Computational Economics
https://doi.org/10.1007/s10614-024-10709-y

Ensemble Methods for Bankruptcy Resolution Prediction: A New Approach

Agustín J. Sánchez-Medina[1] · Félix Blázquez-Santana[1] · Daniel L. Cerviño-Cortínez[2,1] · Mónica Pellejero[1]

Accepted: 30 August 2024

Abstract
When a company goes bankrupt, it generates an extremely important uncertainty for all stakeholders as to whether the company will be reorganized or liquidated. This study aims to provide a successful methodology to predict whether a bankrupt SME will reorganize or liquidate. This could prevent significant economic and social losses and would contribute to reduce the number of SMEs that are helped to reorganize when they have little chance of success or that are liquidated when they could be viable. This useful and valid methodology applies algorithms (e.g., k-nearest neighbors) and techniques of ensemble learning and performance evaluation algorithms for the first time, considering the reviewed literature. By applying this methodology, it is possible to achieve a performance far superior to that known in the literature, specifically with an average accuracy of 94 percent using a data set with only financial variables of 1683 Spanish SMEs in the period 2011–2019.

Keywords Bankruptcy · Reorganization · Prediction · Artificial intelligence · Ensemble learning

Mathematics Subject Classification G33 · G34

✉ Daniel L. Cerviño-Cortínez
daniel.cervino@pdi.atlanticomedio.es

Agustín J. Sánchez-Medina
agustin.sanchez@ulpgc.es

Félix Blázquez-Santana
felix.blazquez@ulpgc.es

Mónica Pellejero
monica.pellejero@ulpgc.es

1 University of Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas de Gran Canaria, Spain

2 Universidad del Atlántico Medio, Carretera de Quilmes, 37, 35017 Tafira Baja, Las Palmas de Gran Canaria, Spain

Published online: 08 January 2025

1 Introduction

Bankruptcy as a representation is associated with substantial economic losses for investors, government agencies, employees, and society in general (Bernstein et al., 2019; Gestel et al., 2006; Wu, 2010). In this regard, it is remarkable that, although creditors are among the stakeholders affected by this situation (Jabeur et al., 2021), the loss of employment caused by bankruptcy generates such social alarm that some researchers question whether financial objectives should prevail over social problems derived from bankruptcy (Blazy et al., 2011). In addition to these negative effects, the judicial process of bankruptcy generates fear of several outcomes, including loss of reputation, business failure, loss of business partner, economic maintenance of insolvent companies, and loss of jobs (Prusak et al., 2019).

When a company enters bankruptcy, it faces a judicial resolution process that is justified by the conflict of interest among debtors, creditors, and equity holders of the company (Fisher & Martel, 2009). This conflict can result in the liquidation of viable companies or the reorganization of nonviable ones in a way that generates large economic losses (Laitinen, 2011).

The structure of the laws governing the resolution processes is extremely important (Silva & Saito, 2020; Wang, 2012), where structures favorable to the debtor or creditor are found. The former favor the reorganization of unviable companies, while the latter increase the probability of premature liquidation of the company.

In this sense, it is crucial to be able to discriminate between those companies that, although bankrupt, may be economically efficient and those that are not (Eidenmüller, 2017). Thus, inefficient companies should be liquidated and their assets distributed among their creditors for more productive uses (Yu & He, 2018). If the reorganization were to fail, it would increase the damage to the creditors, and society to see their losses increase (He et al., 2020a, 2020b). In contrast, it would be essential to help those companies that show potential for their continuity, thus avoiding the loss of employment and the improvement of the economy in general (Bernstein et al., 2019; Blazy et al., 2011). Therefore, as stated by Antill (2022), and Wang (2012), it is necessary to propose a filtering mechanism in the resolution of bankruptcies that would ensure the liquidation of inefficient enterprises and the reorganization of those that are viable. Furthermore, this would help eliminate the uncertainty faced by both debtors and creditors about the future value of assets while reorganization plans are being discussed (Antill & Grenadier, 2019).

However, as stated by Barniv et al. (2002) and Camacho-Miñano et al. (2015), there are few studies that explore the post-bankruptcy scenario, an issue that would undoubtedly be of great interest to investors, creditors, and other stakeholders.

In view of this, this research focuses on identifying, sufficiently in advance, those companies that, despite having entered judicial proceedings for bankruptcy, have the capacity to continue with their activity or that should be liquidated.

This work contributes to the academic literature on the resolution of the bankruptcy process in the following ways: (a) It offers a useful, valid, and easily implementable methodology for prediction. It is based on artificial intelligence algorithms whose performance reaches 94 percent accuracy, a result considerably superior to that achieved in previous academic works. (b) Artificial intelligence techniques and algorithms such as ensemble learning, support vector machines, random forest, k-nearest neighbors, and neural networks are used for the first time. In addition, different techniques and metrics are used to evaluate the performance of algorithms such as stratified cross-validation based on k folds or grid search. (c) A data set consisting of accounting data from 1683 Spanish SMEs (1222 liquidated and 461 reorganized) is used. It is worth mentioning that the accounting ratios of this study are easily accessible to practitioners and academics because they are calculated using accounting information from highly standardized international databases. This facilitates and makes the proposed methodology more usable. The fact that an accurate result can be achieved with accounting data only is relevant because authors like Laitinen (2011) have stated that in this type of research, non-financial information is more informative than financial information. However, non-financial information can rarely be obtained directly, at least in the Spanish context, requiring in many cases a process of elaboration and contextualization that makes it difficult to generalize the results obtained. Moreover, achieving good results using only accounting data contradicts Barniv et al. (2002), who argued that a correct prediction of bankruptcy resolution could not be made using only this type of data.

A sample of Spanish small and medium-sized enterprises (SMEs) has been used for this study due to the important role that this type of company plays in the economy not only in Spain, but also in Europe. In the European Union they constitute 99% of the total number of companies. In comparison, in Spain they reach 99.8%, contributing 61% of Gross Value Added and 66% of employment (Ministerio de Industria Comercio y Turismo de España, 2021). It is worth mentioning that, due to the characteristics of this type of companies, they face significant challenges during the bankruptcy process since, unlike large companies, they have fewer resources and serious restrictions for access to the bank credit (De Blick et al., 2024; Du & Nguyen, 2022; Nicolas, 2022). This situation has recently been aggravated by significant inflationary pressures that have made access to this important source of financing even more difficult (OECD, 2024).

In addition, bankruptcy filings have intensified in the European Union in recent years, reaching their highest level since the start of data collection in 2015, with an 8.4% increase in the second quarter of 2023 over the previous quarter (Eurostat, 2023).

This study is particularly relevant because it provides a robust methodology for predicting bankruptcy resolution using accessible and standardized accounting data. The proposed methodology could be applied in all European Union member countries, thanks to the standardization of the accounting data elaborated under International Reporting Standards (IFRS), which ensures the comparability of the financial ratios used in the study. This allows the generalization and application of the methodology, not only in the member states of the European Union, but also in other countries that adopt or are aligned with IFRS.

2 Literature Review

Bankruptcy is a judicial process arising from the application, by creditors or debtors, of a company's or individual's insolvency to creditors (Rashid, 2019; White, 2016). Such a proceeding is resolved by liquidation, reorganization, or dismissal of the petition. Liquidation results in the total or partial sale of the company's assets to meet its debts, while reorganization allows the restructuring and survival of the company through a plan and prior agreement between debtors and creditors (Stef, 2022).

The bankruptcy resolution process will depend on the structure and design of the bankruptcy laws (Silva & Saito, 2020; Wang, 2012). Thus, a debtor-friendly structure allows unviable companies to reorganize (e.g., laws in the US, (Aguiar Díaz & Ruiz Mallorquí, 2015a)), while a creditor-friendly structure increases the likelihood of premature liquidation [e.g., laws in the UK, (Keasey et al., 2015)].

For their part, European countries have been adapting their bankruptcy codes to that of the US, giving more power to the entrepreneur to restructure his or her outstanding financial contracts and thus avoid the opening of a liquidation phase (Tarantino, 2013). Therefore, there is a clear trend for European legislators to seek to protect companies and employment (Blazy et al., 2011). However, their judicial process is considered more restrictive than those applied in the US, as in some cases they appoint a bankruptcy administrator to manage the entire reorganization or liquidation process. Spain, the focus of this study, takes an intermediate position between debtor-oriented and creditor-oriented systems (Aguiar Díaz & Ruiz Mallorquí, 2015a).

This work focuses on the moment when companies have already entered the legal process of bankruptcy and, therefore, a bankruptcy administrator has been assigned to manage the company until the final resolution. At the beginning of this process, the judge can make two decisions: the first one would be to liquidate the company if he detects that there is no possibility of an agreement with the creditors, while the second one would be to start a negotiation phase. At the end of this stage, the company will be reorganized (healthy) or liquidated depending on the future viability perceived by the creditors and the bankruptcy administrator. In this study, two types of companies have been considered: those that were liquidated after a judicial sentence, either at the beginning or at the end of the process, and those that the judge terminated the bankruptcy proceeding and, consequently, were reorganized.

In any case, as stated by Laitinen (2011), the objective of bankruptcy laws should be to improve the efficiency of the judicial proceedings, trying to reorganize only those companies that have real possibilities of continuing their activity. Additionally, as stated by Antill (2022), and Wang (2012), it is necessary to propose a filtering mechanism in the resolution of bankruptcies that would ensure the liquidation of inefficient enterprises and the reorganization of those that are viable.

The judicial process of bankruptcy and an incorrect decision-making process together generate a set of harmful negative effects for different stakeholders, such as creditors, debtors, employees, and society (Bernstein et al., 2019; Prusak et al., 2019). These negative effects include (a) high uncertainty about the resolution of the judicial process (Antill & Grenadier, 2019); (b) job losses (Bernstein et al., 2019; Blazy et al., 2011; Prusak et al., 2019); (c) economic losses arising from the value of the assets, the company's reputation, and the maintenance of an unviable company that should be liquidated (Prusak et al., 2019); and (d) loss of trust by business partners (Prusak et al., 2019).

Hence, it would be useful to develop methodologies capable of predicting those companies that, having entered bankruptcy, can successfully be reorganized, or liquidated. In this way, the judicial systems that regulate bankruptcy would have tools that would provide relevant information on the probabilities of reorganization or liquidation, thus increasing the efficiency of the decisions taken.

With this in mind, a methodology based on artificial intelligence algorithms is proposed that uses only easily obtainable accounting information. Such a methodology was evaluated using 1,683 Spanish SMEs during the period 2011–2019, achieving an average accuracy of 94 percent in its predictions.

In this area, most of the existing studies focus mainly on identifying those variables that have a significant impact on the reorganization or liquidation of bankrupt companies, for example, by analyzing (a) the composition of the company's assets (e.g., Casey et al., 1986; James, 2016; Rose-Green & Lovata, 2013); (b) the auditor's role in the resolution process (e.g., Casterella et al., 2000; Kim et al., 2008); (c) the efficiency and the role played by bankruptcy laws in relation to the country in which the company operates (e.g., Aguiar Díaz & Ruiz Mallorquí, 2015a; Blazy et al., 2011); (d) the relationship and type of debt existing between banks and corporations (e.g., Blazy et al., 2014; Demiroglu & James, 2015; Leyman et al., 2011); and (e) the role of the court and its characteristics within the bankruptcy resolution process (e.g., Blazy & Esquerré, 2021; Blazy et al., 2011).

However, after analyzing the literature and following Barniv et al. (2002) and Camacho-Miñano et al. (2015), the scarce number of academic papers on the prediction of bankruptcy resolution stands out. These studies can be found in Table 1, which have been extracted from platforms such as Web of Science and Scopus after searching for combinations of the following keywords: "bankruptcy", "prediction", "distress", "firms", "companies", "resolution", "emergence" and "liquidation".

Initially, the literature shows that classical statistical methods such as probit models based on regression have been applied (e.g., Campbell, 1996; Casey et al., 1986; Kennedy & Shaw, 1991), and subsequently focused on logistic regression (e.g., Barniv et al., 1997; Kim et al., 2008). More recently, more advanced artificial intelligence models that employ decision trees have been employed, such as PART (e.g., Camacho-Miñano et al., 2015), as well as classical methods such as the use of probit models (e.g., Gupta et al., 2022; Stef & Bissieux, 2022). It should be noted that although neural networks (e.g., Barniv et al., 1997; Kumar et al., 1997), and hybrid genetic classifier (e.g., Kumar et al., 1997), they have not been used again for this problem again.

The study that achieved the highest predictive ability was that of Gupta et al. (2022) with 0.916 of AUC-ROC. In addition, the work of Casey et al. (1986) is the only one with Kumar et al. (1997) that have used only financial variables, obtaining an accuracy of 70.8 percent and 69.73 percent respectively. Most studies have used US firms as a sample (e.g., Barniv et al., 2002; Campbell, 1996; Casey et al., 1986).

Table 1 Studies on the prediction of the bankruptcy resolution

Authors/year	Sample size	Models	Period of time	Type of variables	Accuracy	Country
Casey et al. (1986)	113 companies	Probit regression analysis	1970–1981	Financial	70.80 percent	USA
Kennedy & Shaw (1991)	165 companies	Probit regression analysis	1973–1985	Financial and non-financial	72.88 percent	USA
Campbell (1996)	121 companies	Probit regression analysis	1987–1992	Financial and non-financial	78.50 percent	USA
Barniv et al. (1997)	237 companies	Artificial neural networks, multi-state ordered logistic regression and nonparametric multiple discriminant analysis	1980–1991	Financial, court-related, market and non-financial	54.4 percent	USA
Kumar et al. (1997)	76 companies	Fischer's linear discriminant analysis, neural networks and hybrid genetic classifier	1993	Financial ratios	69.73 percent	USA
Barniv et al. (2002)	237 companies	Logistic regression	1980–1995	Financial and non-financial	75,10 percent	USA
Kim et al. (2008)	59 companies	Logistic regression	1991–2003	Financial, non-financial and audit reports	81.40 percent	South Korea
Jacobs et al. (2012)	518 companies	Multinominal logistic regression	1985–2007	Financial, non-financial and audit reports	62.69 percent	USA
Camacho-Miñano et al. (2015)	1387 companies	PART and Rough Sets	2010	Financial and non-financial	79.15 percent*	Spain
Gupta et al. (2022)	574 companies	Probit, stepwise regressions and LASSO	1994–2019	Financial, macroeconomic, case, judicial and geographic	AUC ROC = 0.9164	USA
Stef & Bissieux (2022)	3488 companies	Multinomial probit	2019–2021	Financial and non-financial	66.03 percent	France

*Cross-validated weighted accuracy

In contrast to the rest of the works exhibited, in the study of Stef and Bissieux (2022) the bankruptcy process was analyzed with four possible states: friendly liquidation, judicial liquidation, safeguard procedure, legal resource procedure. Consequently, this study cannot be compared with the rest, where only two states—resolution and liquidation—are used.

3 Methodology

In this work, the methodology presented in Fig. 1 has been followed. This methodology is similar to what is usually used in this type of research (e.g., Bulman et al., 2021; Morales et al., 2021; Rafique et al., 2019).

First, data were extracted from the SABI platform (Iberian Balance Analysis System: Bureau van Dijk). Then, data cleaning tasks were performed. In addition, various preprocessing tasks were performed to ensure the quality and the relevance of the dataset. Thus, once the financial information was obtained, the observations with missing values were eliminated and the accounting ratios were generated. Next, a selection of characteristics was made. For this, Variance Inflation Factors (VIF) were calculated to detect and mitigate multicollinearity problems. This action helped to improve the speed of execution and reduce the computational complexity of the algorithms. This approach not only optimizes model performance, but also simplifies the analysis and interpretation of the results. In addition, various preprocessing

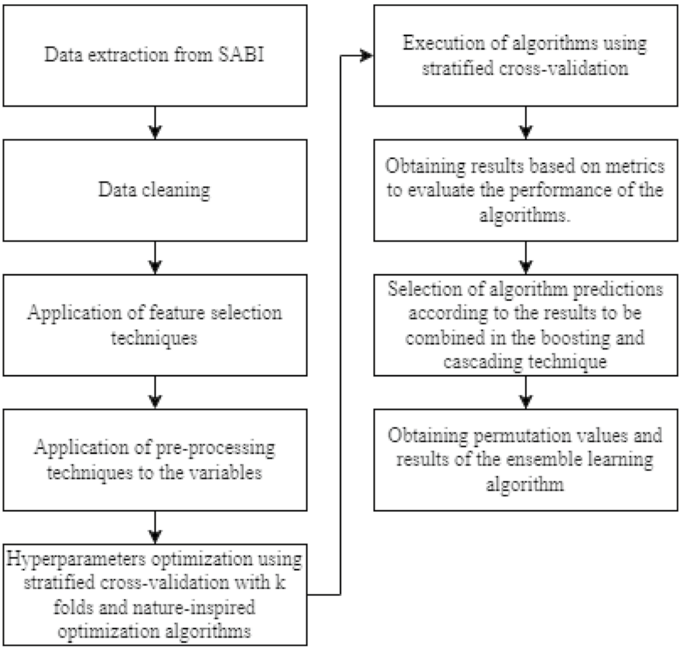

Fig. 1 Methodology flowchart

techniques were implemented to improve the performance of the predictive algorithms. For the numerical variables, the standardization technique was applied using a standard scaler, which ensures that all variables have a mean of zero and a standard deviation of one. This is crucial for algorithms sensitive to the scales of the variables. On the other hand, for categorical variables, such as economic sector, a label encoder was used, which efficiently transforms categories into numerical values. These preprocessing techniques are important to optimize the performance of the machine learning models and ensure more accurate and robust results.

Is important to note that different data sets were used to make the predictions. Thus, the first of these was composed only of the accounting ratios for year N-3, with N being the year in which the status of the resolution of the process was checked. The second contained data only for year N-2. The third contained data only from year N-1. The fourth consisted of the union of data from years N-1, N-2, and N-3. The fifth combined data from N-3 and N-2, and the sixth combined data from N-2 and N-1. Then, the configuration hyperparameters of the different algorithms used in the study were optimized. For this purpose, stratified cross-validation based on k folds where k = 10 (Train = 90 percent and Test = 10 percent of the sample) was used together with the application of nature-inspired optimization algorithms. In this work, algorithms such as the dwarf mongoose optimization algorithm (DMO) (Agushaka et al., 2022), and mountain gazelle optimizer (MGO) (Abdollahzadeh et al., 2022) were used separately. Once the algorithms were correctly configured, the results were obtained again using the previous technique. Subsequently, the results of the metrics for evaluating the performance of the algorithms were obtained. In addition to the steps above, ensemble learning techniques such as boosting and cascading were implemented on top of the catboost algorithm. The input variables for boosting and cascading were the inputs of the other methods and also the predictions of the previous best performing methods, namely support vector machines and k-nearest neighbors. In addition, permutation importance values were calculated using the catboost algorithm. This method allowed us to evaluate the contribution of each variable to the performance of the model.

3.1 Data and Variable Selection

The selected sample included a set of Spanish SMEs that have gone bankrupt and were liquidated or reorganized during the period of 2011–2019. This time period was selected because we aimed to isolate from the study the effects of the financial and economic crisis of 2008 and the one caused by the COVID-19 virus in 2020.

To obtain the sample collection, the SABI database (Iberian Balance Analysis System: Bureau van Dijk) was used. First, a filter was carried out to identify the enterprises that could be classified as SMEs according to their size. According to European regulations, an SME is defined as a company with fewer than 250 employees,and an annual turnover not exceeding 50 million euros, and/or an annual balance sheet total not exceeding 43 million euros (European Comission, 2014). Second, those companies that had submitted accounting information in the 3 years prior to the date of entry into bankruptcy were selected. Third, we filtered those enterprises that, during the period 2008–2019, had gone bankrupt by the end of that period and had been reorganized (in state: active, totaling 461 companies) or liquidated (1222 enterprises). Thus, our total sample comprised 1683 companies.

The description and nature of the independent variables used for the prediction of bankruptcy resolution can be found in Table 2.

In addition, a statistical description of these attributes has been provided (see Appendix A). Most of the accounting variables are financial ratios widely used in the literature focused on bankruptcy and its resolution. Additionally, to avoid problems associated with the multicollinearity of these variables, the variance inflation factor (VIF) was analyzed. Following previous studies that have used this factor in bankruptcy prediction (e.g., Kim & Kang, 2012; Kim et al., 2015), we have applied the most restrictive specification, that the VIF value must be less than 4 for multicollinearity not to exist (see Appendix B), although other authors have considered a limit of 10 (e.g., Agustia et al., 2020; Cho et al., 2021).

Most of the variables used are financial ratios whose relevance has been widely documented in the specialized literature on the prediction of bankruptcy resolution and are fundamental to assessing the financial health of companies. These ratios can be classified into three main categories: (a) profitability, which determines the company's ability to generate profits from its assets or equity; (b) solvency, which evaluates the company's ability to fulfill its long-term obligations; and c) liquidity, which measures the company's ability to satisfy its short-term obligations (e.g., Altman, 1968; Barniv et al., 2002; Beaver, 1966; Du Jardin, 2016; Fisher & Martel, 2009; Kasasbeh, 2021; Kumar & Ravi, 2007; Shetty et al., 2022; Shi et al., 2024; Zmijewski, 1984).

The selection of the ratios utilized in this works was based on the previous studies shown in Table 2. More specifically, we used the liquidity ratio (Current assets/Current liabilities); the solvency ratios: equity to total liabilities (Equity/Total liabilities), and the basic core funding ratio (Equity + Non-current liabilities/Non-current assets + Current operating assets—Current operating liabilities) to measure the capital structure and long-term financial stability. Finally, the profitability ratios used were return on equity (Net income/Equity) and retained earnings to total assets (Retained earnings/Total assets) to measure the company's efficiency in generating profits and its reinvestment capacity.

The dependent variable that reflects the resolution of the bankruptcy, called "Target," is a binary variable whose value is 0 if the company has been liquidated in the resolution process and 1 if the company has been reorganized.

3.2 Modeling Methods

As mentioned above, the objective of this work is to establish a model that can predict bankruptcy resolution with a high accuracy level. Different techniques were used for this purpose: logistic regression (LOGIT), support vector machines (SVM), random forest (RFC), k-nearest neighbors (KNN), fully connected neural networks (FCNN), and ensemble learning techniques such as cascading and boosting with the catboost algorithm. The selection of this set of algorithms and ensemble learning

Table 2 Used variables

Definitions	Formula	Prior studies
Financial ratios		
Bank debt to Total assets	$\frac{Bank\ debt}{Total\ assets}$	Aguiar Díaz and Ruiz Mallorquí (2015a), Leyman et al. (2011)
Liquidity	$\frac{Current\ assets}{Current\ liabilities}$	Aguiar Díaz and Ruiz Mallorquí (2015b), Camacho-Miñano et al. (2015), Huang et al. (2015), Kennedy and Shaw (1991), Radjen and Stanisic (2017)
Current liabilities to Total liabilities	$\frac{Current\ liabilities}{Total\ liabilities}$	Huang et al. (2015)
Equity to Total liabilities	$\frac{Equity}{Total\ liabilities}$	Camacho-Miñano et al. (2015)
Natural logarithm of Total assets	$\ln(Total\ assets)$	Barniv et al., (1997, 2002), Gupta et al. (2022), Huang et al. (2015), Kim et al. (2008), Leyman et al. (2011), Radjen and Stanisic (2017), Rose-Green and Lovata (2013), Wang (2012)
Personal expenses to Total assets	$\frac{Personal\ expenses}{Total\ assets}$	Blazy et al. (2011)
Retained earnings to Total assets	$\frac{Retained\ earnings}{Total\ assets}$	Altman (1968)
Return on equity	$\frac{Net\ income}{Equity}$	Altman (1968), Camacho-Miñano et al. (2015), Jacobs et al. (2012), Kumar et al. (1997)
Trade debt to Total assets	$\frac{Trade\ debt}{Total\ assets}$	Aguiar Díaz and Ruiz Mallorquí (2015a), Leyman et al. (2011)
EBITDA to Total assets	$\frac{EBITDA}{Total\ assets}$	
Basic funding ratio	$\frac{(Equity+Non-current\ liabilities)}{Non-current\ assets} + (Current\ operating\ assets - Current\ operating\ liabilities)$	
Operating earnings to Operating assets	$\frac{Operating\ earnings}{Operating\ assets}$	
Other variables used		
Sector	Spanish Economic Sector	Aguiar Díaz and Ruiz Mallorquí (2015a, 2015b), Blazy et al. (2011), Camacho-Miñano et al. (2015), Gupta et al. (2022), Muñoz-Izquierdo et al. (2019)
Duration	Time in days elapsed from date of incorporation to date of change in the company's status	Aguiar Díaz and Ruiz Mallorquí (2015b), Camacho-Miñano et al. (2015), Muñoz-Izquierdo et al. (2019)

techniques is due to the high performance that they have previously provided in other works related to entry bankruptcy and/or resolution of bankruptcy prediction. Furthermore, these classification techniques are also commonly used in cutting-edge academic works in other fields where machine learning is used (e.g., Mushava & Murray, 2024; Shen et al., 2024). On the other hand, it was considered to use ensemble techniques because they improve the performance of the classification techniques (e.g., Liu et al., 2023; Radovanovic & Haas, 2023). Nonetheless, the algorithms discussed above show some limitations, such as: (a) the sensitivity to noise or outliers that algorithms such as logistic regressions (Adeli et al., 2020), support vector machines (Singla & Shukla, 2020), and k-nearest neighbors (Zhang et al., 2022) have; (b) poor handling of high-dimensional datasets with modeling methods such as random forests (Speiser et al., 2019); (c) the ease of overtraining of algorithms such as logistic regressions (Adeli et al., 2020), and fully connected neural networks (He et al., 2020a, 2020b); (d) sensitivity to hyperparameter adjustment as is the case for methods such as k-nearest neighbors (Zhang et al., 2022), fully connected neural networks (He et al., 2020a, 2020b), and catboost (Hancock & Khoshgoftaar, 2020); and (e) the accuracy of all the modeling methods discussed above may be sensitive to sample size and distribution (Rodriguez et al., 2022; Zhang & Xia, 2022). To deal with the latter issue, a large sample size (1683 companies) was used, being the second study, considering the literature reviewed, with the largest number of companies used for the prediction of bankruptcy resolution (see Table 1). In addition, when running the algorithms, weights are added to adjust for the proportion of liquidated and reorganized companies to obtain balanced results. The software development was performed using the Python programming language (Python Software Foundation, 2021), version 3.8.8.

3.2.1 Logistic Regression

This type of regression seeks to predict a dichotomous dependent variable, normally defined between 0 and 1, through existing relationships with a series of previously identified independent variables. This algorithm has been used in other work related to bankruptcy resolution prediction (e.g., Barniv et al., 1997, 2002; Jacobs et al., 2012; Kim et al., 2008). The model based on logistic regression is shown in Eq. 1:

$$logit(pi) = \ln\left(\frac{p_i}{1-p_i}\right) = \sum_{i=1}^{N} \beta_i X_i + \beta_0 \tag{1}$$

where p_i is the probability that the function takes the value 1, X represents each explained or independent variable, β represents each of its coefficients or parameters to be estimated, and N is the number of observations.

3.2.2 Support Vector Machines

This is a supervised learning algorithm, developed by Vapnik & Cortes (1995) to solve linear and nonlinear problems. It is an algorithm used in studies for similar tasks, such as the prediction of bankruptcy entry (e.g., Antunes et al., 2017; Sun & Li, 2012). The classification function is defined by the Eq. 2:

$$f(y) = sign\left(\sum_{i=1}^{N} y_i p_i K\left(X, X_i\right) + b \right) \tag{2}$$

where *sign* is the sign of the function, p_i y b are parameters that explain the class separator hyperplane, N is the number of observations, and $K\left(X, X_i\right)$ is the kernel function, where the polynomial, sigmoid, or linear functions are worth mentioning.

3.2.3 Random Forest

Random forest is a type of classifier based on decision trees and sampling with replacement of both variables and the training data set. This method is applied to combine a set of randomly selected trees in order to provide strong generalization and high prediction accuracy (Breiman, 2001; Ho, 1995). Recently, random forest has been used for bankruptcy entry prediction (e.g., Antulov-Fantulin et al., 2021; Jabeur et al., 2021), but has not been applied for bankruptcy resolution.

3.2.4 K-Nearest Neighbors

K-nearest neighbors is an algorithm focused on calculating the distance of observations that are inserted in a set of main observations (Altman, 1992; Fix & Hodges, 1951). The classification result arises from most classes of the nearest k elements. The distance can be calculated using different forms such as euclidean, manhattan, and minkowski. Variants of this algorithm have been used by different authors for similar tasks, such as the prediction of entry into bankruptcy (e.g., Chen et al., 2011; Ouenniche et al., 2018).

3.2.5 Fully Connected Neural Networks

This is a type of artificial neural network in which each of the nodes of the different layers is connected to all the nodes of the next layer. After a review of the literature, this algorithm has been applied for the prediction of bankruptcy resolution in 2 times only (e.g., Barniv et al., 1997; Kumar et al., 1997). As shown in Fig. 2, this study used a neural network structure based on one input layer with 14 nodes; three hidden layers, which have 512, 256, and 64 nodes respectively; and an output layer with a single node.

In addition, Fig. 2 shows the application of the network and highlights that each prior node relates to a posterior node through a weight w. In each connection, two functions are performed: one input function, here the weighted sum ($\sum_{j=1}^{N} X_j w_j^i$, where w_j^i is the weight given to the node j of the neuron i and X_j is the value of the input of the node j), and one activation function. In the latter, the connections between the input layer and hidden layers, and between hidden layers themselves, use the

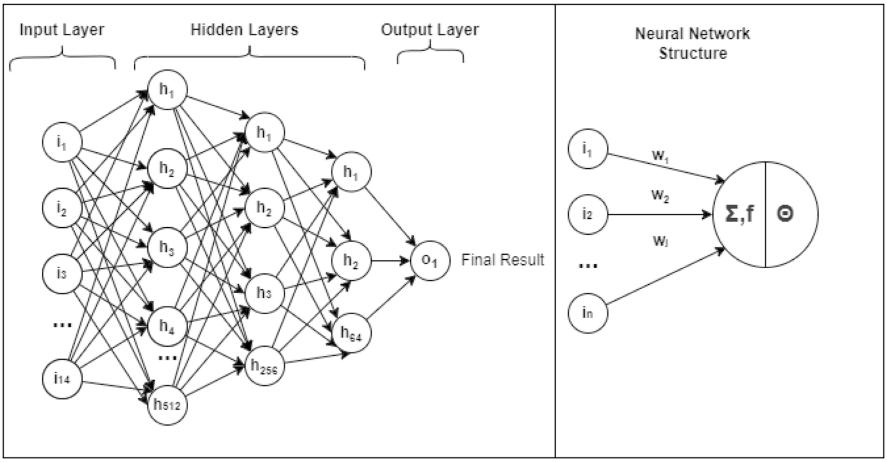

Fig. 2 Model based on FCNN and its structure by node

rectified linear unit (ReLU) function, and for the connection between the last hidden layer and the output layer the sigmoid activation function is used. In addition, dropout layers were included between the connections to avoid overfitting the model.

3.2.6 Ensemble Learning Techniques

With the aim of achieving more robust results, ensemble learning algorithms were used based on boosting and cascading concepts. Boosting is an ensemble learning technique supported on the weighting of prediction error after the use of classifiers in a sequential manner (Zhou, 2009). Algorithms that rely on the idea of boosting have been applied for the prediction of entry into bankruptcy; these algorithms have included catboost (e.g., Jabeur et al., 2021), xgboost (e.g., Du et al., 2020), and gradient boosting machines (e.g., Antulov-Fantulin et al., 2021). However, we are not aware that they have been used in the prediction of bankruptcy resolution. On the other hand, cascading is a method based on the combination of data and predictions made by a set of classifiers to be used by an algorithm in its predictions (Gama & Brazdil, 2000). This significantly improves the results since it uses both data and predictions made by classifiers of another nature. In this study, the catboost algorithm has been applied, using the predictions of the support vector machines and k-nearest neighbors models, since they are the ones that offered the best results in terms of specificity and f1 score.

3.3 Hyperparameters Optimization

To improve the performance of the classifiers, the optimization of their hyperparameters was employed through de algoritmos inspirados en la naturaleza. In this case, optimization algorithms such as dwarf mongoose optimization algorithm (DMO) (Agushaka et al., 2022), and mountain gazelle optimizer (MGO) (Abdollahzadeh et al., 2022) were applied separately due to their high performance compared to other algorithms that were used in bankruptcy prediction works such as particle swarm optimization (PSO) (e.g., Ainan et al., 2024), ant colony optimization (ACO) (e.g., Uthayakumar et al., 2020). This technique tests the best combinations of hyperparameters of the algorithms. For this purpose, a stratified k-fold cross-validation technique was applied together, where k = 10 was applied, and the proportion of the training set was 90 percent, and the test set was 10 percent. Cross validation has been used in bankruptcy resolution prediction research (e.g., Camacho-Miñano et al., 2015; Casey et al., 1986) and aims to generate k models using all parts of the data set so that they are part of each test and training subset and each has the same proportion of observations considering the dependent variable. These methods aim to obtain the hyperparameters that give the best results on the test set by testing each of the hyperparameter combinations. To define the best combination of hyperparameters, we selected those that generated the greatest f1 score on average for the test sets of the 10 folders (see Appendix C).

3.4 Evaluation of Model Performance

Stratified cross-validation based on k-folds was applied again for a different purpose, in this case for the correct evaluation of the performance of the models. It has also been used in bankruptcy resolution prediction research (e.g., Camacho-Miñano et al., 2015; Casey et al., 1986; Kumar et al., 1997). In our study, the best obtained result has a value of k = 10 folders, where for each folder we have 90 percent for the training set and 10 percent for the test set.

The results are calculated as the mean of the results from the 10 models generated by the algorithm for each set of tests. To verify the validity of the model, a series of balanced metrics were used that arise from the confusion matrix and have been weighted to avoid imbalances (see Fig. 3), where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative.

The studies shown in Table 1 lack information related to performance evaluation and model comparison, as the only metric used was accuracy. And although some present the confusion matrix, no ratios are shown for these tools that would allow a more correct assessment of performance. The used metrics are:

		PREDICTED VALUES	
		Reorganization	Liquidation
REAL VALUES	Reorganization	TP	FP
	Liquidation	FN	TN

Fig. 3 Confusion matrix

- $$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{3}$$

- $$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

- $$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

- $$\text{F1 score} = \frac{2 \times precision \times recall}{precision + recall} \tag{6}$$

- $$\text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

- AUC ROC. This is the calculation of the area under the receiver operator characteristics curve.

4 Results

As shown in Table 3, we will differentiate between the results obtained by each of the optimization algorithms applied (DMO and MGO). It should be noted that the performance of algorithms such as random forest (RFC), k-nearest neighbors (KNN), and support vector machines (SVM) is superior to that of the fully connected neural networks (FCNN) and logistic regression (LOGIT). In addition, the ensemble learning techniques called cascading and boosting were applied in the catboost method, achieving results superior to the rest. Thus, the application of these types of techniques has generated more robust and balanced results for the classification of both reorganized and liquidated companies, since it offers better results in metrics such as f1 score (see Table 3). In addition, the application of DMO and MGO in algorithms such as LOGIT, RFC, KNN, and SVM is indifferent because similar results were obtained. However, the results of applying hyperparameter optimization offered by MGO are superior to DMO when applying FCNN. However, the opposite occurs in algorithms such as Catboost, where DMO offers better results than MGO. It should be noted that all these algorithms were used with the data sets mentioned in the methodology section, obtaining the best results with the one containing the union of the data from the 3 years prior to bankruptcy (see Appendix D).

Regarding the main results, those obtained by DMO will be highlighted, since they have been the ones that have obtained the best results with its application. The algorithm that achieved the highest accuracy was catboost, with 94 percent accuracy, followed closely by SVM, with 93.6 percent. In terms of precision, the order is reversed: SVM achieved 97.0 percent, and catboost 96.1 percent. The technique that obtained the best recall result was KNN (95 percent), followed by catboost (91.6 percent). In specificity, SVM and catboost were the most prominent,

Table 3 Algorithm's performance through different metrics

Optimization algorithm	Algorithms	Accuracy	F1 score	Precision	Recall	Specificity	AUC ROC
DMO	FCNN	**0.686** (0.028)	**0.721** (0.015)	**0.659** (0.054)	**0.813** (0.100)	**0.560** (0.151)	**0.784** (0.023)
	LOGIT	**0.660** (0.028)	**0.662** (0.027)	**0.659** (0.033)	**0.666** (0.032)	**0.654** (0.045)	**0.705** (0.038)
	RFC	**0.820** (0.031)	**0.823** (0.030)	**0.814** (0.036)	**0.832** (0.029)	**0.808** (0.041)	**0.901** (0.023)
	KNN	**0.894** (0.013)	**0.899** (0.013)	**0.854** (0.016)	**0.950*** (0.027)	**0.837** (0.023)	**0.890** (0.018)
	SVM	**0.936** (0.019)	**0.933** (0.021)	**0.970*** (0.009)	**0.900** (0.039)	**0.972*** (0.008)	**0.976*** (0.008)
	Catboost	**0.940*** (0.018)	**0.938*** (0.019)	**0.961** (0.010)	**0.916** (0.036)	**0.963** (0.010)	**0.974** (0.013)
MGO	FCNN	**0.725** (0.020)	**0.740** (0.019)	**0.707** (0.046)	**0.786** (0.078)	**0.664** (0.100)	**0.805** (0.018)
	LOGIT	**0.660** (0.028)	**0.662** (0.027)	**0.659** (0.033)	**0.666** (0.032)	**0.654** (0.045)	**0.705** (0.038)
	RFC	**0.820** (0.030)	**0.823** (0.028)	**0.813** (0.035)	**0.833** (0.028)	**0.808** (0.039)	**0.901** (0.023)
	KNN	**0.894** (0.013)	**0.899** (0.013)	**0.854** (0.016)	**0.950*** (0.027)	**0.837** (0.023)	**0.890** (0.018)
	SVM	**0.936** (0.019)	**0.933** (0.021)	**0.970*** (0.009)	**0.900** (0.039)	**0.972*** (0.008)	**0.976*** (0.008)
	Catboost	**0.938*** (0.019)	**0.936*** (0.020)	**0.958** (0.007)	**0.916** (0.037)	**0.960** (0.007)	**0.972** (0.015)

The value in bold is the value of the result of the metric and between brackets the standard deviation. The asterisk indicates the best value among all algorithms for that metric

with 97.2 percent and 96.3 percent, respectively. In terms of f1 score, catboost and SVM repeated in the top position, with 93.8 percent and 93.3 percent, respectively. Finally, with respect to the AUC ROC (see Fig. 4), these two algorithms are once again in the lead, but in inverted positions although with very similar scores: SVM (97.6 percent) and catboost (97.4 percent).

Based on the above results, it seems clear that the algorithms offering superior performance are catboost and SVM. However, if a high recall is required, KNN could also be used.

According to the permutation values of the importance of the variables used for prediction, Table 4 presents those related to the application of the catboost algorithm and the DMO and MGO optimization algorithms, respectively. It should be noted that, as shown in Table 4, the most important variable for prediction is the one composed of SVM predictions. Considering the latter, we have also examined the permutation values of the importance of the features when applying SVM, see Table 5. As is the case in other works, these show that the most important attributes for prediction are duration (e.g., Camacho-Miñano et al., 2015), and sector (e.g., Gupta et al., 2022). Regarding the duration variable, it refers to the age of the company at the time it entered the bankruptcy process, which is a determining factor for the company to emerge successfully from this process, i.e. to reorganize. This greater probability of success in companies with greater longevity is determined by different factors, such as a greater knowledge of the market in which they operate and its dynamics, as well as more deeply rooted relationships with their suppliers, customers, and financial companies, which results in a better predisposition to negotiate conditions that favor the company in its reorganization process (Aguiar Díaz & Ruiz Mallorquí, 2015b; Camacho-Miñano et al., 2015; Muñoz-Izquierdo et al., 2019). The sector variable contains specific market conditions, regulations, and levels of competition that may influence the reorganization capacity of certain companies that have entered bankruptcy process. In this sense, Blazy et al. (2011) show that certain sectors, especially industrial ones, have a higher propensity to reorganize

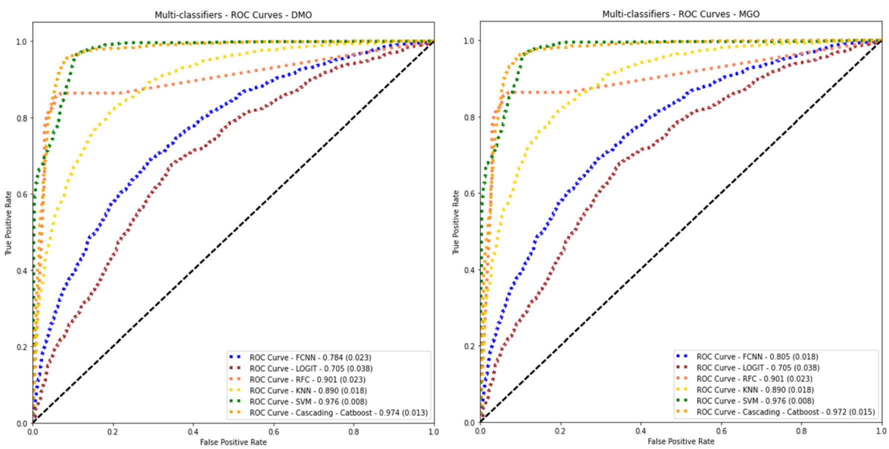

Fig. 4 ROC Curves of the different methods

Table 4 Permutation values of the characteristics applying catboost

Variable	Permutation importance value (DMO)	Variable	Permutation importance value (MGO)
SVM prediction	**0.293** (0.020)	SVM prediction	**0.294** (0.016)
KNN prediction	**0.024** (0.007)	KNN prediction	**0.024** (0.007)
Natural logarithm of Total assets	**0.005** (0.004)	Retained earnings to Total assets	**0.006** (0.004)
Retained earnings to Total assets	**0.004** (0.005)	Duration	**0.004** (0.004)
Return on equity	**0.004** (0.005)	Sector	**0.004** (0.004)
Trade debt to Total assets	**0.003** (0.002)	Natural logarithm of Total assets	**0.004** (0.005)
Sector	**0.003** (0.004)	Trade debt to Total assets	**0.003** (0.004)
Basic funding ratio	**0.003** (0.005)	Return on equity	**0.003** (0.005)
Personal expenses to Total assets	**0.003** (0.005)	Liquidity	**0.002** (0.003)
Duration	**0.003** (0.005)	Basic funding ratio	**0.002** (0.005)
Current liabilities to Total liabilities	**0.002** (0.003)	EBITDA to Total Assets	**0.001** (0.003)
Equity to Total liabilities	**0.002** (0.004)	Personal expenses to Total assets	**0.001** (0.003)
Liquidty	**0.002** (0.005)	Equity to Total liabilities	**0.000** (0.004)
EBITDA to Total Assets	**0.001** (0.004)	Current liabilities to Total liabilities	**0.000** (0.005)
Operating earnings to Operating assets	**0.001** (0.004)	Operating earnings to Operating assets	**−0.000** (0.003)
Bank debt to Total assets	**0.001** (0.004)	Bank debt to Total assets	**−0.000** (0.005)

The value in bold is the value of the result of the metric and between brackets the standard deviation

Table 5 Permutation values of the features applying SVM

Variable	Permutation importance value (DMO)	Variable	Permutation importance value (MGO)
Duration	**0.793** (0.030)	Duration	**0.793** (0.025)
Sector	**0.730** (0.023)	Sector	**0.738** (0.023)
Liquidity	**0.008** (0.005)	Liquidity	**0.007** (0.005)
Natural logarithm of Total assets	**0.002** (0.002)	Basic funding ratio	**0.001** (0.004)
Retained earnings to Total assets	**0.001** (0.001)	Retained earnings to Total assets	**0.001** (0.001)
Equity to Total liabilities	**0.001** (0.001)	Equity to Total liabilities	**0.001** (0.001)
Basic funding ratio	**0.001** (0.004)	Natural logarithm of Total assets	**0.001** (0.001)
Current liabilities to Total liabilities	**0.000** (0.000)	Current liabilities to Total liabilities	**0.000** (0.000)
EBITDA to Total Assets	**0.000** (0.000)	EBITDA to Total Assets	**0.000** (0.000)
Operating earnings to Operating assets	**0.000** (0.000)	Trade debt to Total assets	**0.000** (0.000)
Personal expenses to Total assets	**0.000** (0.001)	Operating earnings to Operating assets	**0.000** (0.001)
Bank debt to Total assets	**0.000** (0.001)	Personal expenses to Total assets	**0.000** (0.001)
Trade debt to Total assets	**0.000** (0.001)	Bank debt to Total assets	**0.000** (0.001)
Return on equity	**−0.002** (0.004)	Return on equity	**−0.002** (0.003)

Note. The value in bold is the value of the result of the metric and between brackets the standard deviation due to factors such as demand stability and entry barriers. For their part, Stef and Bissieux (2022) point out that some sector-specific characteristics, such as the level of innovation and market structure, directly affect the probability of reorganization, especially in capital-intensive sectors. Antill and Grenadier (2019) highlight the influence of the regulatory environment in sectors such as finance and energy, whose legal frameworks can facilitate or hinder reorganization according to government policies. Finally, studies such as Blazy and Esquerré (2021) show that the interaction between industry characteristics and firm financial variables provides a more complete picture for predicting reorganization outcomes.

5 Discussion

Considering the results from other studies (see Table 6), the predictive performance of the model based on SVM, and catboost proposed in this work can be considered to be superior to those obtained previously.

The use of algorithms such as RFC, KNN, FCNN, SVM, and catboost, and techniques such as cascading and boosting based on ensemble learning in tasks related to the prediction of bankruptcy resolution is a novelty in works of this type, not having been detected in previous research.

It should be noted that the results of this work have been obtained using only financial ratios based on accounting information that was easily accessible. Previously this had only been attempted in the study of Casey et al. (1986) and Kumar et al. (1997), where they achieved a 70.8 and 69.7 percent rate of accuracy with a probit regression and artificial neural networks, respectively. Relative to other studies that have used both accounting variables and non-accounting variables, we can highlight that the work of Kim et al. (2008) obtained an 81.4 percent level of accuracy using a logistic regression. and the research of Gupta et al. (2022), which by regression methods obtained 91.6 percent of AUC-ROC. These results are quite far from those achieved with the proposed methodology in this paper, where a 94 percent accuracy level and 97.6 percent AUC-ROC level were achieved.

Then, the study of Campbell (1996) is the one that offers a higher percentage in terms of precision (83.3 percent), which is significantly lower than that obtained in the present research (97 percent using SVM).

For their part, Kim et al. (2008), through logistic regression, achieved the best values of all the studies reviewed in terms of recall (88.6 percent) and f1 score (84.9 percent). These results are considerably lower than those obtained in the present investigation, where a sensitivity of 95 percent was achieved with KNN and an f1 score of 93.8 percent with catboost.

Finally, the work of Casey et al. (1986) is the one that achieved the highest specificity (71.9 percent); however, here SVM achieved superior results (97.2 percent) in this metric.

It should be noted that many of these predictive models of bankruptcy resolution apply data that could introduce biases in determining the prediction. This is because studies such as Jacobs et al. (2012) and Kim et al. (2008) use financial audit

Table 6 Results extracted from similar works

Authors/year	Models	Cross validation	Type of variables	Accuracy	F1 score	Precision	Recall	Specificity	AUC ROC
Casey et al. (1986)	Probit regression analysis	+	Financial	0.708	0.703	0.709	0.696	0.719	–
Kennedy and Shaw (1991)	Probit regression analysis	–	Financial and non-financial	0.729	–	–	–	–	–
Campbell (1996)	Probit regression analysis	–	Financial and non-financial	0.785	0.843	0.833	0.854	0.641	–
Barniv et al. (1997)	Artificial neural networks, multi-state ordered logistic regression and nonparametric multiple discriminant analysis	–	Financial, court-related, market and non-financial	0.544	–	–	–	–	–
Kumar et al. (1997)	Fischer's linear discriminant analysis, neural networks and hybrid genetic classifier	+	Financial ratios	0.697	–	–	–	–	–
Barniv et al. (2002)	Logistic regression	–	Financial and non-financial	0.751	–	–	–	–	–
Kim et al. (2008)	Logistic regression	–	Financial, non-financial and audit reports	0.814	0.849	0.816	0.886	0.708	–
Jacobs et al. (2012)	Multinominal logistic regression	–	Financial, non-financial and audit reports	0.627	–	–	–	–	0.466
Camacho-Miñano et al. (2015)	PART/Rough Sets(RS)	+	Financial and non-financial	0.792	–	–	–	–	–
Gupta et al. (2022)	Probit, stepwise regressions and LASSO	–	Financial, macroeconomic, case, judicial and geographic	–	–	–	–	–	0.916

reports for prediction, which could generate subjectivity (Muñoz-Izquierdo et al., 2019). This has been overcome with the proposed methodology because it uses only accounting data, which allows greater objectivity in the prediction and can be used in any country that employs an accounting system similar to the Spanish one, as is the case of European Union countries, as well as countries regulated under IFRS. Furthermore, this work includes algorithmic techniques to improve the validity and quality of the results, for example, the use of algorithms such as DMO or MGO for hyperparameter optimization, and the application of stratified cross-validation based on k folds.

6 Conclusions, Implications, Limitations, and Future Lines of Research

A bankruptcy situation for any company generates great uncertainty for the different stakeholders. Thus, unnecessary liquidation or a failed reorganization attempt causes significant economic and employment losses and produces social alarm. To safeguard this type of cost, it is crucial to identify at an early stage those SMEs that, despite being in the judicial process associated with bankruptcy, have real options to achieve reorganization or liquidation. Toward this aim, a set of techniques of artificial intelligence were applied to a sample of 1683 Spanish SMEs that went bankrupt in the period 2011–2019 (1222 were liquidated and 461 were reorganized). Of the prediction mechanisms implemented, the one that yields the best results is able to distinguish between liquidated and reorganized companies with an average of 97.6 percent correct predictions. For this purpose, accounting data for 3 years prior to the bankruptcy resolution process were used.

Several theoretical implications stand out from this research. First, knowledge has been contributed to a field of study that has been little explored. Thus, a method based on artificial intelligence algorithms and techniques such as ensemble learning has been presented, which achieves excellent prediction results. Moreover, the method has been designed in such a way that the input variables of the model are easily obtainable and can be implemented in all member states of the European Union, due to the standardization of accounting data by IFRS. This ensures the comparability of the financial ratios used in the study and allows their application not only in the member states of the European Union, but also in other countries that adopt or align with IFRS. This method showed the highest predictive capacity compared to studies in the reviewed academic literature (e.g., Barniv et al., 2002; Camacho-Miñano et al., 2015; Campbell, 1996). In addition, this is the first study to use a methodology that applies ensemble learning techniques, algorithms such as fully connected neural networks, random forest, k-nearest neighbors, catboost, support vector machines, and model performance evaluation such as stratified cross-validation based on k-folds and grid search. Finally, only financial ratios elaborated with easily obtained accounting data have been used. Thus, this answers the standard suggested by Barniv et al. (2002), who stated that accounting data alone are insufficient for the prediction of bankruptcy resolution. In this way, the work contributes to adding value to the role of accounting data in the cited prediction.

Achieving an average accuracy of 94 percent in bankruptcy prediction has important practical implications for stakeholders in bankruptcy procedures. This high accuracy ensures high reliability, thereby significantly improving the ability to identify companies that will be reorganized versus those that should be liquidated, which can optimize the resources and efforts invested. This would allow judges and bankruptcy administrators to make decisions with greater confidence and prioritize those cases with a higher probability of success in the reorganization, improving the efficiency of the process and reducing costs. In addition, for creditors it would allow better risk management and the possibility of maximizing the recovery of their investments; for employees, to know the future situation of their jobs with greater precision; for public administrations, to improve the evaluate the state of the enterprise to which they would provide a possible subsidy; for suppliers, to anticipate with greater accuracy to realize sales strategies; and for customers, to be able to secure supplies.

The main limitation of this study is that only accounting data from Spanish SMEs were used. However, the accounting ratios used in this study can be easily calculated with the balance sheets of companies from other countries that would serve to train classification algorithms. In regard to future lines of research, investigations should be carried out where other types of predictor variables are used, such as other accounting variables and even non-accounting variables such as macroeconomic attributes or those derived from the bankruptcy process, which can be easily obtained and generalized to improve the performance of the modeling techniques. Similarly, with the aim of checking the robustness of the methodology used in this research, it would be interesting to use accounting data from companies in other countries, regions, and sectors and to be able to make comparisons on bankruptcy resolution. As well as looking for the main reasons why companies in certain sectors are more likely to reorganize or liquidate. For example, as is the case with companies in industrial sectors that are more prone to restructure than those outside such sectors (Stef & Jabeur, 2018). In addition, other algorithms could be used to improve the results obtained, for example, convolutional neural networks through image generation with tabular data (e.g., Hosaka, 2019), or other types of optimization algorithms such as Genghis Khan shark (Hu et al., 2023), prairie dog (Ezugwu et al., 2022), or geyser inspired (Ghasemi et al., 2024).

Appendix A: Statistical Description of Independent Variables Used

Variables	Mean	Standard Deviation	Minimun	First quartile	Median	Third quantile	Maximum
Natural logarithm of Total assets	8.113	1.677	0.178	7.004	8.049	9.171	14.675
Liquidity	0.615	0.301	0.000	0.365	0.659	0.891	1.000
EBITDA-Total assets	−0.053	1.427	−27.474	−0.068	0.004	0.048	97.107
Trade debt to Total assets	0.533	6.691	0.000	0.120	0.266	0.483	419.905
Retained earnings to Total assets	−0.176	6.223	−464.864	−0.087	0.003	0.039	30.181
Operating earnings to Operating assets	−0.110	3.359	−89.028	−0.105	−0.004	0.038	154.622
Bank debt to Total assets	0.888	17.666	0.000	0.299	0.504	0.716	1322.009
Personal expenses to Total assets	0.320	1.217	−0.017	0.058	0.160	0.346	79.934
CBF	−72.831	5460.066	−413,264.927	0.595	0.883	1.028	1124.336
Equity to Total liabilities	1.211	31.594	−0.999	0.006	0.157	0.445	1729.439
Current assets to Current liabilities	13.070	606.722	0.000	0.707	1.114	1.832	44,435.200
Return on equity	−2.744	192.912	−14,577.404	−0.161	0.019	0.223	354.931
Sector	4464.044	1641.068	113.000	4110.000	4631.000	4777.000	9900.000
Duration	9184.062	4993.315	735.000	5598.000	8957.000	11,662.000	44,970.000

Appendix B: Variance Inflation Factor (VIF) of the Accounting Variables

Accounting variables	VIF
Natural logarithm of Total assets	3.66
Liquidity	3.51
EBITDA-Total assets	2.90
Trade debt to Total assets	2.75
Retained earnings to Total assets	2.59
Operating earnings to Operating assets	2.39
Bank debt to Total assets	1.74
Personal expenses to Total assets	1.48
CBF	1.03
Equity to Total liabilities	1.01
Current assets to Current liabilities	1.01
Return on equity	1.00

Appendix C: Optimization of Model Hyperparameters

Models	Initial hyperparameters	Optimal hyperparameters (DMO)	Optimal hyperparameters (MGO)
FCNN	Number of epochs = [10–100] Learning_rate = [0.001–0.1] Number of hidden layers = [1–3] Nodes in the hidden layer 1 = [32–512] Nodes in the hidden layer 2 = [32–512] Nodes in the hidden layer 3 = [32–512] Dropout = [0.3–0.7]	Number of epochs = 79 Learning_rate = 0.001 Number of hidden layers = 3 Nodes in the hidden layer 1 = 409 Nodes in the hidden layer 2 = 198 Nodes in the hidden layer 3 = 39 Dropout = 0.3263430888584688	Number of epochs = 73 Learning_rate = 0.008951530243259002 Number of hidden layers = 3 Nodes in the hidden layer 1 = 145 Nodes in the hidden layer 2 = 31 Nodes in the hidden layer 3 = 49 Dropout = 0.3225521674879272
LOGIT	Solver = [‘liblinear’, ‘lbfgs’, ‘saga’,’sag’] Penalty = [‘l1’,’l2’,’elasticnet’] Max_iter = [10–200]	Solver = ‘liblinear’ Penalty = ‘l1’ Max_iter = 146	Solver = ‘liblinear’ Penalty = ‘l1’ Max_iter = 200
RFC	Max_depth = [10–100] N_estimators = [10–1000]	Max_depth = 100 N_estimators = 604	Max_depth = 32 N_estimators = 954
KNN	Algorithm = [‘auto’,’ball_tree’,’kd_tree’,’brute’] P = [1,2], where P = 1 use Manhattan’s distance and P = 2 use Euclidean’s distance Weights = [‘uniform’, ‘distance’] N_neighbors = 2	Algorithm = ’auto’ P = 1 Weights = ’distance’ N_neighbors = 2	Algorithm = ’auto’ P = 1 Weights = ’distance’ N_neighbors = 2
SVM	Gamma = [0.001–0.1] C = [1–100] Kernel = [’linear’, ’poly’, ’rbf’, ’sigmoid’, ’precomputed’]	Gamma = 0.002336023969268924 C = 24.742641633014358 Kernel = ‘rbf’	Gamma = 0.0023420730885665587 C = 55.56346718268796 Kernel = ‘rbf’
Cascading—Catboost	Depth = [1–20] L2_leaf_reg = [1–10] Learning_rate = [0.001–0.1]	Depth = 10 L2_leaf_reg = 6 Learning_rate = 0.09549558079885992	Depth = 7 L2_leaf_reg = 8 Learning_rate = 0.07761075370762993

Appendix D: Performance of the Algorithms Using Accounting Data from Different Prior Years, Where N Being the Base Year Where the Bankruptcy Occurred

Using accounting data of the year N-3

Optimization algorithm	Algorithms	Accuracy	F1 score	Precision	Recall	Specificity	AUC ROC
DMO	FCNN	**0.691** (0.034)	**0.673** (0.042)	**0.717** (0.050)	**0.640** (0.068)	**0.742*** (0.070)	**0.745** (0.045)
	LOGIT	**0.650** (0.058)	**0.647** (0.066)	**0.651** (0.056)	**0.646** (0.087)	**0.653** (0.074)	**0.712** (0.060)
	RFC	**0.719** (0.067)	**0.722** (0.071)	**0.711** (0.058)	**0.735** (0.093)	**0.703** (0.063)	**0.791*** (0.059)
	KNN	**0.564** (0.031)	**0.553** (0.048)	**0.566** (0.030)	**0.544** (0.072)	**0.583** (0.052)	**0.586** (0.036)
	SVM	**0.545** (0.035)	**0.666** (0.025)	**0.526** (0.020)	**0.907*** (0.040)	**0.182** (0.046)	**0.555** (0.056)
	Catboost	**0.728*** (0.060)	**0.725*** (0.063)	**0.734*** (0.065)	**0.720** (0.077)	**0.735** (0.081)	**0.787** (0.058)
MGO	FCNN	**0.666** (0.024)	**0.649** (0.052)	**0.688** (0.056)	**0.634** (0.119)	**0.698** (0.104)	**0.738** (0.048)
	LOGIT	**0.650** (0.058)	**0.647** (0.066)	**0.651** (0.056)	**0.646** (0.087)	**0.653** (0.074)	**0.712** (0.060)
	RFC	**0.721** (0.065)	**0.723** (0.069)	**0.715** (0.058)	**0.735** (0.091)	**0.707** (0.064)	**0.791*** (0.061)
	KNN	**0.564** (0.031)	**0.553** (0.048)	**0.566** (0.030)	**0.544** (0.072)	**0.583** (0.052)	**0.586** (0.036)
	SVM	**0.545** (0.035)	**0.666** (0.025)	**0.526** (0.020)	**0.907*** (0.040)	**0.182** (0.046)	**0.555** (0.056)
	Catboost	**0.728*** (0.057)	**0.725*** (0.064)	**0.729*** (0.048)	**0.725** (0.092)	**0.731*** (0.059)	**0.790** (0.050)

The value in bold is the value of the result of the metric and between brackets the standard deviation. The asterisk indicates the best value among all algorithms for that metric

Using accounting data of the year N-2							
Optimization algorithm	Algorithms	Accuracy	F1 score	Precision	Recall	Specificity	AUC ROC
DMO	FCNN	**0.652** (0.026)	**0.653** (0.017)	**0.657** (0.050)	**0.655** (0.053)	**0.649** (0.092)	**0.709** (0.020)
	LOGIT	**0.635** (0.050)	**0.633** (0.051)	**0.637** (0.057)	**0.631** (0.058)	**0.638** (0.072)	**0.686** (0.046)
	RFC	**0.692** (0.045)	**0.695** (0.047)	**0.689** (0.046)	**0.703** (0.062)	**0.681** (0.065)	**0.774*** (0.060)
	KNN	**0.567** (0.042)	**0.562** (0.038)	**0.571** (0.047)	**0.555** (0.044)	**0.579** (0.073)	**0.592** (0.062)
	SVM	**0.541** (0.042)	**0.666** (0.028)	**0.524** (0.025)	**0.913*** (0.039)	**0.169** (0.064)	**0.549** (0.054)
	Catboost	**0.698*** (0.045)	**0.695*** (0.042)	**0.707*** (0.055)	**0.685** (0.050)	**0.711*** (0.075)	**0.774** (0.063)
MGO	FCNN	**0.657** (0.046)	**0.654** (0.062)	**0.663** (0.055)	**0.662** (0.116)	**0.653** (0.110)	**0.702** (0.036)
	LOGIT	**0.635** (0.05)	**0.633** (0.051)	**0.637** (0.057)	**0.631** (0.058)	**0.638** (0.072)	**0.686** (0.046)
	RFC	**0.694** (0.050)	**0.697** (0.048)	**0.693** (0.052)	**0.703** (0.056)	**0.685** (0.069)	**0.774*** (0.060)
	KNN	**0.567** **(0.042)**	**0.562** (0.038)	**0.571** (0.047)	**0.555** (0.044)	**0.579** (0.073)	**0.592** (0.062)
	SVM	**0.541** (0.042)	**0.666** (0.028)	**0.524** (0.025)	**0.913*** (0.039)	**0.169** (0.064)	**0.549** (0.054)
	Catboost	**0.702*** (0.057)	**0.701*** (0.054)	**0.707*** (0.065)	**0.698** (0.062)	**0.705*** (0.085)	**0.768** (0.073)

The value in bold is the value of the result of the metric and between brackets the standard deviation. The asterisk indicates the best value among all algorithms for that metric

Using accounting data of the year N-1							
Optimization algorithm	Algorithms	Accuracy	F1 score	Precision	Recall	Specificity	AUC ROC
DMO	FCNN	**0.709** (0.043)	**0.704** (0.053)	**0.715** (0.036)	**0.697** (0.076)	**0.722** (0.042)	**0.766** (0.040)
	LOGIT	**0.698** (0.047)	**0.701** (0.050)	**0.696** (0.050)	**0.709** (0.068)	**0.688** (0.068)	**0.752** (0.048)
	RFC	**0.765*** (0.040)	**0.762*** (0.034)	**0.781*** (0.073)	**0.748** (0.041)	**0.781** (0.089)	**0.835*** (0.034)
	KNN	**0.585** (0.048)	**0.579** (0.053)	**0.588** (0.051)	**0.575** (0.076)	**0.594** (0.079)	**0.612** (0.058)
	SVM	**0.548** (0.023)	**0.669** (0.015)	**0.528** (0.015)	**0.913*** (0.039)	**0.182** (0.057)	**0.549** (0.055)
	Catboost	**0.762** (0.034)	**0.758** (0.027)	**0.779** (0.059)	**0.742** (0.035)	**0.783*** (0.077)	**0.832** (0.031)
MGO	FCNN	**0.714** (0.048)	**0.724** (0.039)	**0.705** (0.061)	**0.751** (0.057)	**0.677** (0.100)	**0.770** (0.047)
	LOGIT	**0.698** (0.047)	**0.701** (0.050)	**0.696** (0.050)	**0.709** (0.068)	**0.688** (0.068)	**0.752** (0.048)
	RFC	**0.765*** (0.044)	**0.762*** (0.037)	**0.781*** (0.075)	**0.748** (0.034)	**0.781*** (0.089)	**0.836*** (0.035)
	KNN	**0.585** (0.048)	**0.579** (0.053)	**0.588** (0.051)	**0.575** (0.076)	**0.594** (0.079)	**0.612** (0.058)
	SVM	**0.548** (0.023)	**0.669** (0.015)	**0.528** (0.015)	**0.913*** (0.039)	**0.182** (0.057)	**0.549** (0.055)
	Catboost	**0.755** (0.040)	**0.751** (0.033)	**0.772** (0.066)	**0.735** (0.041)	**0.775** (0.088)	**0.831** (0.027)

The value in bold is the value of the result of the metric and between brackets the standard deviation. The asterisk indicates the best value among all algorithms for that metric

Using accounting data of the years N-3, N-2, and N-1							
Optimization algorithm	Algorithms	Accuracy	F1 score	Precision	Recall	Specificity	AUC ROC
DMO	FCNN	**0.686** (0.028)	**0.721** (0.015)	**0.659** (0.054)	**0.813** (0.100)	**0.560** (0.151)	**0.784** (0.023)
	LOGIT	**0.660** (0.028)	**0.662** (0.027)	**0.659** (0.033)	**0.666** (0.032)	**0.654** (0.045)	**0.705** (0.038)
	RFC	**0.820** (0.031)	**0.823** (0.030)	**0.814** (0.036)	**0.832** (0.029)	**0.808** (0.041)	**0.901** (0.023)
	KNN	**0.894** (0.013)	**0.899** (0.013)	**0.854** (0.016)	**0.950*** (0.027)	**0.837** (0.023)	**0.890** (0.018)
	SVM	**0.936** (0.019)	**0.933** (0.021)	**0.970*** (0.009)	**0.900** (0.039)	**0.972*** (0.008)	**0.976*** (0.008)
	Catboost	**0.940*** (0.018)	**0.938*** (0.019)	**0.961** (0.010)	**0.916** (0.036)	**0.963** (0.010)	**0.974** (0.013)
MGO	FCNN	**0.725** (0.020)	**0.740** (0.019)	**0.707** (0.046)	**0.786** (0.078)	**0.664** (0.100)	**0.805** (0.018)
	LOGIT	**0.660** (0.028)	**0.662** (0.027)	**0.659** (0.033)	**0.666** (0.032)	**0.654** (0.045)	**0.705** (0.038)
	RFC	**0.820** (0.030)	**0.823** (0.028)	**0.813** (0.035)	**0.833** (0.028)	**0.808** (0.039)	**0.901** (0.023)
	KNN	**0.894** (0.013)	**0.899** (0.013)	**0.854** (0.016)	**0.950*** (0.027)	**0.837** (0.023)	**0.890** (0.018)
	SVM	**0.936** (0.019)	**0.933** (0.021)	**0.970*** (0.009)	**0.900** (0.039)	**0.972*** (0.008)	**0.976*** (0.008)
	Catboost	**0.938*** (0.019)	**0.936*** (0.020)	**0.958** (0.007)	**0.916** (0.037)	**0.960** (0.007)	**0.972** (0.015)

The value in bold is the value of the result of the metric and between brackets the standard deviation. The asterisk indicates the best value among all algorithms for that metric

Using accounting data of the years N-3, and N-2							
Optimization algorithm	Algorithms	Accuracy	F1 score	Precision	Recall	Specificity	AUC ROC
DMO	FCNN	**0.692** (0.052)	**0.704** (0.044)	**0.687** (0.062)	**0.736** (0.096)	**0.649** (0.142)	**0.764** (0.045)
	LOGIT	**0.648** (0.033)	**0.645** (0.044)	**0.649** (0.030)	**0.643** (0.070)	**0.653** (0.042)	**0.694** (0.038)
	RFC	**0.770** (0.021)	**0.776** (0.020)	**0.755** (0.025)	**0.800** (0.027)	**0.739** (0.034)	**0.854** (0.024)
	KNN	**0.809** (0.020)	**0.818** (0.021)	**0.780** (0.023)	**0.862*** (0.044)	**0.756** (0.036)	**0.798** (0.024)
	SVM	**0.825** (0.020)	**0.802** (0.025)	**0.923*** (0.033)	**0.710** (0.040)	**0.939*** (0.031)	**0.914** (0.012)
	Catboost	**0.856*** (0.026)	**0.853*** (0.027)	**0.877** (0.042)	**0.833** (0.045)	**0.880** (0.050)	**0.934*** (0.016)
MGO	FCNN	**0.714** (0.039)	**0.702** (0.055)	**0.727** (0.031)	**0.684** (0.090)	**0.743** (0.044)	**0.780** (0.042)
	LOGIT	**0.648** (0.033)	**0.645** (0.044)	**0.649** (0.030)	**0.643** (0.070)	**0.653** (0.042)	**0.694** (0.038)
	RFC	**0.769** (0.021)	**0.775** (0.020)	**0.757** (0.023)	**0.794** (0.027)	**0.744** (0.029)	**0.852** (0.024)
	KNN	**0.809** (0.020)	**0.818** (0.021)	**0.780** (0.023)	**0.862*** (0.044)	**0.756** (0.036)	**0.798** (0.024)
	SVM	**0.825** (0.020)	**0.802** (0.025)	**0.923*** (0.033)	**0.710** (0.040)	**0.939*** (0.031)	**0.914** (0.012)
	Catboost	**0.862*** (0.023)	**0.859*** (0.025)	**0.877** (0.035)	**0.845** (0.050)	**0.878** (0.044)	**0.933*** (0.016)

The value in bold is the value of the result of the metric and between brackets the standard deviation. The asterisk indicates the best value among all algorithms for that metric

Using accounting data of the years N-2, and N-1							
Optimization algorithm	Algorithms	Accuracy	F1 score	Precision	Recall	Specificity	AUC ROC
DMO	FCNN	**0.722** (0.027)	**0.729** (0.033)	**0.712** (0.038)	**0.754** (0.075)	**0.690** (0.072)	**0.784** (0.025)
	LOGIT	**0.652** (0.025)	**0.662** (0.023)	**0.644** (0.028)	**0.681** (0.031)	**0.623** (0.043)	**0.700** (0.025)
	RFC	**0.800** (0.029)	**0.803** (0.029)	**0.791** (0.031)	**0.817** (0.038)	**0.783** (0.039)	**0.872** (0.026)
	KNN	**0.845** (0.027)	**0.857** **(0.025)**	**0.795** (0.029)	**0.932*** (0.038)	**0.758** (0.046)	**0.816** (0.037)
	SVM	**0.905** (0.020)	**0.900** (0.022)	**0.942*** (0.021)	**0.862** (0.035)	**0.947*** (0.020)	**0.956*** (0.015)
	Catboost	**0.914*** (0.026)	**0.912*** (0.026)	**0.937** (0.029)	**0.889** (0.035)	**0.939** (0.029)	**0.953** (0.015)
MGO	FCNN	**0.679** (0.042)	**0.694** (0.036)	**0.675** (0.068)	**0.738** (0.121)	**0.619** (0.173)	**0.766** (0.032)
	LOGIT	**0.652** (0.025)	**0.662** (0.023)	**0.644** (0.028)	**0.681** (0.031)	**0.623** (0.043)	**0.700** (0.025)
	RFC	**0.799** (0.028)	**0.802** (0.028)	**0.791** (0.029)	**0.815** (0.040)	**0.784** (0.037)	**0.872** (0.026)
	KNN	**0.845** (0.027)	**0.857** (0.025)	**0.795** (0.029)	**0.932*** (0.038)	**0.758** (0.046)	**0.816** (0.037)
	SVM	**0.905** (0.020)	**0.900** (0.022)	**0.942*** (0.021)	**0.862** (0.035)	**0.947*** (0.020)	**0.956*** (0.015)
	Catboost	**0.918*** (0.024)	**0.915*** (0.024)	**0.941** (0.026)	**0.892** (0.034)	**0.944** (0.026)	**0.954** (0.015)

The value in bold is the value of the result of the metric and between brackets the standard deviation. The asterisk indicates the best value among all algorithms for that metric

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability The datasets generated and/or analyzed during the current study are not publicly available due to are property of Bureau Van Dijk but are available from the corresponding author on reasonable request.

Declarations

Conflict of Interest The authors have not disclosed any conflict of interests.

References

Abdollahzadeh, B., Gharehchopogh, F. S., Khodadadi, N., & Mirjalili, S. (2022). Mountain Gazelle optimizer: A new Nature-inspired Metaheuristic Algorithm for Global Optimization Problems. *Advances in Engineering Software, 174*, 103282. https://doi.org/10.1016/j.advengsoft.2022.103282

Adeli, E., Li, X., Kwon, D., Zhang, Y., & Pohl, K. M. (2020). Logistic regression confined by cardinality-constrained sample and feature selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(7), 1713–1728. https://doi.org/10.1109/TPAMI.2019.2901688

Aguiar Díaz, I., & Ruiz Mallorquí, M. V. (2015a). Causes and resolution of bankruptcy: The efficiency of the law. *Spanish Review of Financial Economics, 13*(2), 71–80. https://doi.org/10.1016/j.srfe.2015.04.001

Aguiar Díaz, I., & Ruiz Mallorquí, M. V. (2015b). Deuda bancaria versus deuda comercial y tiempo de resolución del concurso. *Investigaciones Europeas De Dirección y Economía De La Empresa, 21*(3), 111–119. https://doi.org/10.1016/j.iedee.2014.09.003

Agushaka, J. O., Ezugwu, A. E., & Abualigah, L. (2022). Dwarf Mongoose optimization algorithm. *Computer Methods in Applied Mechanics and Engineering, 391*, 114570. https://doi.org/10.1016/j.cma.2022.114570

Agustia, D., Muhammad, N. P. A., & Permatasari, Y. (2020). Earnings management, business strategy, and bankruptcy risk: Evidence from Indonesia. *Heliyon, 6*(2), e03317. https://doi.org/10.1016/j.heliyon.2020.e03317

Ainan, U. H., Por, L. Y., Chen, Y. L., Yang, J., & Ku, C. S. (2024). Advancing Bankruptcy forecasting with hybrid machine learning techniques: Insights from an unbalanced polish dataset. *IEEE Access, 12*, 9369–9381. https://doi.org/10.1109/ACCESS.2024.3354173

Altman, E. I. (1968). The prediction of corporate bankruptcy. *The Journal of Finance, XXIII*, 589–609.

Altman, N. S. (1992). An introduction to Kernel and Nearest-neighbor nonparametric regression. *The American Statistician, 46*(3), 175–185. https://doi.org/10.2307/2685209

Antill, S. (2022). Do the right firms survive bankruptcy? *Journal of Financial Economics, 144*(2), 523–546. https://doi.org/10.1016/j.jfineco.2021.07.006

Antill, S., & Grenadier, S. R. (2019). Optimal capital structure and bankruptcy choice: Dynamic bargaining versus liquidation. *Journal of Financial Economics, 133*(1), 198–224. https://doi.org/10.1016/j.jfineco.2018.05.012

Antulov-Fantulin, N., Lagravinese, R., & Resce, G. (2021). Predicting bankruptcy of local government: A machine learning approach. *Journal of Economic Behavior and Organization, 183*, 681–699. https://doi.org/10.1016/j.jebo.2021.01.014

Antunes, F., Ribeiro, B., & Pereira, F. (2017). Probabilistic modeling and visualization for bankruptcy prediction. *Applied Soft Computing Journal, 60*, 831–843. https://doi.org/10.1016/j.asoc.2017.06.043

Barniv, R., Agarwal, A., & Leach, R. (1997). Predicting the outcome following bankruptcy filing: A three-state classification using neural networks. *International Journal of Intelligent Systems in Accounting, Finance & Management, 6*(3), 177–194. https://doi.org/10.1002/(sici)1099-1174(199709)6:3%3c177::aid-isaf134%3e3.3.co;2-4

Barniv, R., Agarwal, A., & Leach, R. (2002). Predicting bankruptcy resolution. *Journal of Business Finance and Accounting, 29*(3–4), 497–520. https://doi.org/10.1111/1468-5957.00440

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*(1966), 71. https://doi.org/10.2307/2490171

Bernstein, S., Colonnelli, E., Giroud, X., & Iverson, B. (2019). Bankruptcy spillovers. *Journal of Financial Economics, 133*(3), 608–633. https://doi.org/10.1016/j.jfineco.2018.09.010

Blazy, R., Chopard, B., Fimayer, A., & Guigou, J. D. (2011). Employment preservation vs. creditors' repayment under bankruptcy law: The French dilemma? *International Review of Law and Economics, 31*(2), 126–141. https://doi.org/10.1016/j.irle.2011.03.002

Blazy, R., & Esquerré, S. (2021). The CV effect: To what extent does the chance to reorganize depend on a bankruptcy judge's profile? *International Review of Law and Economics*. https://doi.org/10.1016/j.irle.2021.105984

Blazy, R., Martel, J., & Nigam, N. (2014). The choice between informal and formal restructuring: The case of French banks facing distressed SMEs. *Journal of Banking and Finance, 44*(1), 248–263. https://doi.org/10.1016/j.jbankfin.2014.04.015

Breiman, L. (2001). *Random forests*. Kluwer Academic Publishers.

Bulman, Z. P., Krapp, F., Pincus, N. B., Wenzler, E., Murphy, K. R., Qi, C., et al. (2021). Genomic features associated with the degree of phenotypic resistance to Carbapenems in Carbapenem-resistant Klebsiella pneumoniae. *mSystems*. https://doi.org/10.1128/msystems.00194-21

Camacho-Miñano, M. D. M., Segovia-Vargas, M. J., & Pascual-Ezama, D. (2015). Which characteristics predict the survival of insolvent firms? An SME reorganization prediction model. *Journal of Small Business Management*, *53*(2), 340–354.

Campbell, S. V. (1996). Predicting bankruptcy reorganization for closely held firms. *Accounting Horizons, 10*(3), 12–25.

Casey, C. J., Mcgee, V. E., & Stickney, C. P. (1986). Discriminating between Reorganized and Liquidated Firms in Bankruptcy. *The Accounting Review*, *61*(2), 249–262. http://www.jstor.org/stable/247256

Casterella, J. R., Lewis, B. L., & Walker, P. L. (2000). Modeling the audit opinions issued to bankrupt companies: A two-stage empirical analysis. *Decision Sciences, 31*(2), 507–530. https://doi.org/10.1111/j.1540-5915.2000.tb01632.x

Chen, H. L., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S. J., & Liu, D. Y. (2011). A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method. *Knowledge-Based Systems, 24*(8), 1348–1359. https://doi.org/10.1016/j.knosys.2011.06.008

Cho, E., Okafor, C., Ujah, N., & Zhang, L. (2021). Executives' gender-diversity, education, and firm's bankruptcy risk: Evidence from China. *Journal of Behavioral and Experimental Finance, 30*, 100500. https://doi.org/10.1016/j.jbef.2021.100500

De Blick, T., Paeleman, I., & Laveren, E. (2024). Financing constraints and SME growth: The suppression effect of cost-saving management innovations. *Small Business Economics, 62*(3), 961–986. https://doi.org/10.1007/s11187-023-00797-9

Demiroglu, C., & James, C. (2015). Bank loans and troubled debt restructurings. *Journal of Financial Economics, 118*(1), 192–210. https://doi.org/10.1016/j.jfineco.2015.01.005

Du, J., & Nguyen, B. (2022). Cognitive financial constraints and firm growth. *Small Business Economics, 58*(4), 2109–2137. https://doi.org/10.1007/s11187-021-00503-7

Du Jardin, P. (2016). A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research, 254*(1), 236–252. https://doi.org/10.1016/j.ejor.2016.03.008

Du, X., Li, W., Ruan, S., & Li, L. (2020). CUS-heterogeneous ensemble-based financial distress prediction for imbalanced dataset with ensemble feature selection. *Applied Soft Computing Journal*. https://doi.org/10.1016/j.asoc.2020.106758

Eidenmüller, H. (2017). Contracting for a European insolvency regime. *European Business Organization Law Review, 18*(2), 273–304. https://doi.org/10.1007/s40804-017-0067-1

European Comission. Regulation (EU) No 651/2014 of 17 June 2014 declaring certain categories of aid compatible with the internal market in application of Articles 107 and 108 of the Treaty, 57 Official Journal of the European Union (2014). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.187.01.0001.01.ENG&toc=OJ%3AL%3A2014%3A187%3ATOC

Eurostat. (2023). *Q2 2023: Business bankruptcies at highest level since 2015*. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20230817-1. Accessed 26 May 2024

Ezugwu, A. E., Agushaka, J. O., Abualigah, L., Mirjalili, S., & Gandomi, A. H. (2022). *Prairie Dog Optimization Algorithm*. *Neural Computing and Applications* (Vol. 34). Springer London. https://doi.org/10.1007/s00521-022-07530-9

Fisher, T. C. G., & Martel, J. (2009). An empirical analysis of the firm's reorganization decision. *Finance, 30*(1), 121. https://doi.org/10.3917/fina.301.0121

Fix, E., & Hodges, J. (1951). Nonparametric discrimination: consistency properties. *Randolph Field, Texas, Project,* (Feb), 21–49.

Gama, J., & Brazdil, P. (2000). Cascade generalization. *Machine Learning, 41*(3), 315–343. https://doi.org/10.1023/A:1007652114878

Gestel, T. V., Baesens, B., Suykens, J. A. K., Van den Poel, D., Baestaens, D. E., & Willekens, M. (2006). Bayesian kernel based classification for financial distress detection. *European Journal of Operational Research, 172*(3), 979–1003. https://doi.org/10.1016/j.ejor.2004.11.009

Ghasemi, M., Zare, M., Zahedi, A., Akbari, M. A., Mirjalili, S., & Abualigah, L. (2024). Geyser inspired algorithm: A new geological-inspired meta-heuristic for real-parameter and constrained engineering optimization. *Journal of Bionic Engineering, 21*(1), 374–408. https://doi.org/10.1007/s42235-023-00437-8

Gupta, J., Kingdom, U., Barzotto, M., Kingdom, U., & Behaviour, S. (2022). Bankruptcy resolution: Misery or strategy bankruptcy resolution. *Misery or Strategy, 2018*(December), 1–61.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: An interdisciplinary review. *Journal of Big Data*. https://doi.org/10.1186/s40537-020-00369-8

He, C., Ma, M., & Wang, P. (2020a). Extract interpretability-accuracy balanced rules from artificial neural networks: A review. *Neurocomputing, 387*, 346–358. https://doi.org/10.1016/j.neucom.2020.01.036

He, D., Yu, K., & Wu, J. (2020b). Industry characteristics, court location, and bankruptcy resolution. *Journal of Management Analytics, 7*(3), 389–423. https://doi.org/10.1080/23270012.2020.1715272

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278–282. https://ieeexplore.ieee.org/abstract/document/598994/

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications, 117*, 287–299. https://doi.org/10.1016/j.eswa.2018.09.039

Hu, G., Guo, Y., Wei, G., & Abualigah, L. (2023). Genghis Khan shark optimizer: A novel nature-inspired algorithm for engineering optimization. *Advanced Engineering Informatics, 58*(October), 102210. https://doi.org/10.1016/j.aei.2023.102210

Huang, J. C., Huang, C. S., & You, C. F. (2015). Bank relationships and the likelihood of filing for reorganization. *International Review of Economics and Finance, 35*(123), 278–291. https://doi.org/10.1016/j.iref.2014.10.007

Jabeur, S. B., Gharib, C., Mefteh-Wali, S., & Arfi, W. B. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change, 166*(January), 120658. https://doi.org/10.1016/j.techfore.2021.120658

Jacobs, M., Karagozoglu, A. K., Layish, D. N., et al. (2012). Resolution of corporate financial distress: An empirical analysis of processes and outcomes. *The Journal of Portfolio Management, 38*(2), 117–135.

James, S. D. (2016). Strategic bankruptcy: A stakeholder management perspective. *Journal of Business Research, 69*(2), 492–499. https://doi.org/10.1016/j.jbusres.2015.05.006

Kasasbeh, F. I. (2021). Impact of financing decisions ratios on firm accounting-based performance: Evidence from Jordan listed companies. *Future Business Journal, 7*(1), 1–10. https://doi.org/10.1186/s43093-021-00061-0

Keasey, K., Pindado, J., & Rodrigues, L. (2015). The determinants of the costs of financial distress in SMEs. *International Small Business Journal: Researching Entrepreneurship, 33*(8), 862–881. https://doi.org/10.1177/0266242614529317

Kennedy, D. B., & Shaw, W. H. (1991). Evaluating financial distress resolution using prior audit opinions. *Contemporary Accounting Research, 8*(1), 97–114. https://doi.org/10.1111/j.1911-3846.1991.tb00837.x

Kim, M. J., & Kang, D. K. (2012). Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. *Expert Systems with Applications, 39*(10), 9308–9314. https://doi.org/10.1016/j.eswa.2012.02.072

Kim, M. J., Kang, D. K., & Kim, H. B. (2015). Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. *Expert Systems with Applications, 42*(3), 1074–1082. https://doi.org/10.1016/j.eswa.2014.08.025

Kim, M., Kim, M., & McNiel, R. (2008). Predicting survival prospect of corporate restructuring in Korea. *Applied Economics Letters, 15*(15), 1187–1190. https://doi.org/10.1080/13504850601018080

Kumar, N., Krovi, R., & Rajagopalan, B. (1997). Financial decision support with hybrid genetic and neural based modeling tools. *European Journal of Operational Research, 103*(2), 339–349. https://doi.org/10.1016/S0377-2217(97)00124-0

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. *European Journal of Operational Research, 180*(1), 1–28. https://doi.org/10.1016/j.ejor.2006.08.043

Laitinen, E. K. (2011). Assessing viability of Finnish reorganization and bankruptcy firms. *European Journal of Law and Economics, 31*(2), 167–198. https://doi.org/10.1007/s10657-009-9136-4

Leyman, B., Schoors, K. J. L., & Coussement Peter, P. (2011). Does court-supervised reorganization work? Evidence from post-confirmation firm failure. *International Review of Law and Economics, 31*(3), 149–168. https://doi.org/10.1016/j.irle.2011.05.001

Liu, W., Fan, H., & Xia, M. (2023). Tree-based heterogeneous cascade ensemble model for credit scoring. *International Journal of Forecasting, 39*(4), 1593–1614. https://doi.org/10.1016/j.ijforecast.2022.07.007

Ministerio de Industria Comercio y Turismo de España. (2021). Marco Estratégico en Política de PYME 2030, 1–122. https://industria.gob.es/es-es/Servicios/MarcoEstrategicoPYME/Marco Estratégico PYME.pdf%0Ahttps://plataformapyme.es/SiteCollectionDocuments/EstrategiaPYME/Marco_Estategico_Politica_PYME_2030.pdf

Morales, G., Sheppard, J., Logan, R., & Shaw, J. (2021). Hyperspectral Band Selection for Multispectral Image Classification with Convolutional Networks. *Proceedings of the International Joint Conference on Neural Networks*, *2021-July*. https://doi.org/10.1109/IJCNN52387.2021.9533700

Muñoz-Izquierdo, N., Camacho-Miñano, M. D. M., Segovia-Vargas, M. J., & Pascual-Ezama, D. (2019). Is the external audit report useful for bankruptcy prediction? Evidence using artificial intelligence. *International Journal of Financial Studies*. https://doi.org/10.3390/ijfs7020020

Mushava, J., & Murray, M. (2024). Flexible loss functions for binary classification in gradient-boosted decision trees: An application to credit scoring. *Expert Systems with Applications, 238*(PC), 121876. https://doi.org/10.1016/j.eswa.2023.121876

Nicolas, T. (2022). Short-term financial constraints and SMEs' investment decision: Evidence from the working capital channel. *Small Business Economics, 58*(4), 1885–1914. https://doi.org/10.1007/s11187-021-00488-3

OECD. (2024). *Financing SMEs and Entrepreneurs 2024*. https://doi.org/10.1787/fa521246-en

Ouenniche, J., Pérez-Gladish, B., & Bouslah, K. (2018). An out-of-sample framework for TOPSIS-based classifiers with application in bankruptcy prediction. *Technological Forecasting and Social Change, 131*(May 2017), 111–116. https://doi.org/10.1016/j.techfore.2017.05.034

Prusak, B., Morawska, S., Kuczewska, J., & Banasik, P. (2019). The role of stakeholders on rejection of bankruptcy applications in the case of "poverty" of the estate: A polish case study. *International Insolvency Review, 28*(1), 63–85. https://doi.org/10.1002/iir.1329

Python Software Foundation. (2021). *Python 3.8.8 documentation*. https://docs.python.org/3/. Accessed 5 September 2021

Radjen, D., & Stanisic, N. (2017). Financial distress and managerial turnover: The case of the Republic of Serbia. *Prague Economic Papers, 26*(6), 646–660. https://doi.org/10.18267/j.pep.628

Radovanovic, J., & Haas, C. (2023). The evaluation of bankruptcy prediction models based on socio-economic costs. *Expert Systems with Applications, 227*(April), 120275. https://doi.org/10.1016/j.eswa.2023.120275

Rafique, W., Zheng, D., Barras, J., Joglekar, S., & Kosmas, P. (2019). Predictive analysis of landmine risk. *IEEE Access, 7*, 107259–107269. https://doi.org/10.1109/ACCESS.2019.2929677

Rashid, S. (2019). Bankruptcy laws: A comparative study of India and Usa. *Journal of Management*. https://doi.org/10.34218/jom.6.2.2019.028

Rodriguez, S. D., Pascual, M. S., Oletto, A., Barnabas, S., Zuidewind, P., Dobbels, E., et al. (2022). Machine learning outperformed logistic regression classification even with limit sample size: A model to predict pediatric HIV mortality and clinical progression to AIDS. *PLoS ONE, 17*(10 October), 1–13. https://doi.org/10.1371/journal.pone.0276116

Rose-Green, E., & Lovata, L. (2013). The relationship between firms' characteristics in the periods prior to Bankruptcy filing and bankruptcy outcome. *Accounting and Finance Research, 2*(1), 97–109. https://doi.org/10.5430/afr.v2n1p97

Shen, F., Yang, Z., Kuang, J., & Zhu, Z. (2024). Reject inference in credit scoring based on cost-sensitive learning and joint distribution adaptation method. *Expert Systems with Applications, 251*(April), 124072. https://doi.org/10.1016/j.eswa.2024.124072

Shetty, S., Musa, M., & Brédart, X. (2022). Bankruptcy prediction using machine learning techniques. *Journal of Risk and Financial Management*. https://doi.org/10.3390/jrfm15010035

Shi, Y., Qu, Y., Chen, Z., Mi, Y., & Wang, Y. (2024). Improved credit risk prediction based on an integrated graph representation learning approach with graph transformation. *European Journal of Operational Research, 315*(2), 786–801. https://doi.org/10.1016/j.ejor.2023.12.028

Silva, V. A. B., & Saito, R. (2020). Corporate financial distress and reorganization: A survey of theoretical and empirical contributions. *Revista Brasileira De Gestao De Negocios, 22*(Special Issue), 401–420. https://doi.org/10.7819/rbgn.v22i0.4057

Singla, M., & Shukla, K. K. (2020). Robust statistics-based support vector machine and its variants: A survey. *Neural Computing and Applications, 32*(15), 11173–11194. https://doi.org/10.1007/s00521-019-04627-6

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications, 134*, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028

Stef, N. (2022). Bankruptcy voting process and corporate reorganization. *Bulletin of Economic Research*. https://doi.org/10.1111/boer.12368

Stef, N., & Bissieux, J. J. (2022). Resolution of corporate insolvency during COVID-19 pandemic. Evidence from France. *International Review of Law and Economics, 70*, 106063. https://doi.org/10.1016/j.irle.2022.106063

Stef, N., & Jabeur, S. B. (2018). The Bankruptcy prediction power of new entrants. *International Journal of the Economics of Business, 25*(3), 421–440. https://doi.org/10.1080/13571516.2018.1455389

Sun, J., & Li, H. (2012). Financial distress prediction using support vector machines: Ensemble vs. individual. *Applied Soft Computing Journal, 12*(8), 2254–2265. https://doi.org/10.1016/j.asoc.2012.03.028

Tarantino, E. (2013). Bankruptcy law and corporate investment decisions. *Journal of Banking and Finance, 37*(7), 2490–2500. https://doi.org/10.1016/j.jbankfin.2013.02.007

Uthayakumar, J., Metawa, N., Shankar, K., & Lakshmanaprabu, S. K. (2020). Financial crisis prediction model using ant colony optimization. *International Journal of Information Management, 50*(November 2018), 538–556. https://doi.org/10.1016/j.ijinfomgt.2018.12.001

Vapnik, V., & Cortes, C. (1995). Support-Vector Networks. *Kluwer Academic Publishers*. http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf

Wang, C. A. (2012). Determinants of the choice of formal bankruptcy procedure: An international comparison of reorganization and liquidation. *Emerging Markets Finance and Trade, 48*(2), 4–28. https://doi.org/10.2753/REE1540-496X480201

White, M. J. (2016). Small business Bankruptcy. *Annual Review of Financial Economics, 8*, 1–30.

Wu, W. W. (2010). Beyond business failure prediction. *Expert Systems with Applications, 37*(3), 2371–2376. https://doi.org/10.1016/j.eswa.2009.07.056

Yu, K., & He, D. (2018). The choice between bankruptcy liquidation and bankruptcy reorganization: A model and evidence. *Journal of Management Analytics, 5*(3), 170–197. https://doi.org/10.1080/23270012.2018.1462112

Zhang, B., & Xia, C. (2022). The effects of sample size and sample prevalence on cellular automata simulation of urban growth. *International Journal of Geographical Information Science, 36*(1), 158–187. https://doi.org/10.1080/13658816.2021.1931237

Zhou, Z.-H. (2009). Ensemble learning. *Encyclopedia of Biometrics*. https://doi.org/10.1007/978-0-387-73003-5_293

Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of AccouNting Research, 22*, 59–82.