potsane 2020

They are suitable for credit scoring since their structure can be easily interpretable and they can discover non-linear relationships between data, that are quite common among data use for credit scoring.

Although these models can perform well, they can be outperformed by ensemble methods that can be considered as an evolution of simple regression trees. One of the most used ensemble techniques is gradient boosting which relies on the idea that a set of different weak models - weak learners that perform just slightly better than a random guessing- can create a model with a better stability and a greater predictive power. The method involves a specific number of steps: first, a shallow regression tree with a weak predictive power is developed over the training data and then a new model is fitted considering the residuals of previous model's predictions. This process is repeated and for each step, data misclassified by the previous model are assigned higher weights than the correctly classified data to be better considered by the next weak learners -the algorithm learns from misclassifications of the various weak learners. In this way Gradient Boosting Trees, GBT, with high accuracy and stability can be obtained as a combination of simplified trees' predictions; however, their structure may result really complex and difficult to interpret.

Extreme Gradient Boosting, or XGBoost, is an improved implementation of the boosting technique known for its fast processing speed, high accuracy and for the surprising performances obtained in different fields of application, including credit scoring. It considers CART models as weak learners and, unlike the traditional gradient boosting, it involves the building of trees in parallel instead of a development in series.

3.4.2 NEURAL NETWORKS

Deep Neural Network models, DNN, are deep learning techniques, a sub-category of machine learning, that can be applied in different fields to improve predictive capacity due to their high ability to reveal complex and difficult-to-detect information -thus they seem suitable for credit scoring.

These models try to recreate the learning of human brain. DNN structure is composed of a series of layers -hidden layers- made of several neurons that interact through synapses with neurons of neighbouring layers. DNNs are basically a sequence of non-linear interactions, where each layer analyses and applies a linear or non-linear function on the data received from the previous layer and transmits the outcome to the next layer that in turn processes the data and transfers the result to the neurons of the following layer. Through this process the model gradually modifies the weights attributed to the connections of variables by the various functions that populate the network. Figure 3 provides an example of two layers neural network structure composed of three input nodes, 4 nodes in the hidden layer an output node which combines the output of previous layer and provides the final result. It is worth noting that each neuron interacts with all the neurons of the previous and following layer, in this way the structure becomes more intricate allowing the identification of highly complex and non-linear patterns and relations difficult to be found even by others machine learning models.

Moving from lower layers to higher layers, it is possible to detect increasingly more complex relationships and thus to improve the performance of the model, however these models tend to overfit data and perform particularly poorly when the depth of the structure exceeds a certain level.

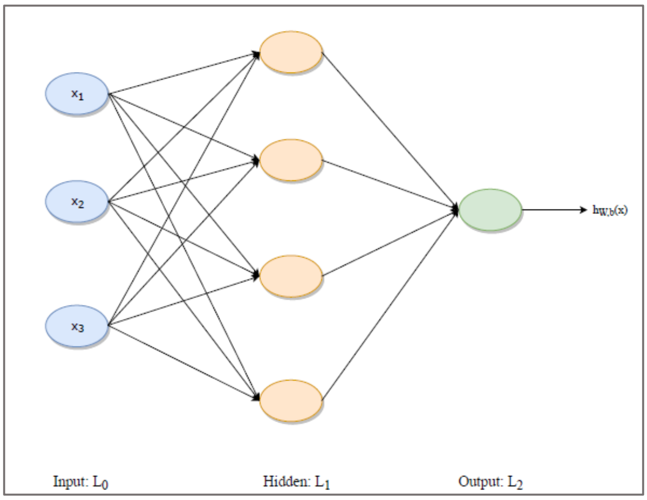
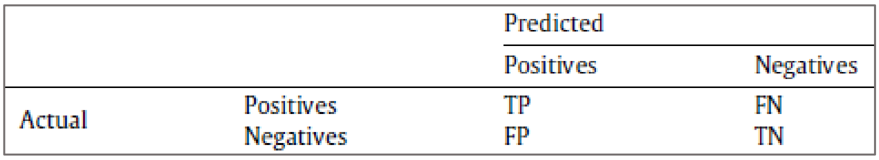


Figure 3: Graphic example from Albanesi and Vamossy 2019

3.5 MEASURES FOR MODEL EVALUATION

In general statistical models can be evaluated considering how well they fit the data -the variability of Y explained by the model- and thus, measures of the goodness of fit of the model, such as R-squared, AIC and BIC criteria, can be used. Considering that credit scoring aims to predict the probability of default of new customer as an estimation of default risk, it is better to evaluate models for their predictive ability by observing the percentage of correctly classified observations.



		Predicted	
		Positives	Negatives
Actual	Positives	TP	FN
	Negatives	FP	TN

Figure 4: Confusion matrix from Dastile, Celik and Potsane 2020

For a specific cut-off, an observation can be predicted as defaulted in case the associated probability of default exceeds that cut-off and as non-defaulted otherwise. In this process two classification errors can be made: Type I, false positive, in case a high-risk observation is classified as non-defaulted and Type II, false negative, in case a low-risk observation is classified as defaulted. When choosing the cut-off, the trade-off between type I and type II errors is considered since each errors represents a cost -i.e. a high cut-off would lead to an increase in type I and decrease in type II. Usually the threshold that minimized potential costs caused by misclassification is chosen.

A common tool useful to evaluate a model performance is a 2x2 confusion matrix as shown in figure 4. The matrix describes the performance of a model by identifying in each cell the number of True Positives -TP-, True Negatives -TN-, False Positives -FP- and False Negatives -FN- and computing the relative frequencies of the entries. Moreover, the confusion matrix can be used to compute different evaluation metrics. The most common one is the accuracy which is the percentage of observations correctly classified and it is computed as the percentage of TP plus the percentage of TN.

Another useful tool is the ROC curve that describes the trade-off between true positive and false positive with different cut-offs, and thus the performance of the model. Figure 5 shows

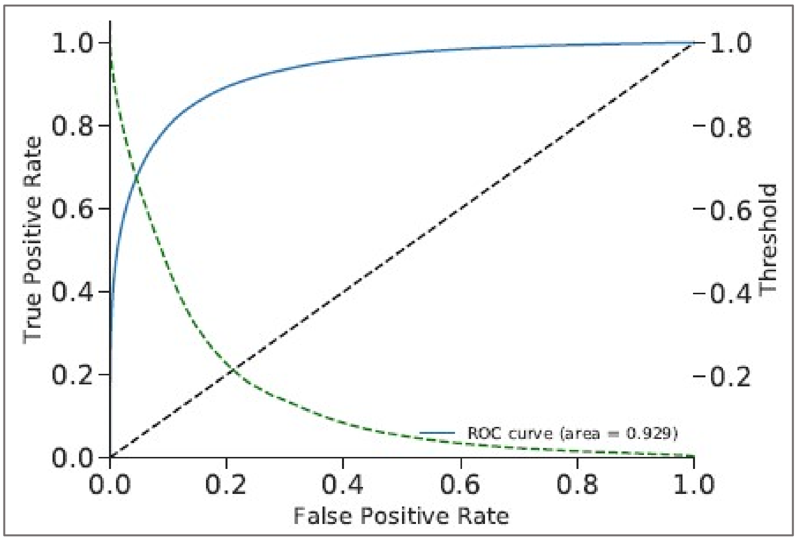


Figure 5: ROC curve from Albanesi and Vamossy 2019

an example of ROC curve and it can be noticed that TP and FP rates vary with variation in the threshold -the green line. The diagonal represents the performance of a random classification and thus the further the curve is from the diagonal and the better the model performance. Following this logic, two of the most common metrics can be computed. The first metric is the AUC score, Area Under the Curve, that provides the overall discriminatory ability of a model by measuring the area underneath the ROC curve -it can be considered "the probability of the classifier assigning a higher probability of being in default to an account that is actually in default" (Albanesi and Vamossy, 2019). The AUC can assume values between 0 and 1 -the closer to 1 the better the model- and it can be easily computed with the following integral:

$$AUC = \int_0^1 y(x)dx$$

The second metric is the Gini coefficient that measures the area between the diagonal and the ROC curve, thus providing the accuracy of the model compared with a random classification. It is computed as $gini = 2 * AUC - 1$. It can assume values between 0 and 1, where 0 corresponds to a random classifier while 1 corresponds to a perfect one.

4. THE HYBRID DNN-GBT MODEL: AN EXAMPLE OF PROGRESS IN CREDIT SCORING

4.1 LIMITATIONS OF CREDIT SCORING MODELS

In this chapter, some limitations of traditional and more complex machine learning models are introduced or recalled in order to build the basis for understanding the advantages brought by the evolution of credit scoring -the discussion is focused on the presentation of a model proposed by Albanesi and Vamossy (2019).

Considering standard models, their main limitation is already discussed in the paper: default behaviour presents complex properties that require a deeper and more sophisticated analysis than the one offered by these models. These properties are the persistent nature of the default status, the non-linear relationships and the multidimensional interactions -high order interactions or variation in the incidence of default due to joint interaction among covariates- between the default and the predictive variables.

A problem affecting credit scoring models used in practice by financial institutions, especially standard techniques, relates to unscorable and invisible customers. Since credit scoring models are fed with traditional data related to credit history of potential customers, loan applicants who don't have credit history -invisible customers- and applicants who don't have a sufficient credit record or do not have recent reported activity -unscorable customers- are quite likely to get rejected from receiving the loan. This creates barrier to accessing credit especially to weaker borrowers -minority, low income and young individuals as discussed in Albanesi and Vamossy 2019- however, new models can mitigate this problem by considering alternative data (World Bank Group, 2019) or requesting a smaller amount of traditional data.

As already discussed, machine learning models present two main issues: overfitting -the model tends to fit the training sample too well and capture specific dynamics of that sample, resulting in a poor out-of-sample predictive performance- and interpretability. The former relates mainly to decision trees and neural networks; in the case of decision trees it can be solved by using ensemble models -boosting techniques mitigate this issue. Interpretability tends to increase as the complexity -thus the predictive power- of the model increases. In fact, while the interpretability of decision trees is quite straightforward its predictive ability is generally modest compared to ensemble and neural network models. Some techniques have been developed for the interpretation of sophisticated models, thus for mitigating this problem. Two common techniques for interpreting machine learning models are the SHAP value, a recent game theoretic approach that allows to estimate the -economic- contribution of each feature to the final output, and the LIME that explains the behaviour of the model through variations in features' values.

4.2 DATA AND MODEL PRESENTATION

Albanesi and Vamossy (2019) proposed a hybrid model -DNN GBT model- for predicting the default probability of customers. It is proven to have better performance compared to theoretical and credit score models used in the industry - in this case it allows to obtain a positive outcome in term of added value for lenders and borrowers - and provides important insights for the credit scoring practice.

Starting from the dataset, important considerations have been made to mitigate possible issues and limitations. The dataset used comes from the Experian credit bureau and covers a period of 12 years from 2004Q1 to 2015Q4. The sample contains data about 1 million households and considers an important amount of variables -more that 200- related to type of credit, credit history, eventual bankruptcy status, and quarterly borrower's credit score. Among all the available variables, the model considers only variables from the credit report that are used by current credit scoring models to be in compliance with anti-discrimination laws and be consistent with these models for a proper performance comparison. Moreover, the exclusion of lagged features allows the model to score the portion of unscorable borrowers with limited credit history, thus expanding the predictive coverage and reducing barriers to accessing credit without the use of alternative data sources that could give rise to privacy and disparity problems (World Group Bank, 2019).

The incidence of default in the dataset is about 34%, computed as the ratio of households with a 90+ days delinquency in the subsequent 8 quarters -this is a common definition of default used in the industry. Checking the incidence of default is important to prevent the dataset from being highly imbalance and therefore compromising and distorting the performance of the model -as pointed out by Dastile, Celik and Potsane (2020) imbalance datasets are a major issue in credit scoring literature.

To train the model and obtain out-of-sample results, the panel data is split into quarters and the training sample is made up of data from 8Q prior the test sample data -thus the model is trained considering one quarter and tested on 8 quarters apart. The hybrid model is built through a combination of results coming from an ensemble learning algorithm and a deep neural network model. The ensemble model chosen is an extreme gradient boosting algorithm while the neural network is built with 5 hidden layers and a dropout technique is applied to the network to mitigate the problem of overfitting -the dropout consists of the random drop of neurons from each layer of the model during the training. The final result of the model, the probability of default of each observation, is computed as the arithmetic mean of the results obtained separately by the DDN and the GBT.

4.3 REVIEW OF THE MODEL ADVANCEMENTS

In the following part, the results obtained by the DNN-GBT model are presented and integrated with considerations to explain advantages brought by machine learning. Its performance is compared to the one of several single models - logistic regression, CART, ensemble models and DNN - and to the one of a conventional credit score. Moreover, a brief presentation of insights coming from the interpretation of the model is made and additional benefits of the model are presented.

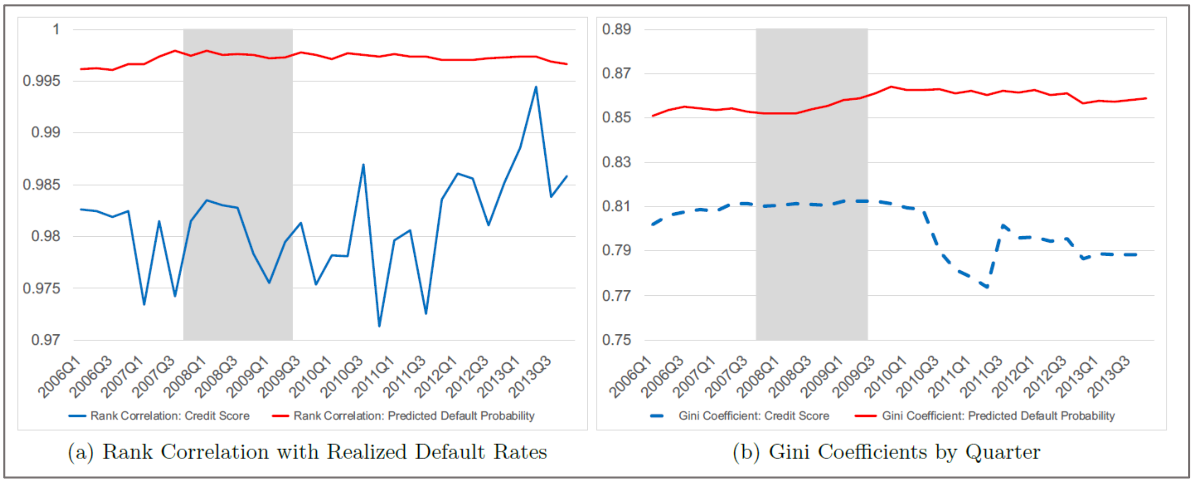
4.3.1 PERFORMANCE COMPARISONS

Overall it is possible to state that the hybrid model shows strong and sound in-sample and out-of-sample performances that surpass other single models' performance. First of all, the model is able to discriminate between delinquent and non-delinquent observations. In fact considering in-sample performance -both training and testing sample cover the same time window from 2004Q1 to 2013Q4-, the model default prediction among defaulted observations is 75.83% whereas the prediction among non-defaulted observations is 11.81%. Considering out-of-sample performances, defaulted observations are associated with predictions between 64.99%-73.67% while non-defaulted predictions range from 12.04% to 15.87%. The strong predictive power is also proved by high levels in the accuracy and in the AUC-score -along with other performance metrics-: the out-of-sample accuracy is always above 86% in all the periods and the AUC-score always exceeds 92%.

The comparison with other single models is conducted by considering as the main performance metric the out-of-sample loss, which measures the divergence between the predicted probabilities and the actual values; thus the lower the loss metric, the lower the distance between prediction probabilities and the actual values, and the higher the performance of the model. The models considered are the most common models used for credit scoring that are GBT, DNN, Random Forest - RF- , CART and logistic regression. Respectively they present the following average loss: 0.3177, 0.3216, 0.3220, 0.3377 and 0.3476. The hybrid model presents the lowest loss -0.3171- among all the models, thus confirming its superior performance. Furthermore, more complex models, GBT and DNN, able to capture complex relations among data, prove to be better than simpler models such as logistic regression - the weakest - and CART. This finding is consistent with the result obtained by Gunnarsson et al. (2021) and other credit scoring literature as proposed by the meta-analysis conducted by Dastile, Celik and Potsane (2020) - of course performances might slightly differ depending on the dataset and the specific features of the model developed.

The model also shows a better performance than conventional credit score. Figure 6 displays correlations between credit score or predicted probability with the realized default rate -in the case of credit score, the correlation is negative- and the Gini coefficient. It can be seen that the model performance is better and more consistent than the one of conventional credit score which presents a higher variability and a reduction in the performance -measured by the Gini coefficient- in the years following the 2007-2008 financial crisis. Thus, the model demonstrates the ability to maintain a sound performance even with changes in the economic conditions -FDIC (2007) presents the limitation related to the effectiveness of credit scoring models with changing economic conditions.

The increase in the performance provided by more complex models translates into practice as added value for lenders and borrowers. Lenders benefit from a reduction in losses caused by borrowers' defaults while borrowers whose risk is misclassified as a higher risk benefit from a reduction in the interest paid. In the hybrid model -compared to logistic regression-, lenders' saving range from 1% to 9% while borrowers added value varies depending on the risk class, for instance it can reach around $1,426 per year for the high-risk class -the savings per capita amount to $40.



(a) Rank Correlation with Realized Default Rates

(b) Gini Coefficients by Quarter

Figure 6 from Albanesi and Vamossy 2019

4.3.2 INTERPRETABILITY

The interpretation to assess the economic importance of variables is conducted through the usage of the SHAP value. This allows to find variables with the highest weight on the output and thus, it is used by lenders to state the most relevant factors used during the creditworthiness assessment -for instance, Bonaccorsi di Patti et al. (2022) reports the SHAP value to be the most used explainability technique by Italian financial intermediaries. Considering the hybrid model, the SHAP value is assigned to each variable or group of those variables who present a high correlation -to prevent from weight splitting across highly correlated variables- and the results of the most significant variables are presented in figure 7. The most important variables or group of variables are worst status on any trades, whose high values are associated with high values of predicted default rates, and credit history features -delinquencies and months since the oldest trade was opened- whose high values go along with low values of predicted probability. Other important variables regard credit utilization and amount of credit -in the latter, high values of the credit correspond to low level of default since the allocation of credit mainly targets borrower with low default risk.

Thanks to interpretability techniques, sophisticated models, like the hybrid model, are able to identify and provide and explanation of the main features influencing the default risk. In this sense, credit scoring models' usefulness does not merely stop at estimating the default of probability, but it also extends to the understanding of the default dynamics, with possible reliable insights for financial institutions in the managing of credit lines, and policy makers in the design of policies.

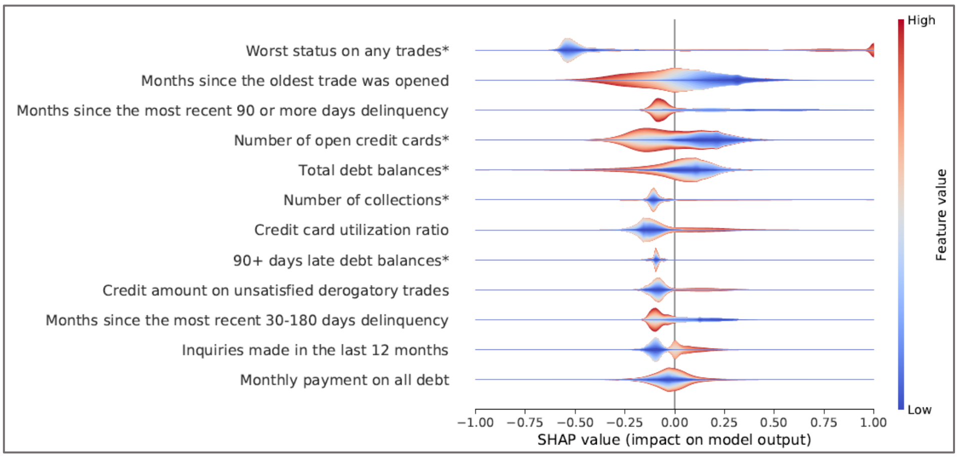


Figure 7: SHAP values of the 12 most important variables. The asterisks identify groups of variables.
Albanesi and Vamossy 2019

4.3.3 SYSTEMIC RISK PREDICTIONS

As other machine learning models, the hybrid model proposed by Albanesi and Vamossy (2019) is able to provide a sound estimation of the absolute risk level linked to a borrower. This is different from traditional techniques used in the industry since, as argued by FDIC (2007), they are used to obtain only a relative risk-ranking of borrowers. Thanks to the absolute level, it is possible to aggregate the probabilities of default of all borrowers and obtain a prediction of the systemic credit risk, and this can have interesting turn ups in macroprudential regulations and policies.

5. CONCLUSIONS

This thesis has presented the evolution of credit scoring models and has shown how the introduction of sophisticated machine learning techniques has been able to significantly expand the potential of these models and thus, attach an ever greater importance to them.

In the last decades, there has been a spread of the use of credit scoring models in the financial industry in various processes such as creditworthiness assessment for credit granting, pricing of financial instruments and setting regulatory capital. Financial institutions have found these models a useful tool for responding to changes in the credit industry, for instance the exponential growth of consumer credit and the increase in prudence toward risk, due to uncertainty caused by financial crisis and economic instability. Moreover, the increase in computational power and innovation in machine learning techniques have brought significant improvements in the reliability and the performance of these models -their importance and relevance are recognized by financial institutions and also by regulatory authorities that have been designed guidelines for the correct usage of these models to provide stability and reliability to the financial system.

The evolution in credit scoring models made it possible to develop sophisticated and complex machine learning models that are able to handle and analyse big amount of conventional and alternative data -i.e., big data- and to capture non-linear relations and hidden patterns of the default dynamic. These characteristics allow to achieve significant improvements in term of performances that translate into a reduction in default losses for banks and thus, more stability. Despite the achievements obtained, the potential of machine learning models has not been fully explored yet and some limitations and challenges still have to be addressed.

Furthermore, these models have demonstrated to be useful in other areas of application. In fact, they can provide insights and suggestions to policy makers and regulatory authorities for the development of policies to prevent and reduce defaults and the prediction of instability periods.

REFERENCES

ALBANESI, S., VAMOSSY, D.F., 2019. Predicting consumer default: a deep learning approach. *NBER Working Paper Series,* No. 26165. Available on: <https://www.nber.org/papers/w26165>.

ALTMAN, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23: 589-609. Available on: <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>

BONACCORSI DI PATTI, E., et al., 2022. *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano (occasional paper) n. 721.* Roma: Banca d'Italia. Available on: <https://www.bancaditalia.it/pubblicazioni/qef/2022-0721/QEF_721_IT.pdf>.

BOUTEILLE, S., and COOGAN-PUSHNER, D., 2012. *The handbook of credit risk management: originating, assessing, and managing credit exposures.* 1° ed. Hoboken, New Jersey: Wiley.

DASTILE, X., CELIK, T., and POTSANE, M., 2020. Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing Journal,* Vol. 91, 106263. Available on: <https://doi.org/10.1016/j.asoc.2020.106263>.

DEL PRETE, S., et al., 2013. *Organizzarsi per prestare in tempo di crisi. Risultati di un'indagine sulle banche (occasional paper) n.154.* Roma: Banca d'Italia. Available on: <https://www.bancaditalia.it/pubblicazioni/qef/2013-0154/QEF_154.pdf>.

EUROPEAN BANKING AUTHORITY, 2020. *Final report – guidelines on loan origination and monitoring.* European Banking Authority. Available on: <https://www.eba.europa.eu/sites/default/documents/files/document_library/Publications/Guidelines/2020/Guidelines%20on%20loan%20origination%20and%20monitoring/884283/EBA%20GL%202020%2006%20Final%20Report%20on%20GL%20on%20loan%20origination%20and%20monitoring.pdf>.

EUROPEAN CENTRAL BANK, 2023. *Economic Bulletin Issue n.5.* European Central Bank. Available on: <https://www.ecb.europa.eu/pub/pdf/ecbu/eb202305.en.pdf>.

EUROPEAN CENTRAL BANK, 2023. *The euro area bank lending survey. Second quarter of 2023.* European Central Bank. Available on:

<https://www.ecb.europa.eu/stats/ecb_surveys/bank_lending_survey/pdf/ecb.blssurvey2023q2~6d340c8db6.en.pdf>.

FEDERAL DEPOSIT INSURANCE CORPORATION, 2007. *Credit Card Activities Manual. Scoring and Modeling (Section VIII).* FDIC. Available on: <https://www.fdic.gov/regulations/examinations/credit_card/pdf_version/ch8.pdf >

GUNNARSSON, B.R., 2021. Deep learning for credit scoring: Do or don't?. *European Journal of Operational Research,* Vol. 295, Issue 1, 292-305. Available on: <https://doi.org/10.1016/j.ejor.2021.03.006>.

LESSMANN, S., 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research. *European Journal of Operational Research*, Volume 247, Issue 1, 124-136. Available on: <https://doi.org/10.1016/j.ejor.2015.05.030>.

ONAY, C., and ÖZTÜRK, E., 2018. A review of credit scoring research in the age of Big Data. *Journal of Financial Regulation and Compliance*, Vol. 26 No. 3, 382-405. Available on: <https://doi.org/10.1108/JFRC-06-2017-0054>.

PROTO, A., ed.by., 2023. *Banking transaction and services.* 1° ed. Torino: G. Giappichelli Editore srl.

RESTI, A., and SIRONI, A., 2007. *Risk management and shareholders' value in banking from risk measurement models to capital allocation policies.* 1° ed. Chichester, West Sussex, England; Hoboken, NJ: Wiley.

STANGHELLI, E., 2009. *Introduzione ai metodi statistici per il credit scoring.* 1° ed. Milano: Springer.

THOMAS, L.C., EDELMAN, D.B., and CROOK, J.N., 2002. *Credit Scoring and Its Applications.* Philadelphia: Society for Industrial and Applied Mathematics.

WORLD BANK GROUP, 2019. *Credit scoring approaches guidelines.* Washington: The World Bank Group. Available on: <https://pubdocs.worldbank.org/en/935891585869698451/CREDIT-SCORING-APPROACHES-GUIDELINES-FINAL-WEB.pdf>.

Number of words: 9795