ASSOCIAÇÃO DE POLITÉCNICOS DO NORTE (APNOR)
INSTITUTO POLITÉCNICO DE BRAGANÇA

Riskit: Investment risk assessment platform

Miguel de Lacerda Pereira

Project report submitted to *Instituto Politécnico de Bragança*
To obtain the Master Degree in Management, specialization in Business Management.

Supervisors:
Ana Paula Carvalho do Monte, Ph.D.
Rui Pedro Sanches de Castro Lopes, Ph.D.

Bragança, december, 2023.

ASSOCIAÇÃO DE POLITÉCNICOS DO NORTE (APNOR)
INSTITUTO POLITÉCNICO DE BRAGANÇA

Riskit: Investment risk assessment platform

Miguel de Lacerda Pereira

Supervisors:
Ana Paula Carvalho do Monte, Ph.D.
Rui Pedro Sanches de Castro Lopes, Ph.D.

Bragança, december, 2023.

Abstract

The current society is volatile, influenced by macro social, economic, geopolitical, and natural phenomena that have a global and deeply interconnected impact. As a result, as unpredictability increases, access to information and decision-support tools becomes increasingly vital in all aspects of social life. The capital market (and companies) is at the forefront of these phenomena, given its volatility and extreme exposure to these macro events.

In this scenario, the objective was to develop a platform that predicts insolvencies. The Riskit: Insolvency Predictor is a web-based platform aimed at assisting the scientific community and investors in predicting the possibility of companies becoming insolvent based on specific financial indicators.

Methodologically, a dataset of 15,000 Portuguese companies was randomly extracted from the Iberian Balance Sheet Analysis System (SABI) database[1]. An analysis was conducted, resulting in the selection of 11 financial indicators used for predictions. To make predictions, the authors conducted a comprehensive study of models commonly used for this type of forecasting and also experimented with some machine-learning models that are not frequently mentioned in the literature. The evaluation of the application's performance in predicting insolvencies is measured by a series of performance benchmarks calculated with the help of a confusion matrix.

It was found that models frequently mentioned in the literature do not always have better performance. The main objectives of this project were achieved, providing both the scientific community and investors with a tool that predicts insolvency using a set of financial indicators and demonstrating the value of machine-learning models for making these predictions. The application can be visited at https://riskit.ipb.pt/.

Keywords: Insolvency prediction, Risk management, Financial indicators, Machine learning models, Web-based application.

[1]The SABI database was made available by the Applied Management Research Unit (UNIAG), through an agreement protocol between UNIAG and COFACE.

Resumo

A sociedade atual é volátil, atravessada por macro fenómenos sociais, económicos, geopolíticos e naturais com impacto à escala global e profundamente interrelacionados. Em consequência, à medida que aumenta a imprevisibilidade, o acesso à informação e a instrumentos de ajuda a decisão são, cada vez mais, vitais em todos os aspetos da vida social. O mercado de capitais (e empresas) está na linha da frente destes fenómenos, pela sua volatilidade e extrema exposição a esses macro fenómenos.

Neste cenário objetivamos desenvolver uma plataforma que faz a previsão de insolvências. A *Riskit: Insolvency Predictor* é uma plataforma baseada na *web* e tem como objetivo auxiliar a comunidade científica e investidores na previsão da possibilidade de empresas se tornarem insolventes, com base em indicadores financeiros específicos.

Metodologicamente foi utilizado um conjunto de dados de 15.000 empresas portuguesas extraídas aleatoriamente da base de dados SABI[2]. Realizou-se uma análise da mesma, resultando na seleção de 11 indicadores financeiros usados para as previsões. Para realizar as previsões, os autores fizeram um estudo compreensivo de modelos habitualmente usados para este tipo de previsão e experimentaram também alguns modelos de *machine-learning* que não são frequentemente mencionados na literatura. A avaliação do desempenho da aplicação na previsão de insolvências é medida por uma série de *benchmarks* de desempenho calculados com a ajuda de uma matriz de confusão.

Verificou-se que os modelos mencionados com mais frequência na literatura nem sempre têm melhor desempenho. Os principais objetivos deste projeto foram alcançados, oferecendo tanto à comunidade científica quanto aos investidores uma ferramenta que prevê a insolvência usando um conjunto de indicadores financeiros e comprovando o valor dos modelos de *machine-learning* para fazer estas previsões. A aplicação pode ser visitada em https://riskit.ipb.pt/.

Palavras-chave: Previsão de insolvência, Gestão de risco, Indicadores financeiros, Modelos de *machine-learning*, Aplicação alojada na *web*.

[2]A base de dados SABI foi disponibilizada pela UNIAG, através de um protocolo celebrado entre a UNIAG e a COFACE.

Dedication

I dedicate this work to my family, especially to my father Fernando, who from a young age developed in me a will to persue knowledge and believe on the scientific method, and always encouraged and supported me in pursuing my goals. To my dear mother Isabel who raised me to be the person I am today. To my wife, who provided encouragement and strength throughout this journey, leading by example and balancing two jobs while also taking a master's degree. To my sister and to my friends, for keeping me company and preserving my sanity throughout this process. I also dedicate this work to my coleagues at ITSector and Polytechnic Institute of Bragança (IPB). To all the students, professors, researchers, scholarship holders, and countless others who strive to make an impact on the world through their knowledge, research and discoveries. Lastly, I dedicate this work to both of my cats, who kept an eye on me and also made extensive contributions to this work stepping on the keyboard and taking long naps.

Acknowledgments

I would like to express my gratitude to my thesis advisor and proposer of this work, Professor Ana Paula Carvalho do Monte, Ph.D. from IPB. The guidance she gave me when the project was starting, for the knowledge she passed to me while I was developing this work, and also for having the vision to concieve this project.

I'm also deeply thankfull to Professor Rui Pedro Sanches de Castro Lopes, Ph.D. from IPB. Not only because of all the technical guidance he gave me on the development of the application, that resulted on a valuable and usefull tool for students, researchers, managers, investors and analysts. But also for using is guidance to empower me and other students to push themselves in order to be better at their work. He truly is a teacher and a mentor.

To my father, Professor Fernando Augusto Pereira, Ph.D. from IPB, for all the patience and help he gave me on all the stages of this work. It really makes it easier to work, when you have someone with so much knowledge and patience next to you.

To Professor Marco Pinheiro from University Institute of Lisbon (ISCTE), for the support and advice on the initial drafts of this application.

A special thank to the representatives of SABI for granting me the access to the database, and made it possible to deliver this work.

To Professor Paula Odete Fernandes, Ph.D. from IPB, to the UNIAG, to my school School of Technology and Management (ESTiG), to my Institution IPB and to Association of Polytechnic Institutes of the Northern Region (APNOR).

To my wife and my family for the emotional support and encouragment they gave me troughout my academic journey.

Table of Contents

List of Figures . . . viii

List of Tables . . . ix

Introduction . . . 1

1 Literature Review . . . 3
- 1.1 Risk and profitability of the financial market . . . 3
- 1.2 Fundamental analysis: variables to consider in the analysis of economic and financial performance of companies . . . 5
- 1.3 Instruments and tools for evaluating the risk of companies and investments . . . 6
- 1.4 Business bankruptcy estimation models . . . 7
- 1.5 Web Platforms and Machine Learning . . . 8
  - 1.5.1 Web Platforms . . . 8
  - 1.5.2 Machine Learning . . . 9

2 Research Methodology . . . 11
- 2.1 Objectives and Research Questions . . . 11
- 2.2 Description of Variables and Models . . . 11
- 2.3 Description of the Platform Skeleton . . . 14

3 Presentation and Analysis of Results . . . 15
- 3.1 Sample characterization . . . 15
- 3.2 Description of Riskit: Insolvency Predictor Platform . . . 16
  - 3.2.1 Main Page - Insolvency Prediction, Input Suggestion, and Model Information 16
  - 3.2.2 Other Pages and components . . . 18
  - 3.2.3 Reflective synthesis . . . 20

Conclusion, Limitations and Future Research Lines . . . 21

Bibliography . . . 21

Appendix . . . 24

List of Figures

Figure 1	Insolvency Predictor - Prediction example	16
Figure 2	Insolvency Predictor - Financial indicators input suggestion	17
Figure 3	Insolvency Predictor - Model info	17
Figure 4	Insolvency Predictor - Financial indicator description	18
Figure 5	Insolvency Predictor - Best Model	18
Figure 6	Insolvency Predictor - Model details page	19
Figure 7	Insolvency Predictor - About page	19
Figure 8	Insolvency Predictor - Header	20
Figure 9	Insolvency Predictor - Footer	20
Figure A1	Principal Component Analysis	37
Figure A2	Exploratory Data Analisys (Link)	38

List of Tables

Table A1 Financial indicators used in the investment appraisal 24
Table A2 Performance benchmarks 26
Table A3 Companies by Economic Activity Code 27
Table A4 Average scores of companies (active vs insolvent) 29
Table A5 Measures of association between companies' Economic Activity Code and status 33

Introduction

The main goal of this project is to develop an application that provides the scientific community and investors with a tool they can utilize for their work. The tool specializes in using machine learning and other models to predict the insolvency of companies based on specific financial indicators. Beyond this goal, the project also seeks to benchmark the performance of various models in insolvency prediction and foster data-driven research in finance, management, and risk.

This document is composed of three main chapters: a theoretical framework, a methodology, and a final chapter dedicated to the application presentation.

The theoretical framework chapter serves as the foundation for the project, diving into the main theoretical concepts that contributed to the application's development. It begins with a brief exploration of the sociological landscape and its impact on decision-making. Among other topics, the framework explores concepts related to risk, financial markets, and insolvency/bankruptcy prediction. In relation to risk, it examines risk management, the risk profile of investors, risk tolerance, and how these factors influence investment decisions. The chapter also covers fundamental analysis, encompassing financial and non-financial factors to consider when analyzing the financial and economic performance of companies. Additionally, it explores tools for evaluating risk in companies and investments, such as scenario analysis or break-even analysis. Furthermore, the chapter scrutinizes business bankruptcy/insolvency estimation models. The final section of the chapter offers a brief overview of web platforms and the integration of machine learning. All of these topics, along with their integration, lay the groundwork for the application's implementation and development.

The methodology explains the approach taken to fulfill the objectives. Among other things, it provides a description of the variables and models used in the application. It begins by explaining the data analysis conducted to reduce the number of financial indicators used in the application compared to the number present in the database. To achieve this, a Principal Component Analysis (PCA) and Exploratory Data Analisys (EDA) were performed, leading to the identification of a subset of indicators with low correlation between them (<=0.6). The next part of the chapter is dedicated to explaining the models included in the application. These models were selected based on their suitability for binary classification (insolvent/not insolvent) and their relevance in the field of insolvency prediction, as learned from the theoretical framework chapter. Each model is then explained, and their strengths and characteristics are highlighted. After that, the chapter explains the various benchmarks used to assess the performance of each model. Lastly, this chapter delves into the architecture of the application (front-end, dataset, and Application Programming Interface (API)), highlighting the technologies used to develop these components. In summary, this chapter provides a comprehensive overview of the entire process involved in developing the application.

The third and final chapter of this work is dedicated to the presentation of the application, containing a sample characterization, a detailed description of the functionalities, and a reflective synthesis. It begins by discussing the sample used for the database, consisting of 15,000 random portuguese companies from various Economic Activity Code (EAC) divisions. The chapter analyzes the most represented EACs and provides average values of economic, financial, and operational variables for both insolvent and not insolvent companies. The statistical significance of mean differences between insolvent and not insolvent companies is also evaluated, as well as the correlation between EACs and company status, using Fischer's phi. Following this, the chapter explores the structure of the application. The main page allows users to perform insolvency predictions by selecting a model and inputting the values for the financial indicators. The application provides a prediction result along with an accuracy score, calculated from the confusion matrix. The input suggestion feature is also explained, suggesting values for other financial indicators based on the user's initial input and measuring its Euclidean distance to find the most similar company data in the database. Additionally, the "Best Model" functionality is described, where the model with the highest accuracy is automatically selected. The chapter further covers the Model Page, providing details and information on the performance benchmarks of each model, as well as the About Page, which offers general information about the application and links to the Model Page. Finally, the chapter includes a brief mention of the information contained in the footer of the application. The Reflective Synthesis section concludes the chapter by highlighting the achievements of Riskit: Insolvency Predictor and acknowledging potential areas for improvement, such as adding more models, offering other types of predictions, and incorporating a worldwide dataset of companies.

1. Literature Review

1.1 Risk and profitability of the financial market

Today's society is characterized as being a society of risk and high reflexivity of phenomena (Beck et al., 2000; Giddens, 2000). An increasingly volatile society crossed by macro social, economic, geopolitical and natural phenomena with impact on a global scale and deeply interrelated. As a result, as unpredictability increases, access to information and decision-making tools are increasingly vital in all aspects of social life. The capital market (and companies) is at the forefront of these phenomena, due to its volatility and extreme exposure to these macro phenomena. In Portugal, in the last decade and a half, phenomena of this nature have profoundly affected the environment of companies and the capital market, namely, as pointed out by the *OECD Capital Market Review of Portugal 2020 Mobilising Portuguese Capital Markets for Investment and Growth* (2020): the financial crisis that started in 2008, the subsequent recovery period, the crisis resulting from Covid-19 and at the present time the effects of the Russia-Ukraine war. The increasing diversity, complexity and availability of data in the digital age also causes profound changes in financial systems (Tian et al., 2023).

Risk is understood as the condition that any activity of an individual or company is subject to the effects of unforeseen events and conditions of incomplete information, which may lead to loss or loss of benefits (Kiseleva et al., 2018). Market Risk or Financial Risk is defined as the risk of changes in the market value of a financial instrument or portfolio, due to unexpected changes in market conditions such as changes in share prices, interest rates, exchange rates and volatility of these variables (Resti & Sironi, 2007).

Soares et al. (2015) refer that risk is associated with the probability of a future financial flow not occurring or occurring in a different amount than expected. They emphasize that associating risk solely with loss, while common, is not entirely accurate. This is because the concept of risk, along with its associated probability of occurrence, applies not only to the possibility of a loss, where the actual realized flow is lower than expected, but also to the potential for gain in the opposite situation-when the realized flow is higher than expected. As for the typology of the companies' risks, Kiseleva et al. (2018) distinguish: macroeconomic and national policy risks; financial risks, regulatory fee risks, technology risks; risks associated with competitors and operational risks. Moutinho and Mouta (2013), quoting Lopes and Flavell (1998), point out as the main strategic risks of companies: the lack of an integrated vision of the business that can lead to the under utilization of resources and their best capacities, running the risk of business fragmentation, and also the problem of non-synergies between design and other activities that can lead to business inconsistency; the same authors also point out the risk of concentration when projects are too large when compared to the size of the company that implements them.

With regard to the risks of financial products, the Securities Markets Commission (CMVM) points out the following: market risks (change in quotations); credit risks (imminent default); currency risk (currency devaluation); liquidity risk (impossibility of redeeming capital); fiscal risk (tax aggravation); political risks (devaluation by the authority of the immanent country); interest risk (possibility of interests being subordinated to the interests of the immanent and/or its intermediaries) (Securities Markets Commission - CMVM, 2012). Teixeira (2022) states that the financial risk linked to the company's financial structure should be analyzed from three perspectives: the coverage of financial charges, which verifies whether the results generated by the activity are capable of covering the financial charges arising from the company's indebtedness; treasury, which represents the short-term financial situation, with the financial risk arising from the possibility that normal receipts from the activity are not sufficient to meet obligations with third parties; and the capital structure, which represents the solvency capacity in the medium and long term in which the financial risk is related to the fact that the indebtedness reaches high proportions that may jeopardize the fulfillment of obligations with third parties in the future. Tian et al. (2023) says that financial risks are influenced by international and national policies, macroeconomic performance and psychological expectations of investors, which makes risk management a very challenging activity.

Bank of Portugal (BDP) proposes a complete risk typology comprising nine risk categories, both financial (credit risk, market risk, interest rate risk and exchange rate risk) and non-financial (operational risk, information systems risk, strategy risk, compliance risk and reputation risk) which, due to its scope, we transcribe in full, as follows.

> *CREDIT RISK: probability of occurrence of negative impacts on results or capital, due to the inability of a counterparty to fulfill its financial commitments to the institution, including possible restrictions on the transfer of payments from abroad. Credit risk exists mainly in credit exposures (including securities), credit lines, guarantees and derivatives;*
> *MARKET RISK: probability of negative impacts on earnings or equity due to unfavorable movements in the market price of instruments in the trading portfolio, caused by fluctuations in share prices, commodity prices, interest rates, exchange rates. Market risk is mainly associated with the holding of short-term positions in debt and capital securities, currencies, commodities and derivatives;*
> *INTEREST RATE RISK: probability of negative impacts on earnings or capital, due to adverse movements in interest rates, due to maturity mismatches or interest rate resetting periods, the absence of perfect correlation between fees received and paid on the different instruments, or the existence of options embedded in financial instruments on the balance sheet or off-balance sheet items;*
> *EXCHANGE RATE RISK: probability of negative impacts on results or capital, due to adverse movements in exchange rates, caused by changes in the price of instruments that correspond to open positions in foreign currency or by changes in the institution's competitive position due to significant changes in exchange rates;*
> *"COMPLIANCE" RISK: probability of occurrence of negative impacts on results or capital, resulting from violations or non-compliance with laws, regulations, contracts, codes of conduct, established practices or ethical principles. It may result in legal or regulatory sanctions, limiting business opportunities, reducing the potential for expansion or the impossibility of demanding compliance with contractual obligations;*
> *OPERATIONAL RISK: probability of occurrence of negative impacts on results or capital, resulting from failures in the analysis, processing or settlement of operations, internal and external fraud, the activity being affected due to the use of resources under an "outsourcing" regime, existence of insufficient or inadequate human resources or the inoperability of infrastructures;*
> *INFORMATION SYSTEMS RISK: probability of occurrence of negative impacts on results or capital, as a result of the unadaptability of information systems to new needs, their inability to prevent unauthorized access, to guarantee data integrity or to ensure continuity of the business in case of failure, as well as due to the continuation of an unsuitable strategy in this area;*

STRATEGY RISK: probability of negative impacts on results or capital arising from inadequate strategic decisions, poor implementation of decisions or inability to respond to changes in the environment, as well as changes in the institution's business environment; REPUTATION RISK: probability of occurrence of negative impacts on earnings or capital, resulting from a negative perception of the public image of the institution, whether justified or not, by customers, suppliers, financial analysts, employees, investors, press bodies or opinion general public (Bank of Portugal - BdP, 2007).

After assuming the existence of risk in business activity and in the financial markets, it is important to consider the attitude towards risk, commonly known as risk tolerance. For Schooley and Worden (2016) risk tolerance increases with the healthy state of the individual, with his status as a self-employed worker, with a higher level of education and, on the other hand, it tends to decrease with age, marriage and the existence of dependent minors. These same authors also warn of the fact that investors are not always rational and may react in an exaggerated and emotional way in particular contexts.

For Wach and Chomiak-Orsa (2021) the investment decision process should be seen in a broader context, distinguishing three main types of context: in conditions of certainty when the choice of an alternative unequivocally determines its own effects (such as machinery with clearly defined performance and operating parameters); in risky conditions the choice is uncertain as to its consequences but there is knowledge about the likelihood of possible consequences and outcomes (for instance, gambling); and in conditions of uncertainty, when the decision is taken in conditions of ambiguity in which the effects are known but not their probability of occurrence and in conditions of ignorance when nothing is known about the effects and their probability.

As for the risk profile of the investor, commonly referred to as "Risk-Appetite", investors may have different attitudes: risk aversion, risk neutral or risk appetite (Illing & Aaron, 2005). The Securities Markets Commission - CMVM (2012) typifies investors as: conservative or prudent - looking for products with guaranteed capital invested and profitability; balanced or moderate - looking for products with guaranteed capital invested but prefers longer terms to better accommodate any adversity; dynamic - looking for products with returns above the market average and assuming medium and long-term investments and accepting the risk of some losses on invested capital; and, daring - those who actively seek products with a return higher than the market average, assumes shorter terms and accepts the risk of total loss or even greater than the invested capital.

1.2 Fundamental analysis: variables to consider in the analysis of economic and financial performance of companies

The literature distinguishes between two types of factors or variables to be considered in the analysis of the economic or financial performance of companies, financial factors and non-financial factors.

As for non-financial factors, Moutinho and Mouta (2013) emphasize that the analysis of non-financial factors such as: strategic, technical, commercial, political, social, environmental, organizational, human resources and management is fundamental to the decision-making process. Moutinho and Lopes (2010), in a study of portuguese companies, show that strategic and technical factors are more relevant than financial and commercial factors and that the least relevant factors are those of a social and political nature.

Regarding financial factors used in investment evaluation, the literature distinguishes between indicators that do not consider the time value of money (update) and those that do. You can find in the appendix of this work in table A1 a the synthesis of these factors (Barros, 2000; Martins et al., 2006; Soares et al., 1999).

1.3 Instruments and tools for evaluating the risk of companies and investments

In real life there is risk associated with estimating cash flows. Financial risk analysis consists of evaluating the uncertainties related to a company's financial operations. To estimate the risk we can use the following.

Break Even Analysis (or Critical Point or Dead Point): can be univariate or multivariate and consists of identifying the determining uncertainty variables for the investment project (for example, sales price, quantities, cost of raw materials, personnel costs, amount of investments, discount rate, etc.), assign new values to that variable(s), recalculate the values of cash flows and decision criteria and analyze the impact on the values of these criteria; in projects with constant cash-flows, the critical point represents the maximum value (for variables negatively related to Net Present Value (NPV), such as costs and discount rate) and minimum (for variables positively related to NPV, such as sales price and volume of activity) for this variable without making the NPV negative (Carneiro, 2017; Martins et al., 2006).

Scenario Analysis: this methodology intends to overcome the deficiencies of the sensitivity analysis (namely the univariate); derives from the multivariate sensitivity analysis. It consists on the construction of a set of scenarios that portray the future evolution of the environment in which the investment fits, paying special attention to the critical variables; a limited set of scenarios is considered (optimistic, pessimistic and moderate), assigning a probability of occurrence to each one of them. If these present a normal (or approximate) distribution, the mean or expected value for the evaluation criteria can be estimated, as well as its maximum, minimum and standard deviation (Carneiro, 2017).

Monte-Carlo model: this methodology is based on the use of random numbers in order to obtain a significant sample of the behavior of a system so that it can be analyzed and from there the overall behavior of that system can be deduced. Simulation is a modeling process of a real system that serves as a basis for carrying out experiments, in order to obtain knowledge about its behavior to support decision-making. It highlights the range of possible results and the likelihood of obtaining results other than those derived from using average values. The appearance of the Monte-Carlo (MC) model is located around the year 1944, having since then adapted and changed over time; Currently, the existing computing power allows extensive simulations to be carried out, which help to obtain more accurate results (Platon & Constantinescu, 2014). In quantitative risk assessment, the Monte Carlo (MC) model is highly recognized and used by the academic community and by professionals in the field (Platon & Constantinescu, 2014). The authors also add that the use of this method allows collecting a distribution of all possible results by repeating the use of the model with different values for the different inputs that go according to our simulation. Even so, it is necessary to take into account that the use of the MC model has some limitations, according to Janekova et al. (2021) the deterministic calculation of financial criteria does not provide a versatile integration of investment risk. Especially because in the stochastic simulation approach, the result is dependent on the evaluator, as this can provide misleading inputs for the decision-making process or selection of risk factors.

Real Options: in a context of great dynamism and uncertainty, the measure of value creation by an investment must value its flexibility, i.e., its ability to respond, in a flexible way, to the future evolution of operating conditions. A Real Option, like financial options, is the right, but not the obligation, to take an action that affects a real physical asset, at a predetermined cost, during a predetermined period of time, for example, to make vary the produced quantity of a product, produce other products, close the activity or postpone investments (Martins et al., 2006).

In risk assessment there are other aspects that we should take into account besides the instruments and tools for this evaluation. The next chapter will focus on a specific aspect of this crucial process: business bankruptcy estimation models. These models represent a fundamental component of risk evaluation, offering valuable insights into the financial stability of companies and investments.

1.4 Business bankruptcy estimation models

When investing in a company, it is important to understand the reasons why some companies go bankrupt and other companies remain solvent (Romão, 2009). According to Altman and Hotchkiss (2005) the literature commonly points to four generic terms to define bankruptcy which, however, present formal distinctions: failure, when the return on invested capital of a company is significantly and continuously lower than the rates of return prevalent in other similar investments; insolvency when a company is unable to fulfill its obligations meaning a lack of liquidity, which may be temporary or lead to bankruptcy; default or non-compliance when a company violates a condition agreed with a creditor; and bankruptcy when the formal declaration of bankruptcy is assumed with the competent legal entity.

The causes that lead companies to bankruptcy are numerous and the risk of bankruptcy cannot be eliminated, but it can be identified before it occurs (Li et al., 2009; Voda et al., 2021). The use of specific indicators of liquidity and financial solvency are essential elements to identify financial problems, providing clues about the company's ability to pay the tax obligations owed, about the extent to which the equity can cover the debt and, also, about the current financial resources available (Voda et al., 2021). In order to maintain the prosperity and competitiveness of a company, it is extremely important to know the financial situation in which it finds itself, since adequate management decisions cannot be made without a high-quality, comprehensive and timely diagnosis, supported by a detailed analysis of the adverse phenomena that threaten the company's operations (Horváthová & Mokrišová, 2020).

Several studies point out that good corporate governance is essential to overcome financial difficulties (distress) that can determine the probability of bankruptcy (Sewpersadh, 2022; Voda et al., 2021). Companies with a consistent shareholder portfolio, with a qualified and committed management body, and with good audit processes are associated with a low probability of distress (Sewpersadh, 2022).

As for business failure prediction models, Wu et al. (2010) identify five main models: the Multivariate Discriminant Analysis (MDA) model based on accounting variables by Altman (1968); the logit model (Logistic Regression Analysis) with accounting indices from Ohlson (1980); the probit model (Logistic Regression Analysis) using accounting data from Zmijewski (1984); the risk model with accounting and market variables by Shumway (2001); and the BSM-Prob model based on accounting and market variables by Hillegeist et al. (2004). The disadvantage of this last model is that it is only applicable to companies listed on the stock exchange (Hillegeist et al., 2004; Romão, 2009). According to Yang et al. (1999) the bankruptcy prediction models have used a variety of statistical methods: linear discriminant analysis (Altman, 1968; Altman, Haldeman & Narayanan, 1977), regression analysis

(Korobow, Stuhr & Martin, 1976), logit regression (Barth , et al., 1985; Pantalone & Platt, 1987), and weighted average maximum likelihood estimation (Zmijewski, 1984). More recently, the use of Neural Networks (NN) based models has been gained acceptance (Atiya, 2001; Yang et al., 1999; Zhang et al., 1997).

Concerning the financial ratios as inputs, according with Atiya (2001) the Working Capital/Total Assets (WCTA); Retained Earnings/Total Assets (RETA); Earnings Before Interest And Taxes/Total Assets (EBITTA); Market Capitalization/Total Debt (MCTD); sales/total assets, are the financial ratios used by Altman model and have been widely used as inputs, even for Neural Network (NN)s and other nonlinear models. In the study by Wu et al. (2010), in addition to these, the following were used as inputs: Net Income Divided By Total Assets (NITA); Current Liabilities To Current Assets (CLCA); Income From Operations After Depreciation Divided By Total Liabilities (FUTL); Total Liabilities To Total Assets (TLTA); Ohlson O-score = log (total assets/Gross National Product (GNP) price-level index), the index assumes a base value of 100 for 1968; Relative Size = log (the number of outstanding shares multiplied by year-end share price then divided by total market value); price = log of closing price at end of previous fiscal year; LagSIGMA = historical idiosyncratic risk; LagExReturn = lagged excess return; logage = log(years for which firm has traded); and Segment = the number of business segments in the firm.

As for the performance and adequacy of these different models, Álvares (2019) underlines that business failure prediction models generally evolved from univariate financial ratio analysis to multivariate models, and then to logit models that offer the opportunity to directly estimate the probability of bankruptcy under less restrictive statistical assumptions. Wu et al. (2010) corroborate this idea in their exhaustive comparative study of different models (sample from 1980 to 2006 and contains 887 bankruptcies and 49724 non-bankrupt firm-year observations), concluding that: the MDA model based on accounting variables of Altman (1968) has the worst performance relative to other models; that non-linear regression econometric models based on accounting variables (logit model by Ohlson, 1980; and Probit model by Zmijewski, 1984), showed good performance during the 1970s but lost performance in more recent periods; that Shumway's (2001) risk model, which uses market data and company characteristics, generally outperforms models based solely on accounting information; and that the more comprehensive models that infer from key accounting information, market data, and company characteristics are the most reliable for predicting bankruptcy cases. In a study with 1099 companies from various sectors of the Croatian economy, chosen as predictor variables the liquidity, profitability, leverage, efficiency and solvency Bogdan and Bareša (2021) conclude that the model based on MDA, although robust, is outperformed by the logit model . Hillegeist et al. (2004), from a study with 14303 industrial firms, suggest that the main advantages of using option-pricing models in bankruptcy prediction are that they provide guidance about the theoretical determinants of bankruptcy risk and they supply the necessary structure to extract bankruptcy-related information from market prices. Finally, according to Atiya (2001), from the many studies existing in the literature, the NNs based models are, generally, superior to other techniques concerning bankruptcy prediction.

1.5 Web Platforms and Machine Learning

1.5.1 Web Platforms

Web platforms are digital environments accessible through the internet, enabling user interaction and the execution of various online activities and services. Presently, they are employed across various sectors and activities, serving as a pivotal and indispensable element of contemporary life. These platforms facilitate communication, collaboration, and global access to information and resources. In fact, web applicatons serve both as product and a consequence of modern reflexivity, as suggested in the introduction. These platforms can be accessed through specific applications or directly using a web browser with an internet connection.

The widespread adoption and use of the internet began in the early 1990s. Pecini (2018) refers to the subsequent decade until the 2000s as Web 1.0, characterized by closed-type portals with internal links that limited users to a passive role as consumers of the provided services. Since the rise of the so-called web 2.0, the production and sharing of content by the common individual have been accompanied by the increasing presence of platforms in everyday life (Pecini, 2018).

The Web 2.0 is the network as a platform, spanning all connected devices; Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an "architecture of participation", and going beyond the page metaphor of Web 1.0 to deliver rich user experiences (O'Reilly, 2007). The Web 2.0 leverages customer self-service and algorithmic data management to reach out to the entire web, to the edges and not just the center, to the long tail and not just the head (O'Reilly, 2007). Still according to O'Reilly (2007), the core competencies of Web 2.0 companies are: services, not packaged software, with cost-effective scalability; control over unique, hard-to-recreate data sources that get richer as more people use them; trusting users as co-developers; harnessing collective intelligence; leveraging the long tail through customer self-service; software above the level of a single device; lightweight user interfaces, development models and business models.

In a more general way, these web platforms can offer a variety of functionalities and services, such as Social Networks, which allow users to create profiles, share content, connect with friends, family, and other people, participate in groups and communities, among other social features; E-commerce platforms that enable the buying and selling of products and services online, with various payment and delivery options; Online Banking Services offered by financial institutions that allow users to conduct banking transactions, check balances, pay bills, and manage their finances online; Web Applications that offer online applications that perform specific functions, such as text editors, spreadsheets, task managers, among others; Cloud Storage that allows users to store and access their files and data over the internet, securely and conveniently; Media Streaming that provides video and music streaming services, allowing users to watch movies, series, listen to music, and access other content in real-time over the internet; E-learning that offers online courses and educational content, enabling distance learning.

1.5.2 Machine Learning

Machine Learning is a subfield of Artificial Intelligence (AI) that focuses on the development of algorithms and techniques that enable computer systems to learn and improve their performance on specific tasks without being explicitly programmed for each of them. Machine Learning is the science (and art) of programming computers so they can learn from data (Géron, 2019).

Machine-learning technology powers many aspects of modern society, from web searches to content filtering on social networks to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones (LeCun et al., 2015). According to the same authors, Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests, and select relevant results of the search. Increasingly, these applications make use of a class of techniques called deep learning.

According to Simeone (2017), machine learning methods may be useful when:

- The task involves a function that maps well-defined inputs to well-defined outputs.
- Large data sets exist or can be created containing input-output pairs.
- The task provides clear feedback with clearly definable goals and metrics.
- The task does not involve long chains of logic or reasoning that depend on diverse background knowledge or common sense.
- The task does not require detailed explanations for how the decision was made.
- The task has a tolerance for error and no need for provably correct or optimal solutions.
- The phenomenon or function being learned should not change rapidly over time.
- No specialized dexterity, physical skills, or mobility is required.

Géron (2019) states that Machine Learning systems can be classified based on:

- Whether or not they are trained with human supervision (supervised, unsupervised, semi-supervised, and reinforcement learning).
- Whether or not they can learn incrementally on the fly (online versus batch learning).
- Whether they work by simply comparing new data points to known data points or instead by detecting patterns in the training data and building a predictive model, much like scientists do (instance-based versus model-based learning).

The machine learning process typically involves the following steps: data collection, data processing, model training, model evaluation, and model optimization. Data collection involves gathering a significant amount of relevant data from various sources such as databases, sensors, event logs, or even data generated by humans. Data pre-processing is necessary to clean and prepare the data in a suitable format. During model training, the machine learning algorithm is fed with the training data to learn from it and adjust its parameters and internal hypotheses to find patterns and representations that enable it to make predictions or classifications. Model evaluation is performed with unseen data to assess how well the model generalizes and whether it can make accurate predictions. Finally, model optimization allows fine-tuning and improving the model's performance and accuracy.

2. Research Methodology

2.1 Objectives and Research Questions

The objective of this project was to develop an application that provided students, researchers, managers, investors and analysts a tool they can utilize for their work. The tool specializes in using machine learning and other models to predict the insolvency of companies based on specific financial indicators.

The main question that this work aims to answer is whether it is possible to make accurate insolvency predictions using machine learning models trained on a dataset of existing companies. Another important topic of this work is to evaluate the performance of the various used models and compare their results, using a set of performance benchmarks.

Initially, a literature review was conducted using the keywords mentioned in the abstract as descriptors.

2.2 Description of Variables and Models

The dataset extracted from SABI contained more than 50 financial indicators. This posed a problem because creating an application that allows the user to insert more than 50 indicators would make the application very complex and extensive. Therefore, there was a need to find an acceptable number of indicators. To reduce this number, a data analysis was performed on the dataset.

Two kinds of analyses were conducted. The first one was a Principal Component Analisys (PCA). From the analysis made, the results for a PCA with two dimensions were [0.29387479; 0.13869835]. From the result we can conclude that the first principal component explains 29.3875% of the total variance, and the second principal component explains 13.8699%. Together, these components explain around 43.2574% of the total variance in the data. This PCA was then plotted, as shown on the appendix of this work (Figure A1).

The second analysis consisted of performing an Exploratory Data Analisys (EDA). First, when comparing the number of not insolvent (represented by 0) versus insolvent companies (represented by 1). It was possible to observe the insolvent companies represented around 35% of the dataset, the rest being the percentage of the active or not insolvent companies (around 65%). The last part of the EDA was to create a correlation matrix between all the variables. It was then decided to use as the indicators for the application all the variables from the correlation matrix where all correlated values were lower or equal to 0.6. On the appendix, in Figure A2, you can see in red the positive correlated values, in blue you can see the negatives. The stronger the color, the stronger the value (on the right side of the figure, you can see a color scheeme that explains this). Because this correlation matrix is very extensive you can find it in the appendix of the work, you can also view this image online, with more detail, the caption of the image contains the link.

After this process we ended up with the following variables/financial indicators:

Total Assets(€): The value of the resources owned by the company, it can comprise both current assets and long-term assets.

Economic Profitability(%): Also known as Return on Investment (ROI), it is used to evaluate the profitability of an investment. It measures the return of profit generated by the value of the investment after deducting all expenses, costs, and taxes.

Financial Profitability(%): The ratio of profit generated in relation to the entity's financial resources.

General Liquidity(%): Also known as liquidity ratio, it is the financial indicator that measures a company's ability to meet its short-term financial obligations. The value is obtained by dividing the total assets by the total liabilities.

Indebtedness(%): Percentage of the company's assets that are financed by debt. The value is obtained by dividing the total debt by the total assets.

Working Capital(€): The value that results from the subtraction of current assets from current liabilities. A low value may indicate that a company could have difficulty meeting its short-term financial obligations.

Financial Results(€): Also known as net income, it represents the financial outcome of a company's operation after accounting for all revenues, expenses, taxes, and other elements. A proposed formula to calculate this indicator is: Financial results = revenues + other income - cost of goods sold - operating expenses - taxes. Positive values indicate a net profit, resulting in profitability, while a negative value indicates a net loss.

Other Equity(€): Also known as reserves and surplus, it represents the company's cumulative profits that have not been distributed as dividends or reinvested.

Provisions(€): Represent estimated future liabilities of the company that are accounted for in advance. They are created when there is uncertainty or risk associated with upcoming obligations or liabilities.

Income Tax Expense(€): The amount of income tax that a company expects to pay, based on the calculation of its taxable income.

Inventories(€): Also known as "stock", refers to the total value of goods and materials a company holds. Represents the total value of all inventory the company possesses at a specific time.

The application provides a set of models that the user can choose from to calculate predictions. Among the many models available, the following were chosen.

Logistic Regression (LR): A statistical model used for binary classification problems. It estimates the probability of an instance belonging to a particular class using a logistic function. LR is simple yet effective and can be extended to handle multiclass classification tasks using techniques such as one-vs-rest or softmax regression.

Decision Tree (DT): A machine learning algorithm for both classification and regression tasks. DTs partition the feature space based on different criteria to create a tree-like model. They are easy to interpret and can handle both categorical and numerical data, making them widely used in various domains.

Random Forest (RF): An ensemble learning method that combines multiple DTs to make predictions. It creates a set of DT by randomly selecting subsets of features and instances. The predictions from individual trees are combined to make the final prediction. RF is known for its robustness and ability to handle high-dimensional data.

NN: A class of machine learning models inspired by the human brain's neural structure. They consist of interconnected nodes (neurons) organized in layers. NNs can learn complex patterns and relationships from data, making them powerful for tasks such as image recognition, natural language processing, and more.

Adaboost: An ensemble learning method that combines multiple weak classifiers to create a strong classifier. It iteratively adjusts the weights of the weak classifiers to focus on difficult instances, improving overall accuracy. Adaboost is particularly effective in handling complex classification problems.

MC: Uses random sampling and statistical analysis to approximate complex problems. In the context of prediction, MC simulations can be used to estimate the probability of a certain outcome by generating random samples from a given model. MC methods are versatile and widely applicable in various domains.

It is also important for the user to know the performance of each model. To achieve this, the application takes a Confusion Matrix (a table showing the number of True Negatives (TN), False Positives (FP), False Negatives (FN) and True Positives (TP)), where positives are insolvent companies and negatives are active companies. With that it's possible to calculate a series of benchmarks, that are as follows.

Accuracy: The proportion of correct predictions (Eq. 1, Accuracy).

$$\frac{\text{TP} + \text{TN}}{\text{Total Predictions}} \qquad [1]$$

Precision: The proportion of true positive predictions over all positive predictions (Eq. 2, Precision).

$$\frac{\text{TP}}{\text{TP} + \text{FP}} \qquad [2]$$

Recall: The proportion of true positive predictions over all actual positive instances (Eq. 3, Recall).

$$\frac{\text{TP}}{\text{TP} + \text{FN}} \qquad [3]$$

F1 Score: The harmonic mean of precision and recall (Eq. 4, F1 Score).

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad [4]$$

ROC AUC Score: The area under the receiver operating characteristic curve. The Receiver Operation Characteristic Curve (ROC) is a graphical representation of the trade-off between the true positive and false positive rates. The Area Under Curve (AUC) is a scalar value that represents the area under ROC.

The models have different performance benchmarks depending on the inputs inserted by the user. However, we can observe that the RF, DT, and Adaboost (that use ensemble) are the models that have the best benchmark scores. On the other hand, LR, NN, and MC seem to have lower performance. This might suggest that models that use ensemble may be handling the database better, and getting better results. Models like LR or NNs might struggle to model complex relationships when the data has non-linear patterns. You can find the performance benchmarks for each model in the appendix (Table A2).

Another important functionality that the application provides is an input prediction based on euclidian distance. This was implemented using the scipy library. With this implementation the application will suggest values for the inputs that the user has not filled, based on the company with the minimum euclidian distance from the inputs that the user passed.

2.3 Description of the Platform Skeleton

The application is divided into three parts: a dataset, an Application Programming Interface (API) and the front-end.

The first part consists of a dataset of 15,000 random portuguese companies extracted from the SABI database, each with more than 50 financial indicators. From this dataset, we selected the 11 indicators mentioned earlier in this chapter. This dataset was then used to train our models.

The front-end is the user-facing component. It was developed using Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), Bootstrap library and JavaScript (JS). It is responsible to make the connection of the logic implemented on the API and the inputs of the user. It represents what's called in web development the User-interface/User-experience (UI/UX).

Another essential part of this application is the API, which acts as a bridge between the other two components. The API is responsible for implementing the logic of the models, feeding them with the dataset, training them, and providing them as services to the front-end of the application.

The interaction of these three components allows the user to perform insolvency predictions based on the values of the financial indicators and a selected model.

3. Presentation and Analysis of Results

3.1 Sample characterization

The sample used was taken from SABI, comprising a total of 15,000 portuguese companies, as already stated. The sample includes companies from 81 EAC divisions, with a predominance of the commerce, real-estate and industry divisions, as shown in the appendices on table A3.

Also on the appendices you can find table A4, where we present the average values of the economic, financial, and operational variables of companies whether they are active or in a situation of insolvency/bankruptcy. All tested variables do not follow normality and the variables economic profitability, general liquidity, revenues and extraordinary gains, costs and extraordinary losses, and extraordinary results do not exhibit variance homogeneity, thus we used the robust Welch's t-test to evaluate the statistical significance of mean differences. In general, the scores of active companies are higher than the scores of companies in a situation of insolvency/bankruptcy, with this difference being statistically significant for all variables except for General Liquidity (p=0.580), Financial Results (p=0.222), Extraordinary Revenues and Gains (p=0.367), Extraordinary Costs and Losses (p=0.224), and Extraordinary Results (p=0.144). Indebtedness is higher in insolvent/bankrupt companies, but this difference is not statistically significant (p=0.135).

In table A5 in the appendices, we present the relationship between the typology of companies (EAC division) and their active or insolvent/bankrupt status. For this purpose, we used the Fischer's phi, which is a correlation coefficient between two binary nominal variables, and the evaluation of the size of adjusted residuals. We found types of companies correlated with the active situation marked in green in the table and companies associated with the insolvent/bankrupt situation marked in red in the table. This latter situation, during the period studied, seems to be more common in the clothing and leather industries, printing industries, real estate activities, civil engineering and construction, land transportation, catering and similar activities, research and security activities, education, and sports, entertainment and recreational activities.

3.2 Description of Riskit: Insolvency Predictor Platform

The Riskit: Insolvency Predictor is a Python-based web application that uses the Flask framework for its development. It leverages a diverse set of models with a range of financial indicators to help the scientific community and investors predict the insolvency of businesses while providing a user-friendly interface. In the next sections, you will gain an understanding of how the application is divided and the functionalities it provides.

3.2.1 Main Page - Insolvency Prediction, Input Suggestion, and Model Information

The main page, which also serves as the home page of the application, is where the insolvency prediction is made. As this is intended to be an educational tool that is easy to use, a direct approach was chosen to take the user immediately to the insolvency prediction.

Insolvency prediction: To make an insolvency prediction, the user needs only two inputs. The first is to select one of the existing models, as explained earlier in the methodology: LR, DT, RF, NN, Adaboost, MC and a "Best Model" (which I'll explain later). After that, the user must insert at least one value for the financial indicators, which are: Total Assets, Economic Profitability, Financial Profitability, General Liquidity, Indebtedness, Working Capital, Financial Results, Other Equity, Provisions, Income Tax Expense, and Inventories. Once the necessary inputs are provided, the application will automatically make a prediction of "insolvent" or "not insolvent" and display the accuracy of the prediction. This accuracy is calculated from the analysis of the confusion matrix and it generally increases as more financial indicators are filled in by the user. You can see an example of a prediction in Figure 1.

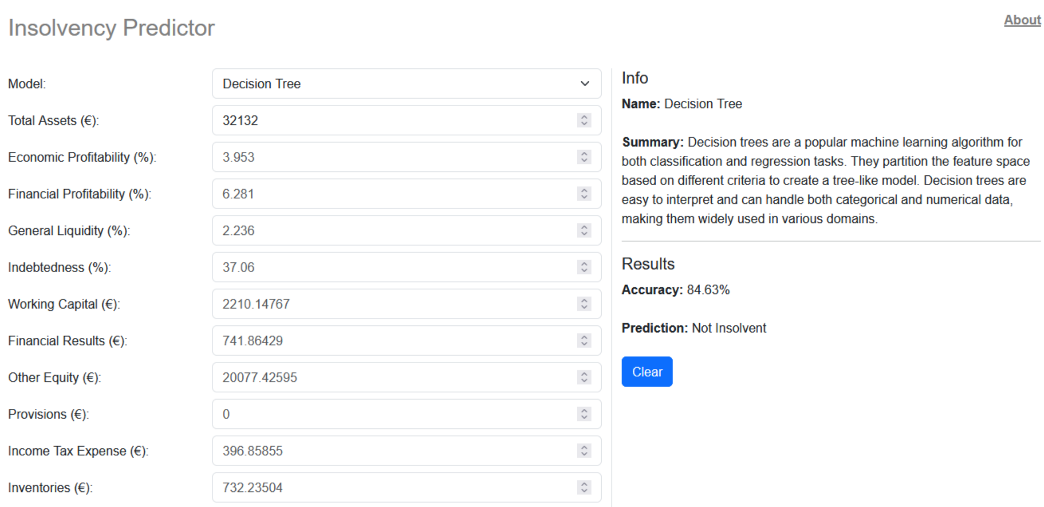

Figure 1: Insolvency Predictor - Prediction example

Input suggestion: Another functionality available to the user on this page is the input suggestion. As soon as the user inserts a value for one of the financial indicators, for instance, Financial Profitability, the application will calculate the Euclidean distance of all companies in the database with the value inserted for this indicator. After finding the company with the lowest distance, the application will suggest the values of the other financial indicators to the user. These suggested values will appear in grey, and the user must double-click on them to accept them, as shown in Figure 2. This is especially useful when the user is not sure of the values that it should input, and can count on a real value for each financial indicator from the company with the lowest Euclidian distance.

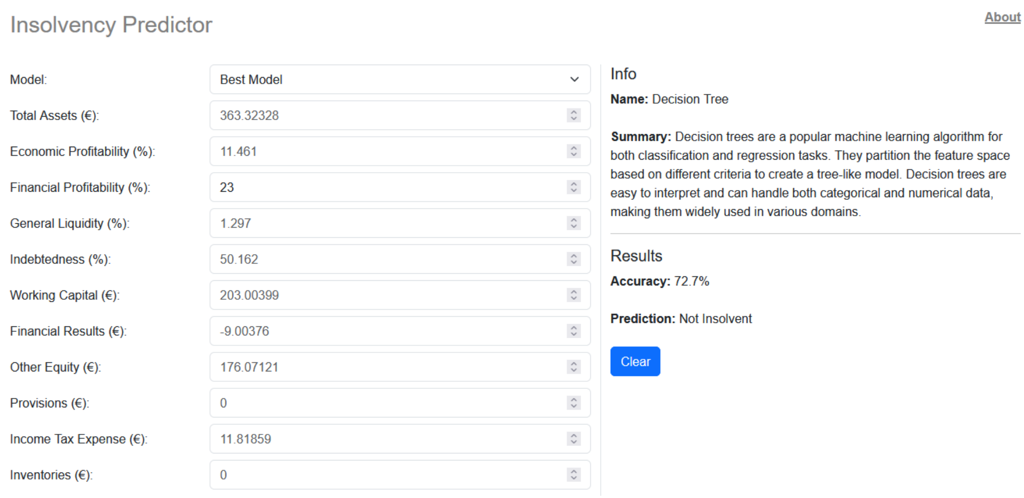

Figure 2: Insolvency Predictor - Financial indicators input suggestion

Model info: In Figure 3 it is possible to see that when the user selects a model, it may be important to provide some information about the model details. On the right, the user will be able to see the name of the selected model and also a brief summary of it.

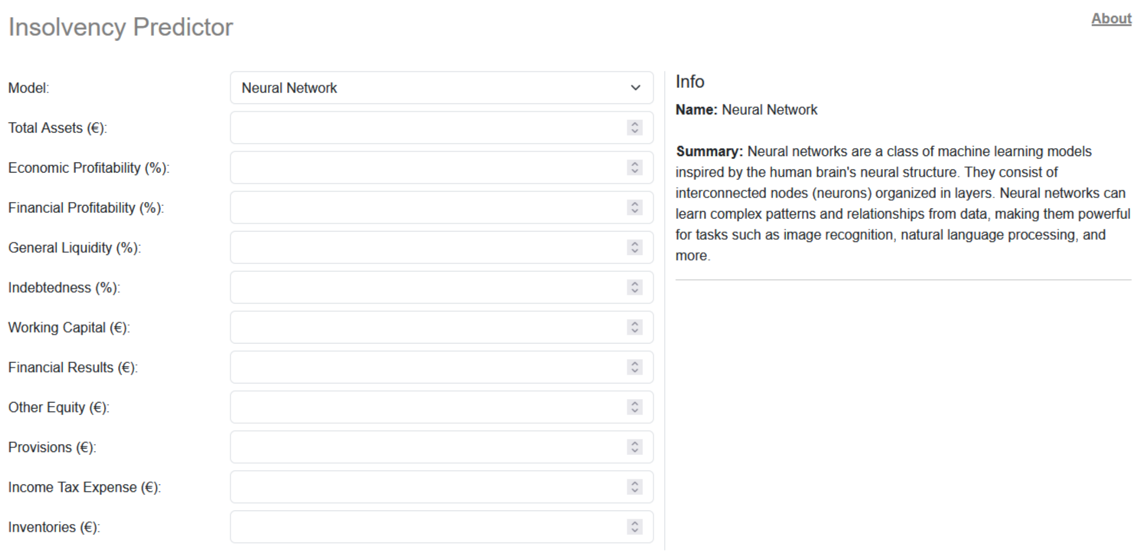

Figure 3: Insolvency Predictor - Model info

Financial indicator description: If the user wants to know more about one of the financial indicators, he can hover the mouse over the label of the financial indicator to obtain a brief description of it, as a tooltip. In the following example shown in Figure 4, the user can find a brief description of the General Liquidity indicator.

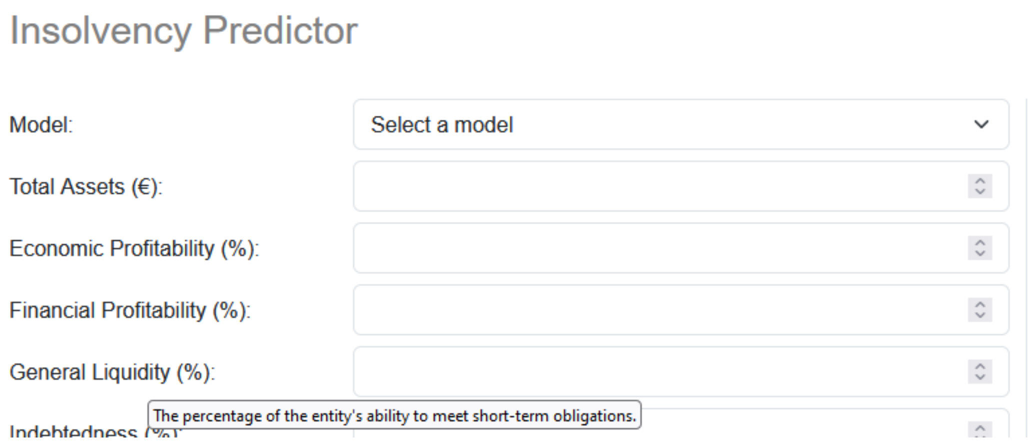

Figure 4: Insolvency Predictor - Financial indicator description

Best Model: In Figure 5 it is possible to see another useful and important functionality. The selection of the Best Model. When the user makes a prediction using this option, all the available models will be tested for that prediction, and the model with the best accuracy will be selected and used to make the final prediction. This can be useful when the user has no preference or idea about which model to select.

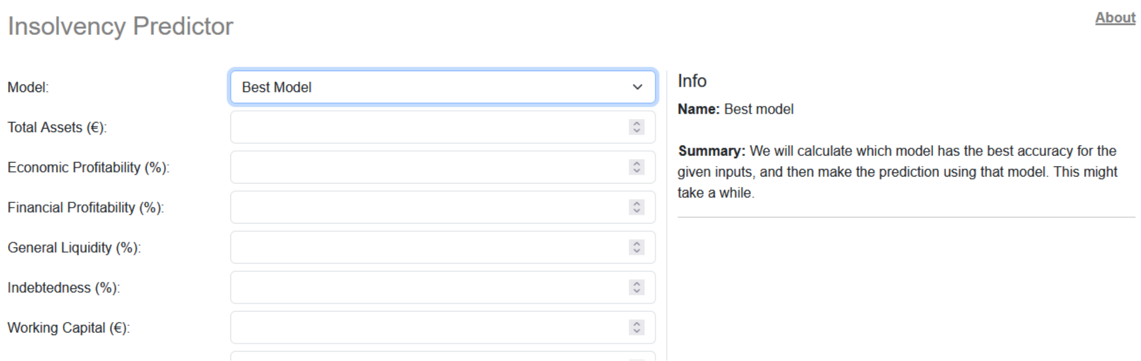

Figure 5: Insolvency Predictor - Best Model

3.2.2 Other Pages and components

On the application there are other pages and components available. They provide support for the main functionalities, and tie up the application. Here's a brief description them.

Model Page: The model page was designed to provide the user with insights into the different models offered by the application. Besides a description of each model (on the left side), the user can see the different performance benchmark metric scores, as explained in the methodology chapter. Also, when the user hovers the mouse over the labels of each metric, a description of each one is displayed as a tooltip, as observed in Figure 6.

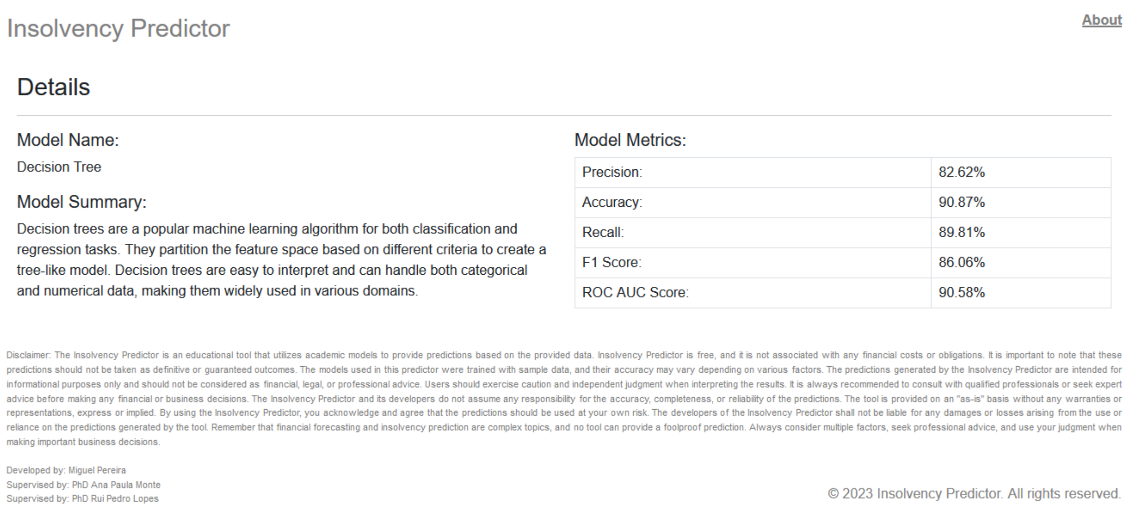

Figure 6: Insolvency Predictor - Model details page

About Page: The about page on Figure 7 is simple and offers general information about the application and the database used (on the left side). It also displays the different models available in the application, providing a link that redirects users to the Model page explained previously.

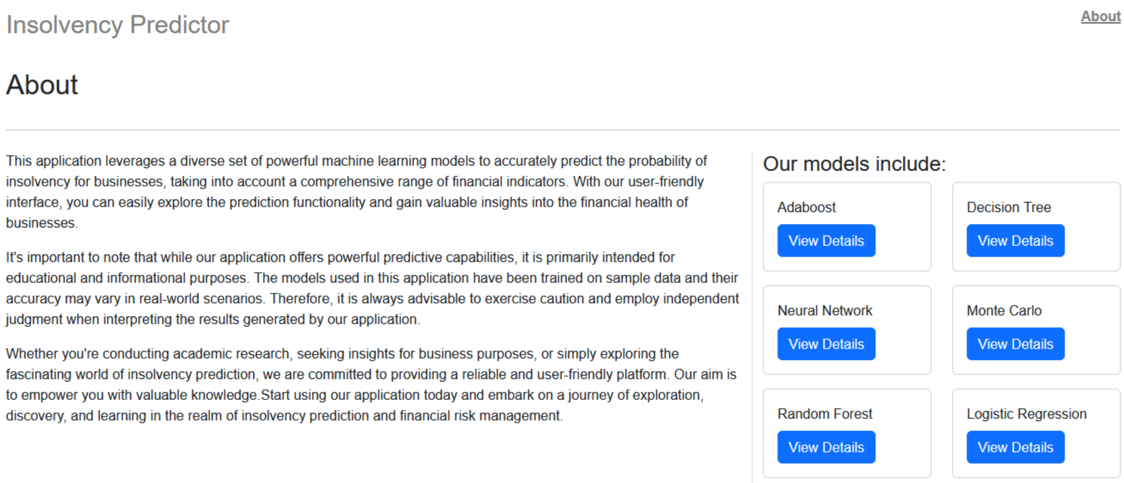

Figure 7: Insolvency Predictor - About page

Header and Footer: The header and footer are present in all the application, they have the function to define the style and design of the website and also some other functions. The header in Figure 8 contains a link to the Main page, and another to the about page. The footer contains a disclaimer message, the names of the developer and supervisors of this project and also some copyright information, as shown on Figure 9.

Figure 8: Insolvency Predictor - Header

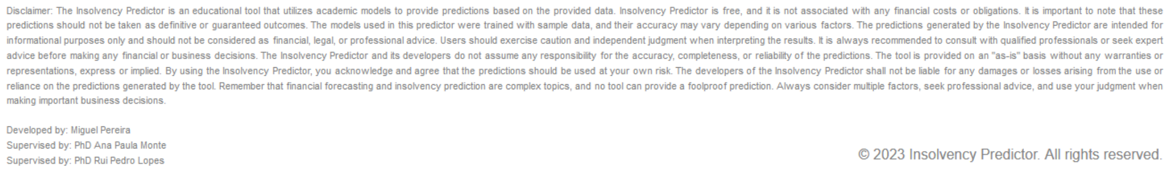

Figure 9: Insolvency Predictor - Footer

3.2.3 Reflective synthesis

The development of this application was a journey during which the researchers acquired a lot of knowledge. From a technological perspective, it was something entirely different from what the main developer usually program in his day-to-day work as a software developer in a fintech company, using different technologies and approaches. On the theoretical side, it was also very exciting as a master's student who had to grasp concepts related to finance, risk, investment, and insolvency.

The initial steps in developing this application involved gathering a usable dataset. To accomplish that we used the SABI database, as mentioned earlier, and extracted the maximum number of randomly selected portuguese companies with all the financial indicators provided by the database. From there, the data was processed, and the financial indicators used for the application were selected, as explained previously.

The next part was developing a console application that could make insolvency predictions on the dataset using the LR model. As soon as that was achieved, it was important to implement different models to validate the results. Next, NNs and MC model were added, as they were studied in the theoretical framework and proposed for this project. Subsequently, research was conducted to find more usable models, culminating in the inclusion of binary machine learning prediction models like RF, DT, and Adaboost.

The Riskit: Insolvency Predictor application evolved from a console application to a Python Flask web-based application with a working API to provide model predictions and other functionalities. As the main goal of the application is research, we decided to keep the UI/UX of the application simple and sober. To accomplish that, HTML, CSS, JS, and Bootstrap were used. Apart from the main page development, an about page was also added, as it is a standard practice in web-based applications, and a model detail page, where the user could gain more knowledge about the models and their performance. As delivering something valuable was the priority, some other functionalities were then developed, including the "Best Model", input prediction, model information and tooltips for the financial indicators and performance benchmarks. We believe that the development of these functionalities really tied the application together and increased the value that researchers can extract from it. As always, there is room for improvement, and some other useful functionalities can be added in the future. We will reflect and write about that in the conclusion of this work.

Conclusion, Limitations and Future Research Lines

In terms of what this work proposed to do, we can assume that it fulfilled its main objective, which was the development of an application that would provide students, researchers, managers, investors and analysts a tool they could utilize for their work. The application is able to make predictions about companies insolvency using the financial indicators and basing that prediction on the trained dataset explained earlier. We can conclude that Riskit: Insolvency Predictor has what we can consider to be a good performance and accuracy even when the user inputs a small number of indicators, proving that it is a solid tool. This work also proposed to benchmark the performance of various models in insolvency prediction, from that it was possible to conclude that the models that use ensembles (RF, DT and Adaboost) seem to have a slight better performance when compared to LR, NN and MC, even though in most cases, all models present good performance in benchmark scores.

It is also important to identify some limitations of the application. First, it is important to note that the end result of the application was way different from what we initially idealized. The initial proposal of this project projected an application that would measure risk and profitability, and would also give a probability of a company going insolvent. After the start of the development, it was understood that this would be too much to achieve, so it was decided to do an application that would focus on predicting insolvency and giving the user useful tools and information for their work. In terms of the technology, there were also some changes, as the application does not have a real database; instead, it uses a trained dataset to make the predictions. This change was made because it was a solution that required less infrastructure, performed really well, and, in general, made more sense.

Due to the limitations mentioned above, there are some future enhancements we would like to propose: incorporating a probability percentage on the insolvency prediction to offer more nuanced insights. More models can be added, like Support Vector Machine (SVM), which can handle both linear and non-linear data, Gradient Boost Machine (GBM)s, which is an ensemble model that we concluded has very good performance benchmarks for this dataset, or K-Nearest Neighbors (KNN), which considers the nearest data points to make its prediction (similar to the Euclidean distance explained earlier). Other types of predictions can be made, like estimating the profitability and risk of a company, which would help the users in risk management and investment strategy. Also, in order to upgrade the application's applicability, it would also be interesting if the dataset could have information about companies worldwide, as having companies only from one country can affect the non-financial factors when evaluating the economic and financial performance of companies, as mentioned in the theoretical framework. Lastly, although we are proud of the simple and intuitive UI/UX the application has, it could use an upgrade.

Bibliography

Altman, E. I., & Hotchkiss, E. (2005). *Corporate Financial Distress and Bankruptcy* (3rd ed.). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118267806

Álvares, P. (2019). *Modelos de previsão de falência empresarial análise crítica do z score de altman*.

Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions On Neural Networks*, *12*, 929–935.

Bank of Portugal - BdP. (2007). *Modelo de Avaliação de Riscos* (tech. rep.). Bank of Portugal.

Barros, C. (2000). *Decisões de Investimento e Financiamento de Projetos* (3rd ed.). Edições Sílabo.

Beck, U., Giddens, A., & Lash, S. (2000). *Modernização Reflexiva: Política, Tradição e Estética no Mundo Moderno* (1st ed.). Celta Editora.

Bogdan, S., & Bareša, S. (2021). Predicting bankruptcy based on the full population of croatian companies. *Ekonomski Pregled*, *72*, 643–669.

Carneiro, B. P. (2017). *Análise de projetos de investimento: Elaboração de uma ferramenta otimizada de avaliação económico-financeira*.

Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow : Concepts, tools, and techniques to build intelligent systems* (R. Roumeliotis & N. Tache, Eds.; 2nd). O'Reilly Media, Inc.

Giddens, A. (2000). *As Consequências da Modernidade* (4th ed.). Celta Editora.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, *9*, 5–34.

Horváthová, J., & Mokrišová, M. (2020). Comparison of the results of a data envelopment analysis model and logit model in assessing business financial health. *Information (Switzerland)*, *11*(3). https://doi.org/10.3390/info11030160

Illing, M., & Aaron, M. (2005). *A brief survey of risk-appetite indexes*. Bank of Canada, Financial System Review.

Janekova, J., Fabianova, J., & Kadarova, J. (2021). Selection of optimal investment variant based on monte carlo simulations. *International Journal of Simulation Modelling*, *20*(2), 279–290. https://doi.org/10.2507/IJSIMM20-2-557

Kiseleva, I. A., Karmanov, M. V., Korotkov, A. V., Kuznetsov, V. I., & Gasparian, M. S. (2018). Risk management in business: Concept, types, evaluation criteria. *Espacios*, *39*(27).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*.

Li, H., Sun, J., & Sun, B. L. (2009). Financial distress prediction based on OR-CBR in the principle of k-nearest neighbors. *Expert Systems with Applications*, *36*(1), 643–659. https://doi.org/10.1016/j.eswa.2007.09.038

Martins, A., Cruz, I., Silva, P., & Gonçalves, P. (2006). *Manual de Gestão Financeira Empresarial* (2nd ed.). Escolar Editora.

Moutinho, N., & Lopes, M. D. S. (2010). The Role of Non-Financial Aspects in Project Appraisal-The Practice of Portuguese Firms. *World Finance Conference – WFC*. https://bibliotecadigital.ipb.pt/handle/10198/4835

Moutinho, N., & Mouta, H. (2013). Investments Projects: Evaluation Tools and Methods. *XVI Congresso da Associação Portuguesa de Investigação Operacional*, 1–4.

OECD Capital Market Review of Portugal 2020 Mobilising Portuguese Capital Markets for Investment and Growth (tech. rep.). (2020).

O'Reilly, T. (2007). What is web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, *65*, 17–387.

Pecini, A. (2018). From the platformization of the web to the platform society: The impact of digital mediation in sociablity and subjectivity. *VI Congresso Internacional de Comunicação e Cultura*.

Platon, V., & Constantinescu, A. (2014). Monte Carlo Method in Risk Analysis for Investment Projects. *Procedia Economics and Finance*, *15*, 393–400. https://doi.org/10.1016/s2212-5671(14)00463-8

Resti, A., & Sironi, A. (2007). *Risk managment and shareholders' value in banking*. John Wiley & Sons, Ltd. https://www.ptonline.com/articles/how-to-get-better-mfi-results

Romão, F. M. E. (2009). *Credit scoring e a previsão de falências no contexto de basileia ii*.

Schooley, D. K., & Worden, D. D. (2016). Perceived and Realized Risk Tolerance: Changes During the 2008 Financial Crisis. *Journal of Financial Counseling and Planning*, *27*(2), 265–276. https://doi.org/10.1891/1052-3073.27.2.265

Securities Markets Commission - CMVM. (2012). *Comissão do mercado de valores mobiliários - recomendações aos investidores*. CMVM. www.cmvm.pt.

Sewpersadh, N. S. (2022). An econometric analysis of financial distress determinants from an emerging economy governance perspective. *Cogent Economics and Finance*, *10*(1). https://doi.org/10.1080/23322039.2021.1978706

Simeone, O. (2017). A brief introduction to machine learning for engineers. http://arxiv.org/abs/1709.02840

Soares, I., Moreira, J., Pinho, C., & Couto, J. (2015). *Decisões de Investimento–Análise Financeira de Projetos*. http://www.silabo.pt/Conteudos/8063_PDF.pdf

Soares, J., Fernandes, A., Março, A., & Marques, J. (1999). *Avaliação de Projetos de Investimento na Ótica Empresarial* (1st ed.). Edições Sílabo.

Teixeira, S. (2022). *Gestão das Organizações* (4th ed.). Escolar Editora.

Tian, C., Niu, T., & Wei, W. (2023). Volatility index prediction based on a hybrid deep learning system with multi-objective optimization and mode decomposition. *Expert Systems with Applications*, *213*, 119184. https://doi.org/https://doi.org/10.1016/j.eswa.2022.119184

Voda, A. D., Dobrotă, G., Tîrcă, D. M., Dumitrascu, D. D., & Dobrotă, D. (2021). Corporate bankruptcy and insolvency prediction model. *Technological and Economic Development of Economy*, *27*(5), 1039–1056. https://doi.org/10.3846/tede.2021.15106

Wach, M., & Chomiak-Orsa, I. (2021). The application of predictive analysis in decision-making processes on the example of mining company's investments projects. *25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, 5058–5066.

Wu, Y., Gaunt, C., & Gray, S. (2010). A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting and Economics*, *6*(1), 34–45. https://doi.org/10.1016/j.jcae.2010.04.002

Yang, Z. R., Platt, M. B., & Platt, H. D. (1999). Probabilistic Neural Networks in Bankruptcy Prediction. *Journal of Business Research*, *44*, 67–74.

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1997). *Theory and Methodology Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis* (tech. rep.).

Appendix

Table A1: Financial indicators used in the investment appraisal

Indicators that do not consider the time value of money
Return On Investment (ROI) is a criterion strictly based on forecast accounting statements, ignoring the updating of cash flows, based on accounting profit and, as such, not comparable with the initial investment.
Payback Period (PP) a criterion that represents the number of years necessary for the estimated operational cash-flows to “recover” the initial investment.
Critical Investment Time (CIT), is a criteria that seeks to overcome part of the limitations of the PP, being calculated taking into account the current value of the operating cash flows generated, at an appropriate discount rate, which "recover" the initial investment.
Indicators that consider the time value of money
Net Present Value (NPV) compares the current value of the cash flows generated by the investment project with the investment made. Projects with NPV equal to or greater than zero must be accepted (the investment will contribute to increasing the wealth of capital holders).
Adjusted Present Value (APV) seeks to overcome the limitations of the weighted average cost of capital as a discount rate for cash flows. Projects that present an APV greater than zero are economically viable, so they are to be accepted. This method is theoretically more flexible in terms of incorporating costs and income inherent to the existence of debt. Furthermore, by separating the value of tax savings, it allows the analyst to have a relatively in-depth perception of the impact of the capital structure on the analysis. The main disadvantage of this method is its operationalization.
Internal Rate of Return (IRR) represents the discount rate that equals the NPV of the project to zero, that is, it is the rate at which the present value of the cash flows generated by the investment project equals the current investment value. This method has as its main advantage, compared to the others already presented, the fact that it seeks to arrive at a single value that summarizes the validity of the project and which will be intrinsic to it, and as such independent of exogenous variables such as the update rate practiced in the market (which will be variable). Disadvantages/limitations: assumes the reinvestment of cash flows generated at the same rate as the IRR; it is not informative of the size and lifetime of the project; does not distinguish situations of financial investment from financing situations (does not apply in situations where cash inflow precedes cash outflow. When positive and negative cash flows alternate, multiple IRR can be obtained); in cases of mutually exclusive (i.e., alternative) projects, the IRR cannot be used. In these cases, the NPV should be used, when projects have different investment amounts or different timings for generating cash flows (the IRR benefits projects with lower investment and quick value generation).
Modified Internal Rate of Return (MIRR) is a complementary criterion to the IRR that aims to remedy the main deficiency of the IRR criterion (the reinvestment rate) and take part of its advantages only. Projects with TIRM equal to or greater than r (opportunity cost of capital) must be accepted. This method has the following main advantages, compared to the others already presented: it is easier to calculate, as it does not involve an iterative process; useful in cases where there are multiple IRR for a given set of flows, as it allows associating a single measure of profitability with such a set.
Profitability Index (PI) is a criterion in which projects that present APV greater than one are economically viable, therefore they are to be accepted (since they present a positive NPV). This method is also known as the Benefit/Cost Ratio, reflecting the idea that expected positive cash flows in the future are benefits while the initial investment is a cost.

Continued on next page

Table A1 (continuation)

Indicators that consider the time value of money
Project Duration (PD) and Successive NPV Profile are criteria that allow analyzing the sensitivity of the project to variations in the discount rate. This assumes that it is possible to divest at the end of each of the years of the estimated useful life of the investment and calculate, at that point, the respective NPV (estimating the residual value of the project in each period). The technique of analyzing successive NPV provides a progressive view, allowing the analysis of the project's sensitivity to variations in the NPV over the various periods of the investment, making it possible to determine the period in which the NPV is maximum. The maximum NPV will correspond to the optimal useful life of the project.
Source: Barros (2000), Martins et al. (2006), and J. Soares et al. (1999).

Table A2: Performance benchmarks

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC Score	Confusion Matrix
Logistic Regression	0.881	0.7627	0.8726	0.8140	0.8786	[[1862, 243], [114, 781]]
Decision Tree	0.9087	0.8262	0.8981	0.8606	0.9058	[[1880, 178], [96, 846]]
Random Forest	0.9273	0.8721	0.9112	0.8912	0.9232	[[1889, 131], [87, 893]]
Neural Network	0.881	0.7500	0.8838	0.8114	0.8818	[[1875, 256], [101, 768]]
Monte Carlo	0.832	0.574	0.897	0.7	0.856	[[1909, 436], [67 , 588]]
Adaboost	0.911	0.845	0.888	0.866	0.905	[[1867, 159], [109, 865]]

Source: Author's own elaboration.

Table A3: Companies by Economic Activity Code

	n	%
Wholesale trade	2685	17.9
Retail trade	1381	9.2
Real estate promotion	986	6.6
Real estate activities	657	4.4
Trade, maintenance and repair of vehicles	549	3.7
Food industries	509	3.4
Manufacture of metal products, except machinery and equipment	455	3.0
Land transport and transport via pipelines	445	3.0
Specialized construction activities	389	2.6
Financial services activities	314	2.1
Restaurants and similar establishments	308	2.1
Clothing industry	307	2.0
Head offices and management consultancy activities	297	2.0
Civil engineering	281	1.9
Human health activities	273	1.8
Warehousing and support activities for transportation	254	1.7
Textile manufacturing	246	1.6
Consultancy and computer programming and related activities	232	1.5
Manufacture of other non-metallic mineral products	220	1.5
Agriculture, livestock production, hunting, and related activities	208	1.4
Accommodation	201	1.3
Manufacture of rubber and plastic products	191	1.3
Leather and leather products industry	189	1.3
Office administrative, office support, and other business support activities	166	1.1
Wood and cork industries and their products	165	1.1
Electricity, gas, steam, hot and cold water supply	154	1.0
Architectural and engineering activities and related technical consultancy	150	1.0
Manufacture of machinery and equipment	148	1.0
Manufacture of chemicals and synthetic fibers	135	0.9
Manufacture of furniture and mattresses	133	0.9
Beverage industry	126	0.8
Manufacture of motor vehicles	106	0.7
Collection, treatment, and disposal of waste; materials recovery	105	0.7
Education	104	0.7
Sports, entertainment, and recreational activities	96	0.6
Manufacture of pulp, paper, paperboard, and articles thereof	87	0.6
Rental and leasing activities	86	0.6
Repair, maintenance, and installation of machinery and equipment	81	0.5
Advertising, market research, and public opinion polling	81	0.5
Employment activities	76	0.5
Building completion and maintenance, landscaping activities	75	0.5
Manufacture of electrical equipment	67	0.4
Printing and reproduction of recorded media	66	0.4

Continued on next page

Table A3 (continuation)

	n	%
Basic metals industries	64	0.4
Travel agency, tour operator, and other reservation service activities	63	0.4
Other consultancy, scientific, technical, and related activities	60	0.4
Publishing activities	57	0.4
Water collection, treatment, and supply	55	0.4
Other personal service activities	55	0.4
Other manufacturing industries	48	0.3
Activities auxiliary to financial services and insurance activities	47	0.3
Social work activities with accommodation	47	0.3
Research and security activities	45	0.3
Telecommunications	42	0.3
Other extractive industries	41	0.3
Manufacture of pharmaceutical products	41	0.3
Information service activities	41	0.3
Legal and accounting activities	38	0.3
Motion picture, video, television, and music production activities	37	0.2
Water transport	35	0.2
Manufacture of computer, electronic, and optical products	34	0.2
Public administration and defense; compulsory social security	34	0.2
Manufacture of other transport equipment	33	0.2
Air transport	31	0.2
Insurance, reinsurance, and pension funding, except compulsory social security	30	0.2
Forestry and logging	28	0.2
Scientific research and development activities	26	0.2
Social work activities without accommodation	24	0.2
Activities of other membership organizations	23	0.2
Theater, music, dance, and other artistic and literary activities	18	0.1
Lotteries and other gambling activities	18	0.1
Postal and courier activities	16	0.1
Radio and television activities	16	0.1
Repair of computers and personal and household goods	16	0.1
Fishing and aquaculture	15	0.1
Collection, treatment, and disposal of wastewater	11	0.1
Library, archives, museums, and other cultural activities	9	0.1
Manufacture of coke, refined petroleum products	6	0.0
Extraction and agglomeration of metal ores	4	0.0
Tobacco industry	3	0.0
Services activities related to extractive industries	2	0.0
Veterinary activities	1	0.0
Total	15000	100,0

Source: Author's own elaboration.

Table A4: Average scores of companies (active vs insolvent)

		Averages	Welch t-test (p)
	Active	36425.1	
Operating Revenues	Insolvent	1842.8	<0.001
	Total	25144.3	
	Active	2902.7	
Current Results	Insolvent	-387.9	<0.001
	Total	1829.3	
	Active	2465.7	
Net Income for the Period	Insolvent	-393.3	<0.001
	Total	1533.1	
	Active	117090.4	
Total Assets	Insolvent	4704.4	<0.001
	Total	80430.1	
	Active	26056.3	
Equity	Insolvent	-363.1	<0.001
	Total	17438.3	
	Active	113.0	
Economic Profit%	Insolvent	-238.7	0.049
	Total	-1.7	
	Active	19.9	
Financial Profit%	Insolvent	-16.1	0.023
	Total	8.2	
	Active	19.4	
General Liquidity%	Insolvent	15.0	0.580
	Total	18.0	
	Active	122.1	
Indebtedness%	Insolvent	2952.1	0.135
	Total	1045.3	
	Active	155.5	
Number of Employees	Insolvent	21.1	<0.001
	Total	111.6	
	Active	3973.3	
Working Capital	Insolvent	1090.9	<0.001
	Total	3033.1	
	Active	2775.0	
Operating Results	Insolvent	-290.8	<0.001
	Total	1775.0	
	Active	29598.7	
Turnover	Insolvent	1770.8	<0.001
	Total	20521.3	

Continued on next page

Table A4 (continuation)

		Averages	Welch t-test (p)
Financial Revenues and Gains	Active	343.3	0.030
	Insolvent	19.8	
	Total	237.8	
Financial Costs and Losses	Active	607.9	<0.001
	Insolvent	116.9	
	Total	447.7	
Financial Results	Active	-264.6	0.222
	Insolvent	-97.1	
	Total	-210.0	
Income Tax	Active	308.6	<0.001
	Insolvent	12.3	
	Total	211.9	
Net Current Results after Taxes	Active	2201.9	<0.001
	Insolvent	-400.1	
	Total	1353.1	
Extraordinary Revenues and Gains	Active	47.7	0.367
	Insolvent	16.1	
	Total	37.4	
Extraordinary Costs and Losses	Active	82.3	0.224
	Insolvent	10.0	
	Total	58.7	
Extraordinary Results	Active	-32.6	0.144
	Insolvent	6.1	
	Total	-20.0	
Cost of Goods Sold and Consumed Materials	Active	17241.6	<0.001
	Insolvent	877.7	
	Total	11903.6	
Personnel Costs	Active	4082.7	<0.001
	Insolvent	373.1	
	Total	2872.6	
Depreciation of the Period	Active	1381.5	<0.001
	Insolvent	77.0	
	Total	956.0	
Other Operating Items	Active	7046.4	<0.001
	Insolvent	574.3	
	Total	4935.2	
Interest Expenses	Active	435.3	<0.001
	Insolvent	100.7	
	Total	326.1	

Continued on next page

Table A4 (continuation)

		Averages	Welch t-test (p)
Cash Flows	Active	3847.2	<0.001
	Insolvent	-316.2	
	Total	2489.1	
Gross Value Added	Active	8673.7	<0.001
	Insolvent	169.8	
	Total	5899.8	
EBIT	Active	2775.0	<0.001
	Insolvent	-290.8	
	Total	1775.0	
EBITDA	Active	4156.5	<0.001
	Insolvent	-213.7	
	Total	2731.0	
Fixed Assets	Active	32323.2	<0.001
	Insolvent	2036.0	
	Total	22443.5	
Intangible Fixed Assets	Active	6095.6	0.006
	Insolvent	101.7	
	Total	4140.3	
Tangible Fixed Assets	Active	8571.7	<0.001
	Insolvent	882.4	
	Total	6063.4	
Other Fixed Assets	Active	17656.0	<0.001
	Insolvent	1052.0	
	Total	12239.8	
Current Assets	Active	21124.2	<0.001
	Insolvent	2669.1	
	Total	15104.1	
Inventories	Active	3327.8	<0.001
	Insolvent	1052.6	
	Total	2585.7	
Liabilities to Third Parties	Active	5087.3	<0.001
	Insolvent	845.6	
	Total	3703.6	
Other Current Assets	Active	12709.1	<0.001
	Insolvent	770.8	
	Total	8814.8	
Bank Deposits and Cash	Active	4166.6	<0.001
	Insolvent	110.8	
	Total	2843.6	

Continued on next page

Table A4 (continuation)

		Averages	Welch t-test (p)
Capital	Active	8970.8	<0.001
	Insolvent	956.2	
	Total	6356.4	
Other Shareholders' Equity	Active	12859.9	<0.001
	Insolvent	-1319.3	
	Total	8234.6	
Long-Term Liabilities	Active	14188.9	<0.001
	Insolvent	2005.7	
	Total	10214.8	
Liabilities to Third Parties (Long-Term)	Active	9728.3	<0.001
	Insolvent	1650.3	
	Total	7093.2	
Other Long-Term Liabilities	Active	4460.2	0.002
	Insolvent	355.4	
	Total	3121.2	
Provisions	Active	863.1	<0.001
	Insolvent	109.3	
	Total	617.2	
Short-Term Liabilities	Active	17427.8	<0.001
	Insolvent	3061.8	
	Total	12741.6	
Short-Term Financial Liabilities	Active	5283.5	<0.001
	Insolvent	1102.3	
	Total	3919.6	
Other Short-Term Liabilities	Active	4441.8	<0.001
	Insolvent	807.3	
	Total	3256.2	
Other Short-Term Liabilities	Active	7709.6	<0.001
	Insolvent	1158.0	
	Total	5572.5	
Total Equity and Liabilities	Active	53447.4	<0.001
	Insolvent	4704.4	
	Total	37547.5	

Source: Author's own elaboration.

Table A5: Measures of association between companies' Economic Activity Code and status

		Active	Insolvent
Agriculture, animal production, hunting, and related service activities	Count	138	70
	Adjusted Residues	-,3	,3
Forestry and logging	Count	19	9
	Adjusted Residues	,1	-,1
Fishing and aquaculture	Count	10	5
	Adjusted Residues	-,1	,1
Extraction and preparation of metallic minerals	Count	3	1
	Adjusted Residues	,3	-,3
Other extractive industries	Count	28	13
	Adjusted Residues	,1	-,1
Activities related to extractive industries	Count	2	0
	Adjusted Residues	1,0	-1,0
Food industries	Count	373	136
	Adjusted Residues	2,9	-2,9
Beverage industry	Count	105	21
	Adjusted Residues	3,8	-3,8
Tobacco industry	Count	3	0
	Adjusted Residues	1,2	-1,2
Manufacture of textiles	Count	172	74
	Adjusted Residues	,9	-,9
Clothing industry	Count	163	144
	Adjusted Residues	-5,4	5,4
Leather and leather product industry	Count	110	79
	Adjusted Residues	-2,7	2,7
Wood and cork industries and their products	Count	106	59
	Adjusted Residues	-,9	,9
Manufacture of pulp, paper, paperboard, and articles	Count	77	10
	Adjusted Residues	4,2	-4,2
Printing and reproduction of recorded media	Count	25	41
	Adjusted Residues	-5,1	5,1
Manufacture of coke, refined petroleum products	Count	6	0
	Adjusted Residues	1,7	-1,7
Manufacture of chemicals and synthetic fibers	Count	117	18
	Adjusted Residues	4,8	-4,8
Manufacture of pharmaceutical products	Count	36	5
	Adjusted Residues	2,8	-2,8
Manufacture of rubber and plastic products	Count	158	33
	Adjusted Residues	4,6	-4,6
Manufacture of other non-metallic mineral products	Count	167	53
	Adjusted Residues	2,7	-2,7
Basic metal industries	Count	57	7
	Adjusted Residues	3,7	-3,7

Continued on next page

Table A5 (continuation)

		Active	Insolvent
Manufacture of fabricated metal products, except machinery and equipment	Count	293	162
	Adjusted Residues	-1,4	1,4
Manufacture of computer, electronic, and optical products	Count	31	3
	Adjusted Residues	3,0	-3,0
Manufacture of electrical equipment	Count	53	14
	Adjusted Residues	2,1	-2,1
Manufacture of machinery and equipment	Count	119	29
	Adjusted Residues	3,4	-3,4
Manufacture of motor vehicles	Count	94	12
	Adjusted Residues	4,7	-4,7
Manufacture of other transport equipment	Count	27	6
	Adjusted Residues	1,8	-1,8
Manufacture of furniture and mattresses	Count	63	70
	Adjusted Residues	-4,9	4,9
Other manufacturing industries	Count	35	13
	Adjusted Residues	,8	-,8
Repair, maintenance, and installation of machinery and equipment	Count	51	30
	Adjusted Residues	-,9	,9
Electricity, gas, steam, hot and cold water, and air conditioning	Count	148	6
	Adjusted Residues	7,6	-7,6
Water collection, treatment, and distribution	Count	55	0
	Adjusted Residues	5,2	-5,2
Collection, treatment, and disposal of wastewater; waste material recovery	Count	10	1
	Adjusted Residues	1,7	-1,7
Collection, treatment, and elimination of waste; material recovery	Count	93	12
	Adjusted Residues	4,6	-4,6
Real estate activities	Count	305	681
	Adjusted Residues	-25,3	25,3
Civil engineering	Count	162	119
	Adjusted Residues	-3,5	3,5
Specialized construction activities	Count	168	221
	Adjusted Residues	-10,3	10,3
Trade, maintenance, and repair of motor vehicles and motorcycles	Count	406	143
	Adjusted Residues	3.3	-3.3
Wholesale trade	Count	2167	518
	Adjusted Residues	16.3	-16.3
Retail trade	Count	934	447
	Adjusted Residues	0.2	-0.2
Land transport and transport via pipelines or gas pipelines	Count	243	202
	Adjusted Residues	-5.8	5.8
Water transport	Count	31	4
	Adjusted Residues	2.7	-2.7

Continued on next page

Table A5 (continuation)

		Active	Insolvent
Air transport	Count	27	4
	Adjusted Residues	2.3	-2.3
Warehousing and support activities for transportation	Count	221	33
	Adjusted Residues	6.7	-6.7
Postal and courier activities	Count	13	3
	Adjusted Residues	1.2	-1.2
Accommodation	Count	132	69
	Adjusted Residues	-0.5	0.5
Food and beverage service activities	Count	50	258
	Adjusted Residues	-19.3	19.3
Publishing activities	Count	32	25
	Adjusted Residues	-1.8	1.8
Motion picture, video, television program production activities	Count	21	16
	Adjusted Residues	-1.4	1.4
Radio and television activities	Count	15	1
	Adjusted Residues	2.3	-2.3
Telecommunications	Count	38	4
	Adjusted Residues	3.2	-3.2
Computer programming and related activities	Count	210	22
	Adjusted Residues	7.6	-7.6
Information service activities	Count	33	8
	Adjusted Residues	1.8	-1.8
Financial service activities	Count	279	35
	Adjusted Residues	8.2	-8.2
Insurance, reinsurance, and pension funds	Count	30	0
	Adjusted Residues	3.8	-3.8
Activities auxiliary to financial services and insurance activities	Count	34	13
	Adjusted Residues	0.7	-0.7
Real estate activities	Count	350	307
	Adjusted Residues	-7.9	7.9
Legal and accounting activities	Count	22	16
	Adjusted Residues	-1.2	1.2
Head office activities; management consultancy activities	Count	244	53
	Adjusted Residues	5.5	-5.5
Architectural and engineering activities; related technical consulting	Count	84	66
	Adjusted Residues	-3.0	3.0
Research and experimental development	Count	25	1
	Adjusted Residues	3.1	-3.1
Advertising, market research, public opinion polling	Count	49	32
	Adjusted Residues	-1.3	1.3
Other professional, scientific, and technical activities	Count	39	21
	Adjusted Residues	-0.4	0.4

Continued on next page

Table A5 (continuation)

		Active	Insolvent
Veterinary activities	Count	1	0
	Adjusted Residues	0.7	-0.7
Rental and leasing activities	Count	67	19
	Adjusted Residues	2.1	-2.1
Employment activities	Count	57	19
	Adjusted Residues	1.4	-1.4
Travel agency, tour operator activities	Count	38	25
	Adjusted Residues	-1.2	1.2
Investigation and security activities	Count	24	21
	Adjusted Residues	-2.0	2.0
Activities related to buildings, landscape services	Count	37	38
	Adjusted Residues	-3.3	3.3
Office administrative, office support services	Count	119	47
	Adjusted Residues	1.2	-1.2
Public administration, defense, compulsory social security	Count	33	1
	Adjusted Residues	3.7	-3.7
Education	Count	49	55
	Adjusted Residues	-4.4	4.4
Human health activities	Count	206	67
	Adjusted Residues	2.9	-2.9
Social work activities with accommodation	Count	35	12
	Adjusted Residues	1.0	-1.0
Social work activities without accommodation	Count	16	8
	Adjusted Residues	-0.1	0.1
Theatre, music, dance, and other arts activities	Count	10	8
	Adjusted Residues	-1.1	1.1
Library, archives, museums, and other cultural activities	Count	8	1
	Adjusted Residues	1.4	-1.4
Lotteries and betting activities	Count	15	3
	Adjusted Residues	1.4	-1.4
Sports, entertainment, and recreation activities	Count	41	55
	Adjusted Residues	-5.2	5.2
Activities of other membership organizations	Count	22	1
	Adjusted Residues	2.9	-2.9
Repair of computers and personal and household goods	Count	8	8
	Adjusted Residues	-1.5	1.5
Other personal service activities	Count	9	46
	Adjusted Residues	-8.1	8.1
	Count	10106	4892

Source: Author's own elaboration.

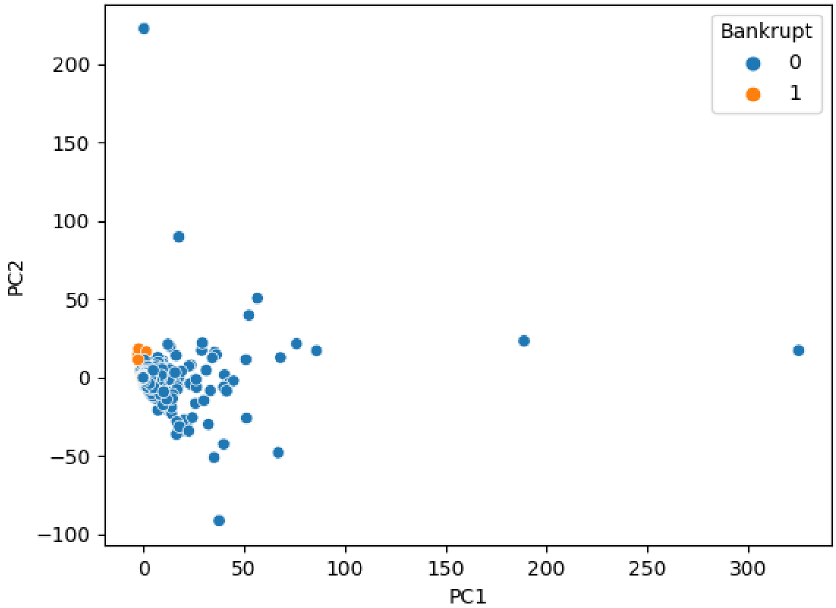

Figure A1: Principal Component Analysis

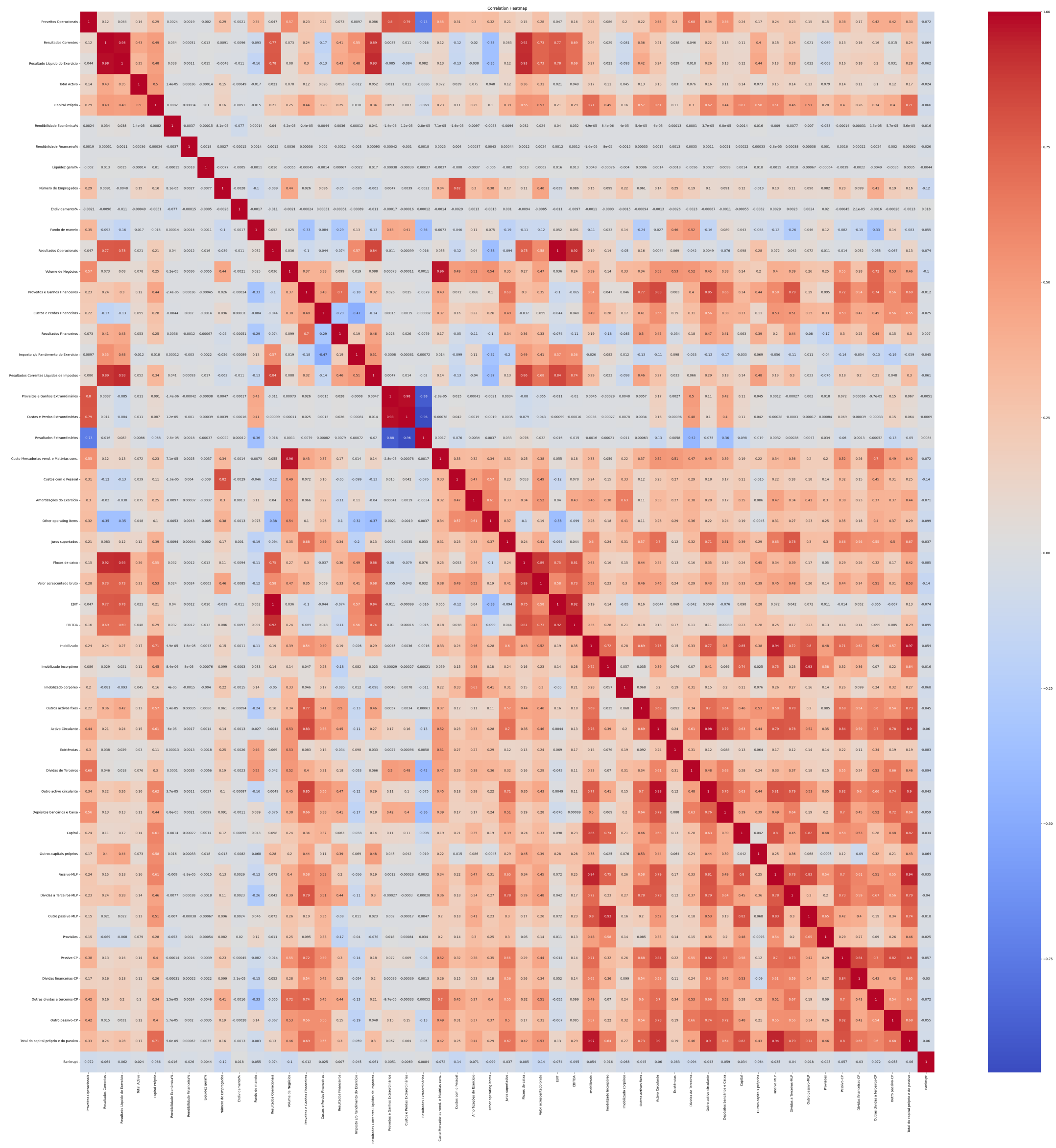

Figure A2: Exploratory Data Analisys (Link)