



HAL
open science

Bankruptcy prediction using machine learning and Shapley additive explanations

Hoang Hiep Nguyen, Jean-Laurent Viviani, Sami Ben Jabeur

► **To cite this version:**

Hoang Hiep Nguyen, Jean-Laurent Viviani, Sami Ben Jabeur. Bankruptcy prediction using machine learning and Shapley additive explanations. *Review of Quantitative Finance and Accounting*, In press, 10.1007/s11156-023-01192-x . hal-04223161

HAL Id: hal-04223161

<https://hal.science/hal-04223161>

Submitted on 7 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bankruptcy prediction using Machine learning and Shapley Additive Explanations

(This version - August 2023 has been published in the
Review of Quantitative Finance and Accounting)

Hoang Hiep Nguyen¹ | Jean-Laurent Viviani¹ | Sami Ben Jabeur²

Abstract

Recently, ensemble-based machine learning models have been widely used and have demonstrated their efficiency in bankruptcy prediction. However, these algorithms are black box models and people cannot understand why they make their forecasts. This explains why interpretability methods in machine learning attract attention from many artificial intelligence researchers. In this paper, we evaluate the prediction performance of Random Forest, LightGBM, XGBoost, and NGBoost (Natural Gradient Boosting for probabilistic prediction) for French firms from different industries with the horizon of one to five years. We then use Shapley Additive Explanations (SHAP), a model-agnostic method to explain XGBoost, one of the best models for our data. SHAP can show how each feature impacts the output from XGBoost. Furthermore, single prediction can also be explained, thus allowing black box models to be used in credit risk management.

Keywords: Shapley Additive Explanations; Explainable Machine learning; Bankruptcy prediction; Ensemble-based model; XGBoost.

JEL classifications: C82, D81, G33.

¹Univ. Rennes, CNRS, CREM - UMR6211, F-35000 Rennes, France.

²Institute of Sustainable Business and Organizations, Sciences and Humanities Confluence Research Center - UCLY, ESDES, 10 place des archives, 69002, Lyon, France.

1. Introduction

In the last 60 years, predicting bankruptcy has been an interesting topic that has attracted the attention of many researchers. Beaver (1966) utilized a univariate analysis; Altman (1968) employed a multivariate technique; and machine learning and deep learning techniques have recently been widely implemented. However, there is no mature or clear theory of why businesses fail (Wang et al., 2014). Through empirical studies, researchers in many countries continue to try to find the best way to predict the fate of a failed firm. The main problems in the bankruptcy prediction field are improving model accuracy, selecting appropriate variables, increasing the prediction time frame, and interpreting bankruptcy prediction models. Bellovary et al. (2007) reviewed 165 bankruptcy prediction studies, many of which achieved good predictive performance. The researchers raised an interesting question, "Why do we continue to develop new and different models for bankruptcy prediction?" and suggested that "future research should consider how these models can be applied and, if necessary, refined".

The most challenging aspect of using artificial intelligence to predict bankruptcies is interpreting and explaining the results. In Europe, according to the "right to explanation" rule, customers who are denied a loan have the right to demand explanations from financial institutions. Nevertheless, du Jardin (2017) concluded that ensemble models cannot be used as a decision-making tool in practice because they cannot be interpreted. Following Son et al. (2019), there are two primary reasons why machine learning models are not utilized in practice. First, the prediction accuracy does not significantly exceed that of statistical models. Second, the results cannot be interpreted. In addition, most studies dealing with distress prediction focus on improving the accuracy of the model and pay insufficient attention to the model's interpretation (Alfaro et al., 2008).

In recent years, a growing number of studies have used model-agnostic methods, including Partial Dependence Plot, Local interpretable model-agnostic explanations (LIME), and Shapley Additive Explanations (SHAP), to interpret machine learning models. Among these methods, SHAP, developed by Lundberg and Lee (2017), stands out as the only one based on game theory, an economic theory. SHAP measures the contribution of each explanatory variable at each point prediction of a machine learning model, making it useful for predicting the failure of companies and potentially transforming the industry's use of ensemble models for bankruptcy forecasting. Despite its potential, most studies that have employed SHAP to interpret their models prioritize improving the accuracy of their bankruptcy prediction models. For instance, Perboli and

Arabnezhad (2021) sought to achieve short-term prediction accuracy improvements while also making accurate mid- and long-term predictions. Yıldırım et al. (2021) utilized graph theory in default prediction, and Sigrist and Leuenberger (2023) introduced a hybrid econometric-machine learning model to model multi-period cumulative and forward corporate default probabilities. These studies have used SHAP to enhance the credibility of their black box models. However, financial analysis in bankruptcy prediction using SHAP has yet to be comprehensively examined. To address this knowledge gap, we use SHAP to perform financial ratio analysis for predicting bankruptcy. This local model-agnostic method allows us to understand the impact of each financial ratio on bankruptcy prediction in our data. Our study contributes to the field of corporate failure prediction in three ways. Firstly, we analyze the effects of financial ratios on financial distress in a cross-industry sample of French companies, using SHAP to interpret the relationship between each ratio and bankruptcy prediction. Our study provides insights into ambiguous relationships by showing interaction visualizations. Secondly, we employ the NGBoost model, which shows promise in bankruptcy prediction despite the XGBoost model's robustness. We are the first researchers, to our knowledge, to use this model in bankruptcy prediction. Finally, while it is recommended to use companies from the same industry to avoid producing misleading results, our study demonstrates that ensemble-based machine learning algorithms can predict bankruptcy accurately across various French industries.

This paper is structured as follows: section 2 contains a review of the literature that clarifies our research question. In section 3, we present the sample and variables. In section 4, we briefly introduce the machine learning models used in our experiments. In section 5, we summarize our results before concluding in Section 6.

2. Literature review

The main objective of many corporate bankruptcy prediction studies is to improve model accuracy. There are two main categories: statistical techniques and artificial intelligence techniques. Statistical techniques include discriminant analysis, logistic regression, etc. Artificial intelligence techniques consist of an artificial neural network, a support vector machine, ensemble models and deep learning. Comparing model accuracy among these techniques is a common goal for many researchers (Barboza et al., 2017; Charalambous et al., 2022, 2023; Staňková, 2023)

Following Hung and Chen (2009) and Wang et al. (2014), there is no consistent conclusion that one specific technique is better than another. However, Altman et al. (2020) concluded that logistic

regression and neural networks outperform support vector machines, gradient boosting and decision trees. In recent years, ensemble-based techniques have performed better than single models in corporate failure prediction (Alfaro et al., 2008; du Jardin, 2019; Nanni & Lumini, 2009). Jones et al. (2017) recommended using ensemble techniques because they predict significantly better than all other classifiers on both cross-sectional and longitudinal test samples. We compile a list of recent studies using complex machine learning algorithms in Table 1. The number of variables used in these studies varies from 16 to 50. Some researchers employ fewer than 20 independent variables, but their models can provide a high level of accuracy in bankruptcy prediction (Perboli and Arabnezhad 2021; Bussmann et al. 2021; Smith and Alvarez 2022; Sigrist and Leuenberger 2023).

Table 1

Explanatory variables.

Study	Models	Number of variables	Sample size	Time period	Variable types	Interpretable methods
du Jardin (2010)	NN	14	1,020	2002-2003	Financial ratio	None
du Jardin (2015)	NN, SOM	36	18,620	2003-2012	Financial ratio, Market-based variables	None
du Jardin (2016)	Bagging, Boosting, Random subspace,	35	17,660	2003-2012	Financial ratio,	None
Jones et al. (2017)	NN, SVMs, AdaBoost, RF	35	30,129	After the year 2000	Financial ratio	Relative variable importance
Veganzones and Séverin (2018)	NN, SVM, RF	10	1,500	2013-2014	Financial ratio	None
Sigrist and Hirsenschall (2019)	Grabit, Boosted MC Logit, Boosted Logit, RF, NN	50	155	2014-2016	Financial ratio, SME characteristics	Partial dependence plots
Perboli and Arabnezhad, (2021)	RF, NN, XGBoost	15	8,959	2001-2018	Financial ratio	SHAP
Bussmann et al. (2021)	XGBoost	19	15,045	2015	Balance-sheet variables	SHAP
Smith and Alvarez (2022)	XGBoost, LightGBM, NN, RF, SVM	17	64,057	1992-2016	Financial ratio Balance-sheet variables	Feature importance score Partial dependence plot

Sigrist and Leuenberger (2023)	TreeBoost, NN, RF, Ensemble	16	21,637	1961-2020	Financial ratio and Industry dummies	SHAP
--------------------------------	-----------------------------	----	--------	-----------	--------------------------------------	------

NN Neural Networks, *RF* Random Forest, *SVM* Support Vector Machine, *SOM* Self Organizing Map.

Although the variable selection method in a bankruptcy prediction model plays an important role in enhancing the model's performance, there is a limited theoretical framework for selecting the best variables (Balcaen and Ooghe 2006). Following Bellovary et al. (2007), 752 different variables are used in bankruptcy prediction models, classified as financial ratio, market-based variables, textual data (firm's textual disclosure) and non-financial variables. The number of variables considered in a study ranges from 1 to 57. Likewise, du Jardin (2010) found that more than 500 different ratios were used in nearly 200 scientific papers. Furthermore, recent research has introduced innovative predictors of failure, such as human resources, earnings management, communicative value of annual report, and macroeconomic indicators (Almaskati et al., 2021; T.-K. Chen et al., 2023; Hernandez Tinoco & Wilson, 2013; Kellner et al., 2022; Schalck & Yankol-Schalck, 2021; Séverin & Veganzones, 2021; Veganzones et al., 2023; Zhao et al., 2022). Numerous studies (Ben Jabeur et al., 2023; Cho et al., 2010; du Jardin, 2010; Jabeur & Serret, 2023; Wang et al., 2014) have demonstrated enhanced prediction accuracy in bankruptcy models by employing meticulous variable selection from the extensive pool of bankruptcy predictors.

Increasing the horizon of bankruptcy prediction is another subject of studies on business failure. The shorter the horizon, the more obvious the signs of bankruptcy that any statistical method can predict accurately. As the horizon increases and symptoms of bankruptcy become weaker, the differences in accuracy among the methods can vary. Following du Jardin and Séverin (2011), for a period of three to six years, a self-organizing map is more accurate than other techniques such as discriminant analysis, logistic regression, neural networks and survival analysis. Altman et al. (2020) found that a logistic regression model can perform as well as a neural network and even better than a support vector machine over a 10-year period.

The last objective in a business failure study is how to interpret and explain bankruptcy prediction using ensemble-based models. These models can be replicated easily by market practitioners because they require minimal intervention for data preparation, selection of variables and specification of the model architecture (Jones et al., 2017). However there is a trade-off between accuracy and interpretability called the black box problem (du Jardin, 2017; Lundberg & Lee, 2017). Researchers try to overcome the black box problem by measuring the importance of features

and the relationship between features and the predicted outcome. Appendix A summarizes the important features resulting from various methods used to explain and interpret ensemble-based models. These studies use several important variables. Jones et al. (2017) facilitated model interpretation by measuring the importance of variables using the relative variable importance method in random forest. Other studies calculated an importance score for each feature and used a partial dependence plot to illustrate their effects on the model's output (Jabeur et al., 2021; Sigrist & Hirnschall, 2019; Smith & Alvarez, 2022; Son et al., 2019). Some researches focused on the importance of explaining individual predictions in practice (Bussmann et al., 2021; Perboli and Arabnezhad, 2021; Sigrist and Hirnschall, 2019; Smith and Alvarez, 2021).

Recently, several studies introduced SHAP, a novel consistent and accurate method, which can explain any particular prediction to interpret XGBoost and LagaBoost in bankruptcy prediction (Moen 2020; Perboli and Arabnezhad 2021; Bussmann et al. 2021; Schalck and Yankol-Schalck 2021; Sigrist and Leuenberger 2023). SHAP is the only explanation method based on a solid theory (Molnar, 2020). Thus, SHAP has also been widely employed in other domains, such as drought prediction (Dikshit & Pradhan, 2021), power system (Zhang et al., 2020) and air quality prediction (Vega García & Aznarte, 2020). Because interpretable machine learning has been developed only recently, studies analyzing financial ratio impacts from black box models remain rare. Almost all studies applying SHAP simply identify the important features with no further interpretation (Bussmann et al. 2021; Yildirim et al. 2021; Sigrist and Leuenberger 2023). Surprisingly, the aspect of financial analysis has not been closely examined in prior works. In fact, Sigrist and Leuenberger (2023) utilized SHAP to highlight the most important feature for their novel model, LaGaBoost, without interpretation. Their study provided the SHAP dependence plot but did not describe how a variable affected the prediction. In the same way, Yildirim et al. (2021) employed SHAP to illustrate which variables contributed most to the prediction. Perboli and Arabnezhad (2021) utilized SHAP to provide the most important features of their model for a random company. However, a financial analysis cannot be conducted because their financial ratios are not clearly disclosed.

Compared with prior work in this research stream, our study provides a comprehensive analysis of financial ratios for bankruptcy prediction. First, we demonstrate the relationship between each financial ratio and failure prediction. We attempt to explain these relationships using financial theories and compare them with the findings of previous research. Second, for certain ratios for

which the relationship with bankruptcy prediction is not defined, we interpret the interaction between the values of these ratios and their Shapley values to illustrate how these ratios contribute to the failure prediction. Third, we investigate the causes of bankruptcy for different companies in the dataset. These analyses allow us to better understand the bankruptcy process and enrich the list of important financial ratios used by black box models in bankruptcy prediction.

3. Data and variables

Our study data is based on a sample of French companies from various industries collected by Diane Bureau van Dijk, which provides the complete data of French companies. We select 24 available financial ratios from 6 categories (financial structure, liquidity-solvency, efficiency, profitability, rotation and contribution) computed by Diane Bureau van Dijk. These financial variables are used in numerous previous researches and have demonstrated their bankruptcy prediction ability in machine learning (du Jardin, 2017, 2019; Jabeur et al., 2021). Table 2 represents the ratios used in our research, which were collected annually from 2013 to 2017.

Following du Jardin (2010), both young, failed and non-failed companies have the same financial structure. Therefore, we only collect companies that were established before January 1, 2014 (in operation for at least 4 years). For size criteria, all companies had total assets of less than 43 million euros in 2018. The bankruptcy occurred in 2018 and non-failed companies were still in business in 2019. From our criteria, the sample includes 4,054 available bankrupt firms in 2018, 5,600 non-failed firms that were chosen randomly among approximately 1.4 million firms available in the database. We opted to use a subsample of non-failed firms, which is a commonly used technique in previous bankruptcy prediction studies (Altman, 1968; Beaver, 1966; du Jardin, 2010; Jabeur et al., 2021; Smith & Alvarez, 2022), as it allows for a more manageable dataset to work with while maintaining the representativeness of the sample. For instance, in practice, the Banque de France (the central bank of France) constructs a credit score using annual samples of approximately 5,000 companies, one third of which are failed firms (Bardos, 1995). This score has been tested on annual samples of around 30,000 companies, and the technique is still in use today (du Jardin, 2019). Additionally, using a subsample allows for better control over the data and reduces the computational resources required for modeling and analysis. Overall, our decision to use a subsample is in line with standard industry practice and previous research in bankruptcy prediction.

We divide our data into a training sample (80% of the population) to estimate the parameters of the machine learning models and a test sample (20% of the population) to evaluate the accuracy of the models. The proportion of failed firms in both our training and test samples is approximately 40%, which is consistent with other studies in the field (du Jardin, 2010, 2016; Wang et al., 2014). Zmijewski (1984) demonstrated that the overall classification rate of a bankruptcy model using logistic regression or probit regression is stable when the proportion of failed and non-failed firms falls within the range of 0.5 to 0.04.

Appendix B shows the descriptive statistics for all independent variables from 1 to 5 years before failure. In general, non-failed firms always have a higher median profitability than failed firms, such as R17 (Profit or Loss/Shareholder Funds), R18 (Profit or Loss/Capital Employed). One year before the bankruptcy, the deterioration symptom becomes clearer. Failed firms have negative average profitability and efficiency ratios, for example R18 (Profit or Loss/Capital Employed), R13 (EBIT/Operating Revenues) and R15 (Operating Cash flow/Value-Added). Non-failed firms, however, show their better performance by their positive mean profitability ratios (Appendix B, Table 6). In addition, Appendix C presents the correlation between the independent variables one, two, three, four and five years before their bankruptcy. We use Principal Component Analysis (PCA) in Discriminant analysis and Logistic regression to handle correlated variables. To handle outliers in our analysis, we adopt the Winsorization method at the 5% level. By limiting the effects of outliers, we can obtain more accurate and reliable results in our analysis while also ensuring that our conclusions are not distorted by a few anomalous observations.

Table 2

Explanatory variables.

Financial dimension	Variable	Formula	Variable value
Financial Structure	R01	Capital Employed/Fixed Assets	Absolute value
	R02	(Shareholder Funds/Capital Employed)*100	Percentage
	R03	(Financial Debt/Capital Employed)*100	Percentage
	R04	(Shareholder Funds/Total Assets)*100	Percentage
Liquidity-solvency	R05	Working Capital Requirement/Current Assets	Absolute value
	R06	Current Asset/Current Liabilities	Absolute value
	R07	Financial Debt/Cash Flow	Absolute value
Efficiency	R08	(Cash flow/Total Sales)*100	Percentage
	R09	(Working Capital Requirement/Total Sales)*360	Absolute value (days)
	R10	(Net Current Assets/Total Sales)*360	Absolute value (days)
	R11	Total Sales/ Tangible and Intangible Assets	Absolute value
	R12	Total Sales/Total Assets	Absolute value
	R13	(EBIT/Operating Revenue)*100	Percentage
	R14	(Value-Added/Total Sales)*100	Percentage
	R15	(Operating Cash flow/Value-Added)*100	Percentage

Profitability	R16	(Profit or Loss before Tax/Operating Revenue)*100	Percentage
	R17	(Profit or Loss/Shareholder Funds)*100	Percentage
	R18	(Profit or Loss/Capital Employed)*100	Percentage
Rotation	R19	(Stock/ Total Sales)*360	Absolute value (days)
	R20	(Debtors/Operating Revenue)*360	Absolute value (days)
	R21	(Creditors/Operating Revenue)*360	Absolute value (days)
Contribution	R22	(Cost of Labor/Value-Added)*100	Percentage
	R23	(Taxes/Value-Added)*100	Percentage
	R24	(Interest/Value-Added)*100	Percentage

EBIT: Earnings before interest and taxes.

4. Methodology

4.1 Modeling method

In this paper, we divide eight models into three groups. The first group includes traditional models with Logistic Regression and Discriminant Analysis. The second group, called single machine learning models, includes Support Vector Machine and Neural Network. The last group consisting of ensemble models contains Random Forest, Gradient Boosting Model such as LightGBM, XGBoost, and Natural Gradient Boost.

4.1.1 Discriminant Analysis

Altman (1968) uses multiple discriminant analysis to distinguish between failed and non-failed firms by calculating the Z score function for each firm:

$$Z = w_0 + \sum_{i=1}^n w_i x_i$$

Here x_i represents the explanatory variables such as: Working Capital/Total assets, Retained Earnings/Total Assets, Earnings Before Interest and Taxes/Total Assets, Market Value of Equity/Book Value of Total Debt and Sales/Total Assets. The coefficients w_i are calculated to maximize between-class variances and minimize within-class variances. The disadvantage of the method is that it requires restrictive statistical assumptions, which is sometimes impossible for financial data.

4.1.2 Logistic regression

After discriminant analysis, Ohlson (1980) was the first researcher to use logistic regression in bankruptcy prediction. The method calculates the probability of default for each company through a sigmoid function as follows:

$$Z = 1 + \frac{1}{1 + e^{-(w_0 + \sum_{i=1}^n w_i x_i)}}$$

where x_i represents the explanatory variable and w_0 , w_i are the intercept and the regression coefficient. Compared with discriminant analysis, logistic regression does not require a strong statistical condition regarding linearity, normal distribution and homoscedasticity. However, a dependent variable must follow a logistic distribution and the problem of multicollinearity among explanatory variables must be avoided.

4.1.3 Support Vector Machine

Cortes and Vapnik (1995) were the first to introduce the support vector machine (SVM) as a binary classifier. The idea of SVM is to create an optimal separating hyperplane in order to classify data into two groups (bankruptcy and non-bankruptcy). The hyperplane maximizes the distance between the nearest observation of each group and its decision boundary. In case of a non-linear separable dataset, SVM transforms the dataset using kernel functions into a higher dimension space in which a hyperplane might separate two classes linearly. Recently, SVM has been used widely in bankruptcy prediction as one of references of performance comparison among many machine learning classifiers (Alaka et al., 2018; Jabeur et al., 2021).

4.1.4 Neural Network

An artificial neural network is a state-of-the-art technique for many applications. In bankruptcy prediction, the algorithm was first applied in 1990 (Odom & Sharda, 1990). Beside discriminant analysis, logistic regression and SVM, a neural network is also considered to be a reference measure used to evaluate the performance of other ensemble methods. The neural network contains multiple layers. The inputs (financial ratios) from the first layer are processed to the output layer (failed or non-failed company) through hidden layers to calculate the weight of each node. In the back propagation stage, the node weights are recalculated in order to reduce the classification error. The neural network is a non-linear classifier so the algorithm does not require any assumptions about input or output. However, it is difficult to define relationships between input variables and output classification.

4.1.5 Random forest

The random forest is an ensemble learning method proposed by Breiman (2001). The random forest classifier consists of a large number of individual decision trees. Based on bootstrap sampling, each individual decision tree generates a class prediction. The outcome of the random forest is the class prediction which receives the most votes from an individual tree. Following Krauss et al. (2017), Loureiro et al. (2018) and Mercadier and Lardy (2019), random forest classifiers

demonstrate their effectiveness in economics and finance. Random forest models outperform traditional models so this classifier has been widely used in recent years (Babar et al., 2020).

4.1.6 XGBoost

XGBoost is an ensemble tree model based on gradient boosting that was first introduced by Chen and Guestrin (2016). The final model of XGBoost is described as:

$$Z = F(x_i) = \sum_{i=1}^T f_t(x_i)$$

where x_i represents independent variables, and $f_t(x_i)$ denotes the output function of each tree. XGBoost is more efficient and faster than Gradient Boosting thanks to a regularized learning objective. The regularization term penalizes the complexity of each tree function in order to avoid overfitting. The regularization term is included in the loss function as follows:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i; y_i) + \sum_k \Omega(f_k)$$

and $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$

where l is loss the function measuring the difference between the prediction and the real value, Ω is the regularization function calculated based on parameters γ, λ , number of leaves T and the score w at each leaf. If the regularization parameters are zero, the objective function simply becomes the original Gradient boosting. XGBoost is not only credited by winning many Kaggle competitions but has also been used in various disciplines. In bankruptcy prediction, XGBoost demonstrates its efficiency through numerous experimental studies in Poland (Zięba et al., 2016), Korea (Son et al., 2019), Spain (Smith & Alvarez, 2022), Europe (Climent et al. 2019; Bussmann et al. 2021) and the United States (Carmona et al., 2019).

4.1.7 LightGBM

Like XGBoost, the Light gradient boosting machine or LightGBM is a new efficient implementation of the Gradient boosting decision tree. This popular algorithm was proposed by Ke et al. (2017) and has been used for classification tasks in many research areas, such as peer-to-peer network loan default (Ma et al., 2018), cryptocurrency price prediction (Sun et al., 2020), chemistry (C. Chen et al., 2019).

LightGBM differs from other gradient boosting decision tree algorithms by using gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB). GOSS is used to partition the optimal node by calculating the variance gain. EFB increases the speed of the training process by

clustering many exclusive features into less dense features. These GOSS and EFB techniques allow the algorithm to grow the tree vertically instead of horizontally like XGBoost. LightGBM might therefore improve the training speed and accuracy especially for large datasets.

4.1.8 NGBoost

NGBoost is a natural gradient boosting algorithm for probabilistic forecasting developed recently by Duan et al. (2020). The algorithm differs from other algorithms in that it returns the conditional probability of the output instead of the expected value in the regression task. In classification problems, NGBoost predicts the probability of each class like most other classification algorithms. The new algorithm consist of a base learner, a parametric probability distribution and a scoring rule. According to Duan et al. (2020), NGBoost provides similar performance compared with existing methods for probabilistic prediction while offering advantages in terms of flexibility, scalability and ease of use.

4.2 Choice of tuning parameters

The Grid search is an important step in hyperparameter tuning before the training process. This popular optimization technique can provide optimal performance for a machine learning algorithm by combining different hyperparameter values. We use the grid search during the cross-validation phase and employ AUC (area under the curve) as a measure of success for each combination of hyperparameters. In cross-validation, the training sample is divided randomly into k groups (called folds), where one group is considered as a test data set and the remaining $k-1$ folds are considered as a training set. Each machine learning model is a fitted training set and evaluated on the test set. This process is repeated until all the fold data become test set to avoid the overfitting problem. In our research, we tune hyperparameters with 10-fold cross-validation. The optimal set of hyperparameters found after grid search is run again with 10-fold cross-validation in a test sample to verify the performance. The hyperparameters that give the best performance of SVM, Neural network, Random forest, XGboost, LightGBM and NGBoost are presented in Appendix D.

4.3 Shapley Additive Explanations

SHAP is an approach developed by Lundberg and Lee (2017) that can provides explanations for any machine learning model. One of its key contributions is the ability to provide explanations for individual predictions. SHAP assigns Shapley values, which are important values from game theory to each feature. These values indicate the contribution of each feature to the final prediction.

Shapley values are represented as an additive feature attribution method based on a linear function and can be expressed as follows:

$$f(x) = g(x') = \phi_0 + \sum_i^M \phi_i x'_i$$

In this formula, f represents the original prediction model, while g represents the explanation model. The term $x = (x_1, x_2, \dots, x_p)$ in the function $f(x)$ represents a vector of p variables or features used as input for the machine learning models. The function $f(x)$ computes the prediction or output of the model. However, the explanation model g use a simplified input x' obtained via a mapping function $x = h_x(x')$. Here, M represents the number of simplified input variables in a single prediction. The term ϕ_0 denotes the base value, which is the average prediction of the model across all inputs, and ϕ_i represents the contribution (or Shapley value) of each individual feature x'_i .

The prediction function $f(x)$ is expressed as the sum of a base value ϕ_0 and the contributions ϕ_i from individual features. In other words, the prediction $f(x)$ is obtained by summing the base value ϕ_0 with the weighted contributions $\phi_i * x'_i$ for each feature. This equation enables the decomposition of predictions into feature-level explanations, facilitating an understanding of the importance of each feature in influencing the model's output. The Shapley value ϕ_i of the feature x'_i can be calculated as follow:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(F - |S| - 1)!}{F!} [f_x(S \cup \{i\}) - f_x(S)]$$

In the formula, F represents the set of all features, and $|S|$ denotes the number of variables in variable subset S that does not include the variable x'_i . For each subset S , the formula calculates the difference between two predictions: $f_x(S \cup \{i\})$ and $f_x(S)$. Here, $f_x(S \cup \{i\})$ represents the model's prediction when the feature x'_i is included along with the features in subset S , while $f_x(S)$ represents the model's prediction with only the features in subset S . The formula further incorporates combination coefficients. The numerator includes $|S|!(F - |S| - 1)!$, which accounts for the number of possible orderings of the features in subset S within the overall set of all features (F). The denominator, $F!$, represents the total number of possible orderings of all features in set F .

The Shapley value ϕ_0 is obtained by summing these contributions over all subsets S . This calculation considers all possible combinations of features and their effects on the model's predictions. It provides a fair and consistent way to attribute importance values to individual features, accounting for their interactions with other features.

In SHAP, global feature importance can be assessed by calculating the average of the absolute Shapley values for each variable across all observations. Variables with higher average Shapley values are considered more important. The global feature importance for the i -th variable, denoted as I_i , is calculated as the average across all observations or the sample size:

$$I_i = \frac{1}{n} \sum_{j=1}^n |\phi_i^{(j)}|$$

Here, n represents the number of observations or the sample size, and $\phi_i^{(j)}$ is Shapley value of variable i in the j -th observation of the sample.

There are a number of methods that use the additive feature attribution method, such as LIME, DeepLIFT and Layer-Wise Relevance Propagation. However, following [Lundberg and Lee \(2017\)](#), each method has its own way of calculating or approximating ϕ_i but only the Shapley value can guarantee the following properties:

Property 1 (Local accuracy): The explanation model $g(x')$ matches the original model $f(x)$ when $x = h_x(x')$.

Property 2 (Missingness): Missingness constraint features ($x'_i = 0$) have no attributed impact.

Property 3 (Consistency): If a model changes so that the marginal contribution of a feature value increases or stays the same (regardless of other features), the Shapley value also increases or stays the same.

The SHAP method contains multiple ways for calculating the Shapley value (ϕ_i) including Kernel SHAP, Deep SHAP and TreeSHAP. TreeSHAP can be applied to understand tree ensemble methods, such as random forest or gradient boosting trees.

Feature importance scores can be measured based on several methods implemented in tree ensemble software packages, including Gain, Split Count and Permutation (Lundberg & Lee, 2018). These methods have been widely used in previous research to explain bankruptcy prediction models (Carmona et al., 2019; Climent et al., 2019; Jabeur et al., 2021; Sigrist & Hirnschall, 2019; Smith & Alvarez, 2022; Son et al., 2019). However, following Lundberg and Lee (2018), the

feature importance scores from the Gain and Split Count are not reliable due to their inconsistency, and only the SHAP and Permutation methods are able to resolve this problem. The difference between the two methods is how they measure feature importance. On the one hand, Permutation feature importance uses model performance. On the other hand, SHAP is based on the magnitude of feature attributions (Molnar, 2020). Consequently, SHAP is clearly the solution for our goal in this study: to display attributions of important financial ratios to the bankruptcy prediction for each company.

5. Results

5.1 Performance of the models

We use average prediction accuracy, Type-I error, Type-II error and AUC (area under the curve) ROC (receiver operation characteristics) in order to compare the performance of our methods.

The average prediction accuracy is calculated by dividing the sum of True Positives and True Negatives by the total number of companies in the test set. The Type-I error is the percentage of failed firms that were wrongly classified in the non-failed firm class. The Type-II error is the percentage of non-failed firms that were incorrectly identified in the bankruptcy firm class. The AUC-ROC curve is widely used for performance measurement of classification models. ROC is a curve with True positive rate on the y-axis and False positive rate on the x-axis, while AUC is the area under the ROC curve representing the degree of separability. The AUC takes values from 0 to 1; a model with a higher AUC value is better at distinguishing between bankrupt companies and healthy companies.

Table 3

Correct classification rates.

	Ensemble models			Single models			Traditional models	
	NGBoost	LightGBM	XGBoost	RF	SVM	NN	DA	LR
1Y	86%	87%	88%	87%	84%	85%	72%	81%
2Y	83.5%	86%	86.5%	85.5%	83%	82.5%	69.5%	74.5%
3Y	82%	83%	83.5%	83.5%	80%	81%	65.5%	72%
4Y	81%	82%	82%	81.5%	80%	80%	67.5%	72.5%
5Y	81.5%	83%	83.5%	82.5%	78.5%	79.5%	70.5%	71.5%
Average	82.8%	84.2%	84.7%	84%	81.1%	81.6%	69%	74%

DA Discriminant Analysis, *LR* Logistic regression, *NN* Neural Networks, *RF* Random Forest, *SVM* Support Vector Machine.

Table 3 presents the correct classification rates for all models. Overall, XGBoost always achieves the best performance over the 1-year to 5-year horizon. Ensemble models provide the highest accuracy but these classifiers do not significantly outperform single models. Logistic regression performs fairly well in the short term but the accuracy decreases rapidly when the horizon prediction exceeds 2 years. These findings correspond with the result of du Jardin (2015). In terms of average correct classification rates, there is only a small difference between model accuracies in a given group.

Table 4

Type-I error and Type- II error.

		Ensemble models			Single models			Traditional models	
		NGBoost	LightGBM	XGBoost	RF	SVM	NN	DA	LR
1 Y	Type-I error	17%	22%	16%	18%	15%	24%	70%	39%
	Type-II error	11%	7%	9%	9%	15%	9%	6%	8%
2 Y	Type-I error	21%	22%	18%	19%	16%	27%	71%	41%
	Type-II error	12%	9%	10%	11%	16%	11%	7%	14%
3 Y	Type-I error	23%	23%	20%	22%	18%	24%	77%	43%
	Type-II error	14%	12%	13%	13%	21%	15%	8%	17%
4 Y	Type-I error	23%	20%	16%	21%	18%	23%	75%	41%
	Type-II error	16%	15%	19%	16%	21%	16%	7%	17%
5 Y	Type-I error	29%	28%	25%	28%	19%	33%	66%	46%
	Type-II error	11%	10%	10%	11%	22%	12%	8%	16%
Average	Type-I error	22.6%	23%	19%	21.6%	17.2%	26.2%	71.8%	42%
	Type-II error	12.8%	10.6%	12.2%	12%	19%	12.6%	7.2%	14.4%

DA Discriminant Analysis, LR Logistic regression, NN Neural Networks, RF Random Forest, SVM Support Vector Machine.

Table 4 illustrates the Type-I and Type-II errors of our classifier models. The inequality between Type-I and Type II errors was found in numerous studies (Alfaro et al., 2008; Nanni & Lumini, 2009; C. Tsai & Wu, 2008). In our study, almost all of these models except SVM identify non-failed firms better than failed firms. It is also the only model that can outperform XGBoost in the Type-I error, although the difference is not significant. On the other hand, XGBoost is much better than SVM in classifying non-failed companies. In the case of bankruptcy, the cost of Type-I error is greater than that of a Type-II error (Altman et al., 1977). Therefore, in terms of Type-I and Type-II errors, we cannot conclude that ensemble models outperform SVM.

Table 5

Area under the ROC curve (AUC).

AUC	Ensemble models			Single models			Traditional models	
	NGBoost	LightGBM	XGBoost	RF	SVM	NN	DA	LR
1Y	93%	94%	94%	93%	92%	91%	73%	84%
2Y	91%	92%	93%	92%	90%	88%	69%	78%
3Y	89%	90%	91%	90%	88%	87%	66%	78%
4Y	87%	90%	90%	89%	87%	88%	68%	78%
5Y	89%	91%	91%	90%	86%	87%	74%	78%

DA Discriminant Analysis, *LR* Logistic regression, *NN* Neural Networks, *RF* Random Forest, *SVM* Support Vector Machine.

Table 5 shows the AUC values of our classifier models from one to five years before the bankruptcy. In terms of AUC, XGBoost shows the best performance in every single year. Almost all of the models have the highest predictive power one year before failure. We also found that ensemble models outperform traditional models. The best AUC values of Logistic Regression and Discriminant Analysis are only 84% and 74%, respectively. Meanwhile the best performance for ensemble models are 94% for LightGBM, XGBoost and 93% for NGBoost and Random forest. The ensemble models group also slightly outperforms single models and has provides stable accuracy over time. The performance of ensemble methods and XGBoost in particular supports previous findings of (Carmona et al., 2019; Climent et al., 2019; Smith & Alvarez, 2022).

5.2 Importance of features with SHAP Summary Plot

For a global interpretation, we employ SHAP Feature Importance (on the left side of Fig.1) and SHAP Summary Plot (on the right side of Fig.1). The SHAP Feature Importance is a visualization tool that enables us to understand the importance of each feature in a machine learning model. This graph displays the features ranked according to their mean absolute SHAP values, which represents their impact on the model's predictions. The SHAP summary plot is another powerful visualization tool that can help understand the contribution of each feature to the output of a black box model.

The left side of Fig. 1 shows the most prominent features that explain the XGBoost model prediction for one year before bankruptcy in our test sample. These features are sorted vertically based on their average impact on the prediction, with the most important feature appearing at the top of the graph. For instance, R23 (tax ratio) is the most significant feature since it has the highest

global impact, which is calculated by the average of the absolute Shapley values of R23 in each firm. The SHAP Feature Importance graph provides a quick and intuitive way to identify the most important features and their impact on the model's predictions. This can be especially useful in finance, where understanding the drivers of a model's predictions is essential for making informed decisions. By using the SHAP Feature Importance graph, we can prioritize our attention on the most significant features and ensure that our decisions are based on the most important variables. Our analysis showed that a number of financial ratios played an important role in predicting failure in our dataset. Specifically, the tax ratio (R23), cost of labor to value-added (R24), asset turnover (R14), interest level (R24), and debtor days (R20) were the most important ratios for predicting failure. This finding is consistent with previous research on corporate bankruptcy prediction, which has identified these ratios as important indicators of financial distress. Interestingly, our results also showed that the most important ratios for the XGBoost models were consistent across different prediction horizons, including the 2-year, 3-year, 4-year, and 5-year horizons. These ratios were the tax ratio (R23), interest level (R24), debtor days (R20), and working capital requirement ratio (R05), as illustrated in Appendix E, Figs. 8-11.

On the right side of Fig. 1, the SHAP summary plot displays the Shapley values for each feature on the x-axis and the corresponding feature values on the y-axis. Each dot is a feature value for one of the companies in our test sample. The values range from high (red) to low (blue). The Shapley value measures the contribution of each feature to the model's output, where a positive value indicates that the feature increases the probability of failure and a negative value indicates that the feature decreases the probability of bankruptcy. By examining the SHAP summary plot, we can identify the most important features for predicting bankruptcy and their direction of impact. The plot enables us to quickly determine which features have the greatest influence on the model's predictions, allowing for an improved interpretation of the model's behavior and increased confidence in its predictions.

The SHAP Summary Plot on the right side of Fig. 1 not only shows us the most important features for predicting bankruptcy but also provides insights into the relationship between feature values and their impact on the prediction. As expected, a low tax ratio (R23), which is often used as a proxy for profitability, is the most important feature in predicting bankruptcy. Additionally, low profitability ratios (R16, R18) are also predictors of high bankruptcy probability. These findings are consistent with previous research that has identified financial profitability ratios and efficiency

ratios as critical factors in predicting corporate failure (Bussmann et al., 2021; Jabeur et al., 2021; Schalck & Yankol-Schalck, 2021; Sigrist & Hirnschall, 2019).

Our analysis of the SHAP summary plot reveals that working capital management is also crucial for predicting bankruptcy. Specifically, high/low values of creditor/debtor days (R20, R21) are among the most important features. In a study with a 2-year horizon, [Jabeur et al. \(2021\)](#) found that the collection period is one of the most important ratios in the CatBoost model. High values of the working capital requirement ratios (R05 and R09) and labor cost (R22) also complement the prediction of bankruptcy.

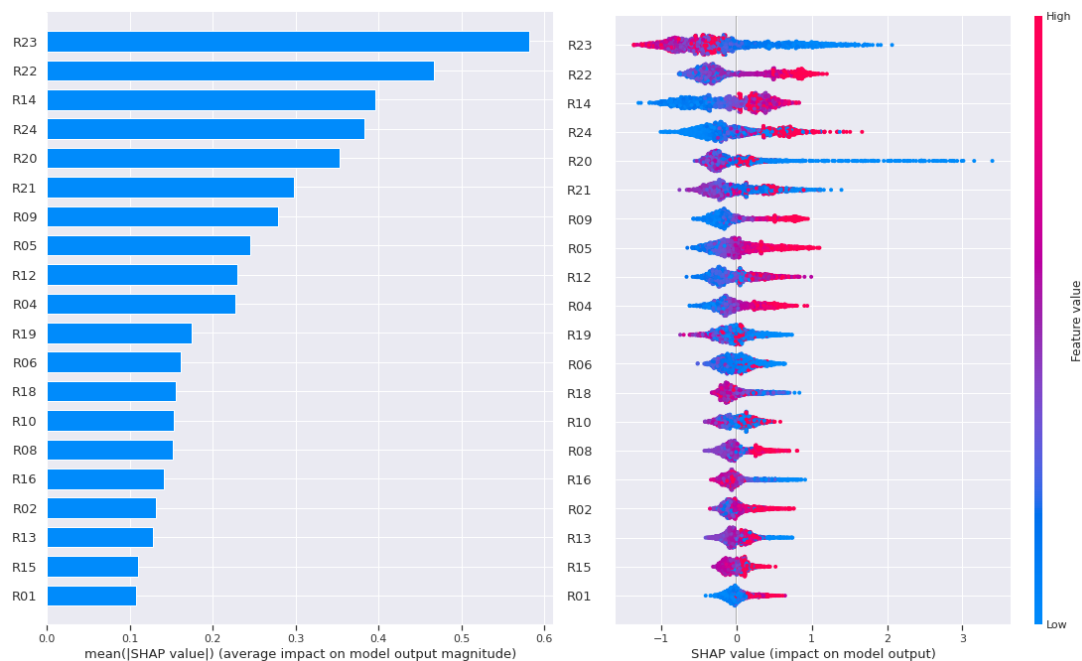


Fig. 1 On the left, SHAP Feature Importance for a 1-year horizon. On the right, SHAP Summary Plot for a 1-year horizon.

In our sample, companies with high asset turnover (R12) or high value-added creation (R14) are more likely to increase the probability of bankruptcy. The asset turnover ratio is a significant component of the Dupont chain of ratios, and higher turnover is generally associated with higher profitability. Altman (1968) also used this as one of five ratios to predict bankruptcy. However, the comparison is only valid within the same industry. For example, wholesalers have high turnover and low margins, while the luxury sector has low turnover and high margins. Additionally, the mean and median turnover ratios of failed firms are higher than those of non-failed firms in our sample (Appendix B, Table 6-10). If sales generate high value-added (R14),

this is usually indicative of a differentiation strategy, which Porter (2008) considers a generic source of competitiveness. However, it can also be a sign of a longer production cycle and potential liquidity problems. Yıldırım et al. (2021) found a similar surprising finding in their SHAP summary plot, where high gross profit to net sales is associated with high Shapley value. We analyze the impact of asset turnover (R12) and value-added creation (R14) on bankruptcy risk in more detail in the next section with SHAP dependence plots. It's also important to look at how debt affects the chance of going bankrupt. As expected, high interest (R24) makes the chance of failing higher. Smith and Alvarez (2022) found that the debt to interest expense ratio (CL.FinExp) and the interest coverage ratio (EBIT.FinExp) are among the variables that contribute the most to the XGBoost model with data in Spain.

We also use the SHAP Summary Plot with LightGBM, our second-most accurate model. Figure 12 in Appendix E shows the most important features of LightGBM one year prior to failure. The most important financial ratios of LightGBM are similar to those of XGBoost, including R23, R24, R20, and R12.

5.3 SHAP Dependence Plot

Prior research primarily utilized Partial dependence plots to illustrate the relationship between a feature and the model's output prediction (Jabeur et al., 2021; Sigrist & Hirnschall, 2019; Smith & Alvarez, 2022; Son et al., 2019). However, SHAP dependence plots provide additional information by showing the local effect of a feature as opposed to the average effect across all feature values (Molnar, 2020). SHAP dependence plots can also reveal non-monotonic relationships between a feature and the model's output, which is useful for identifying specific instances in which the relationship is non-linear or complex. While SHAP dependence plots provide insights into the relationship between features and model output, they do not establish causality because machine learning algorithms are designed to exploit associations rather than model causal relationships (Molnar et al., 2022).

In this section, we use a SHAP dependence plot to visualize the relationship between the most important ratios, R23 (tax ratio), R22 (cost of labor), R14 (value-added creation), R24 (interest level), R20 (debtor days), R21 (days payable outstanding), R09 (working capital requirement to total sales), R05 (working capital requirement to current assets), and their Shapley values.

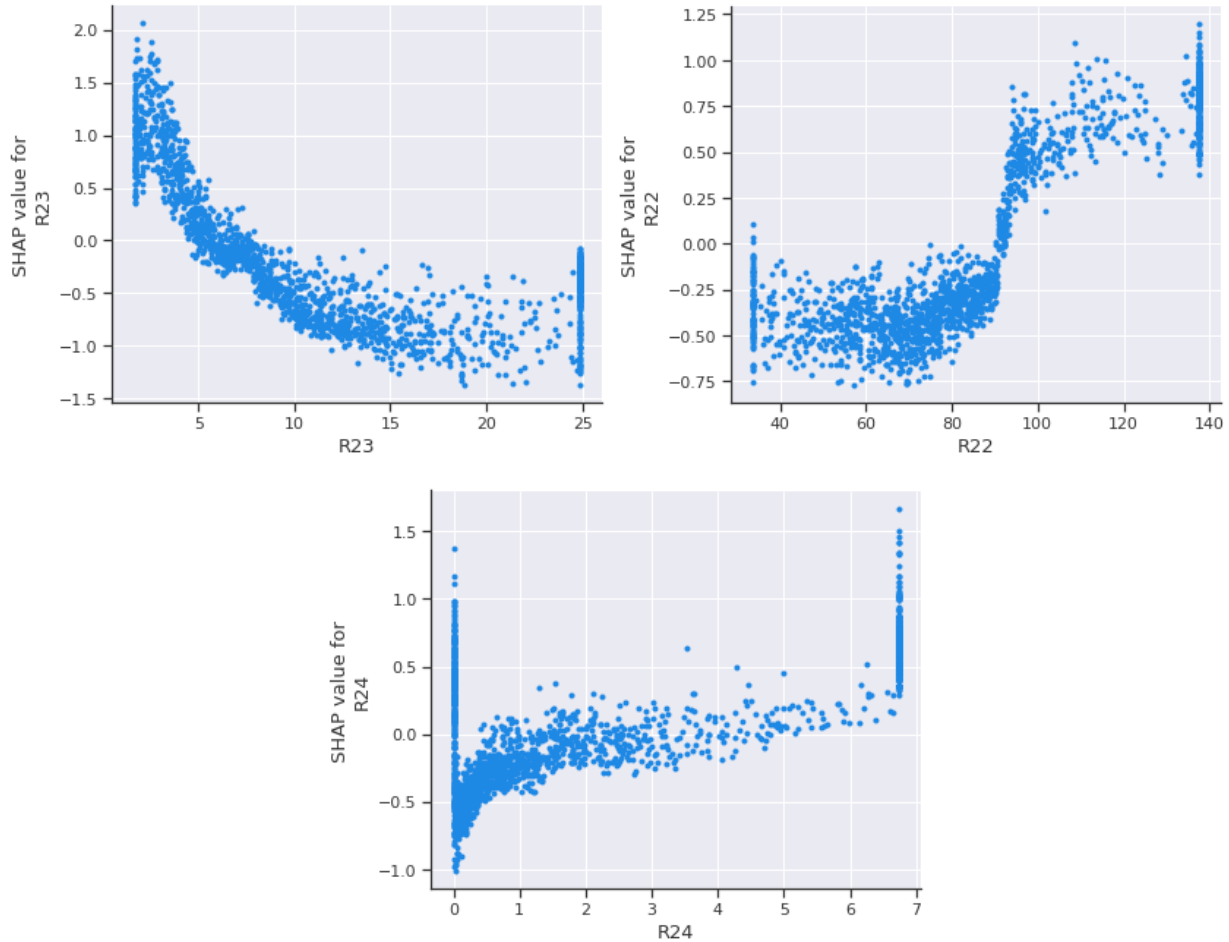


Fig. 2. SHAP dependence plot of R23, R22 and R24 for a 1-year horizon.

The value-added generated by a company is typically determined by subtracting its cost of goods sold from its revenue. This value is then allocated to various stakeholders, such as employees, the state, lenders, and investors, in the form of wages, taxes, interest, and profits. Figure 2 illustrates the impact of different components of value-added on failure prediction, namely R23 (tax ratio), R22 (cost of labor to value-added), and R24 (interest level), which respectively represent the contributions of the state, employees, and lenders in the value-added distribution. These ratios are identified as the most important, second-most important, and fourth-most important features, respectively, in the decision-making process of the XGBoost model.

The graph located in the top left of Figure 2 illustrates the influence of the state's contribution to value-added (R23) on the probability of bankruptcy. In general, a higher proportion of the state's involvement in value-added is considered a positive indicator for companies, as it reflects a healthy financial position. Within our test set, a proportion of state ranging from 5% to 25% is associated

with a reduced probability of bankruptcy, while R23 values below 5% are indicative of a higher risk of failure.

The graph displayed in the top-right corner of Figure 2 indicates a non-linear relationship between the ratio of labor costs to value-added (R22) and the risk of bankruptcy. Specifically, when the labor cost-to-value-added ratio exceeds 90%, there is a significant increase in the probability of bankruptcy. This is likely due to the fact that a high ratio of labor costs to value-added indicates low profit margins, which may not be sustainable over the long term. Conversely, a ratio of labor costs to value-added lower than 80% is associated with a lower risk of failure, as it indicates that the company is generating more value per unit of labor cost, leading to higher profitability and financial stability. These results suggest that there may be an optimal range of the labor cost-to-value-added ratio that minimizes the risk of bankruptcy, and companies should strive to balance labor costs and value-added to achieve financial stability. It is important to note that this optimal ratio may vary depending on the industry and market conditions, as well as other factors such as the company's size and capital structure.

The graph on the bottom of the Fig.2 illustrates the complex relationship between R24 (Interest/Value-Added) and its Shapley value. In our test sample, a low level of interest (less than 1% of value-added) contributes to a decrease in the probability of bankruptcy. However, a very low level of interest (around 0%) can have either a low or high Shapley value, which means it can either increase or decrease a company's failure risk. These two results indicate that there are several configurations (situations) that lead to failure: in one case, it is high leverage, while in the other, it is the inability to raise funds. This contradiction of low leverage also appeared in the study by Bussmann et al. (2021) using SHAP for XGBoost. The study shows that the ratio of total assets to total liabilities is the most explanatory ratio and has a positive importance with predicting failure.

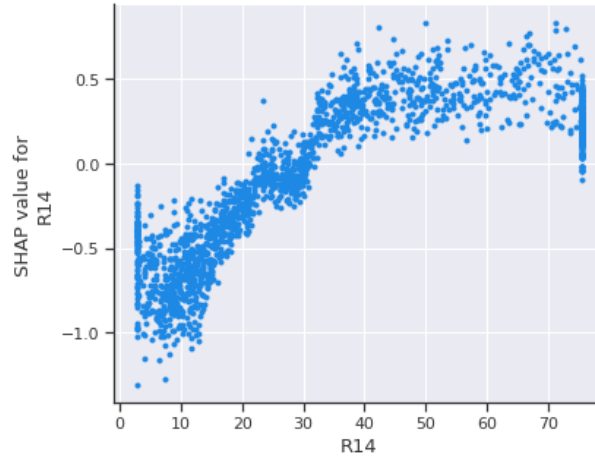


Fig. 3. SHAP dependence plot of R14 for a 1-year horizon.

The (Value-Added/Total Sales) ratio is a widely used measure of a company's efficiency in generating value per unit of sales. Nevertheless, its impact on bankruptcy risk is not straightforward. In our sample of mixed industries, we found that a ratio lower than 20% contributes to a lower failure risk, while a ratio from 20% to 80% is surprisingly associated with a higher bankruptcy risk. However, a (Value-Added/Total Sales) ratio above 80% seems to contribute to a lower failure risk.

One plausible explanation for the finding that a (Value-Added/Total Sales) ratio lower than 20% contributes to a lower failure risk is that such companies tend to have lower fixed costs, which can provide greater flexibility and enable them to respond more quickly to changes in the market environment. Another possible explanation is that companies with lower ratios may be operating in less competitive industries with higher profit margins, which can contribute to their financial stability. These factors may enable companies with lower ratios to better withstand adverse economic conditions and maintain their solvency.

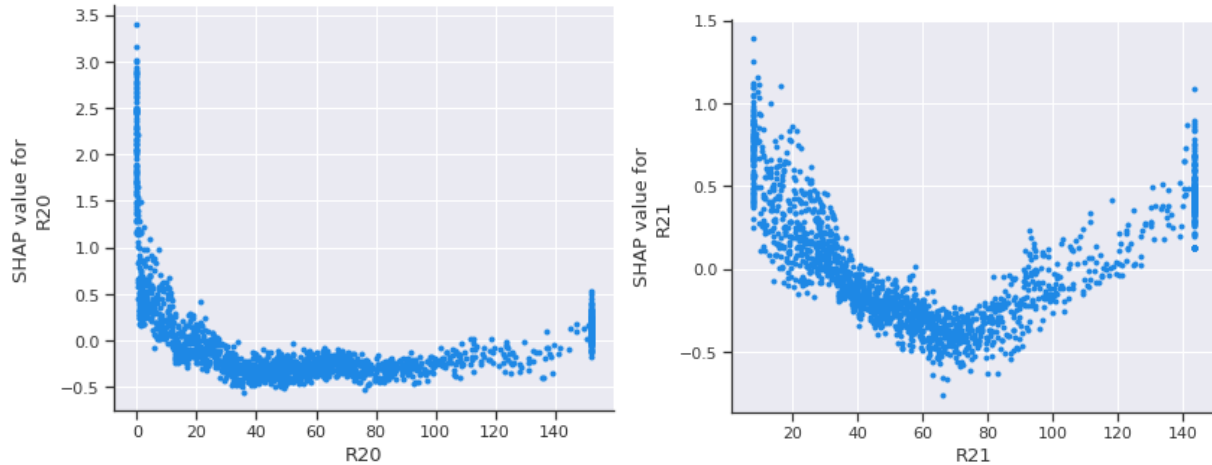


Fig. 4. SHAP dependence plot of R20 and R21 for a 1-year horizon.

The graph on the left side of Fig.4 demonstrates the impact of debtor days (R20) on its Shapley values. Our analysis shows that companies with debtor days ranging from 20 to 200 days have a decreased likelihood of bankruptcy. Conversely, a lengthy collection period (more than 200 days) raises the risk of bankruptcy. Similarly, a very short collection period also leads to a higher failure risk. It is worth noting that R20 (debtor days) is just one factor among several that contribute to a company's financial health and should be assessed in conjunction with other relevant metrics.

The relationship between the days payable outstanding (DPO or R21) and the corresponding SHAP values is depicted in the graph on the right of Fig.4. Our findings suggest that a firm's failure risk may increase when it takes more than 140 days to pay its suppliers or when the firm is asked to pay its bills very quickly (less than 20 days), as indicated by positive SHAP values. Additionally, our analysis revealed that a credit payable period ranging from 30 to 110 days was associated with a reduced probability of bankruptcy. However, it is worth noting that the specific factors driving this relationship may differ depending on the industry. For instance, in the construction industry, contractors typically pay their suppliers after the project is completed. Therefore, longer payables periods may not necessarily indicate poor financial management or an increased risk of bankruptcy. These findings emphasize the importance of carefully managing payables periods while considering industry-specific factors to avoid the risk of financial distress and, ultimately, bankruptcy.

The two graphs in Fig.5 clearly demonstrate a negative relationship between the working capital requirement (R09, R05) and bankruptcy risk. In general, a low requirement for working capital is perceived as a positive indicator of a company's financial health. In our sample, we observed that

a negative working capital requirement to current assets (R05) has a negative SHAP value, which significantly reduces the likelihood of bankruptcy. The graph also shows that a working capital requirement to current assets (R05) of 0.5 represents a critical threshold, above which the risk of bankruptcy increases. These results emphasize the importance of effective management of working capital to maintain a company's financial health and mitigate the risk of bankruptcy.

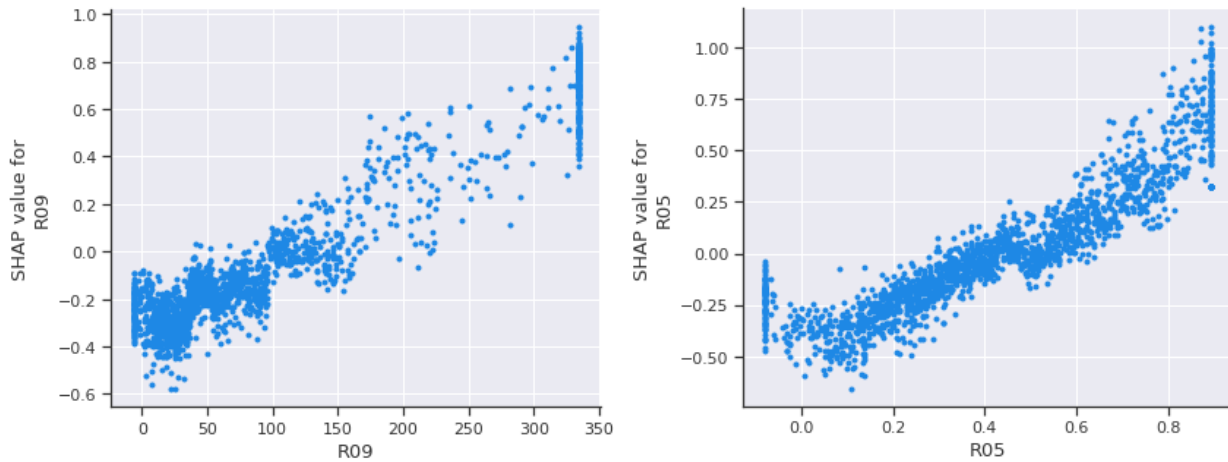
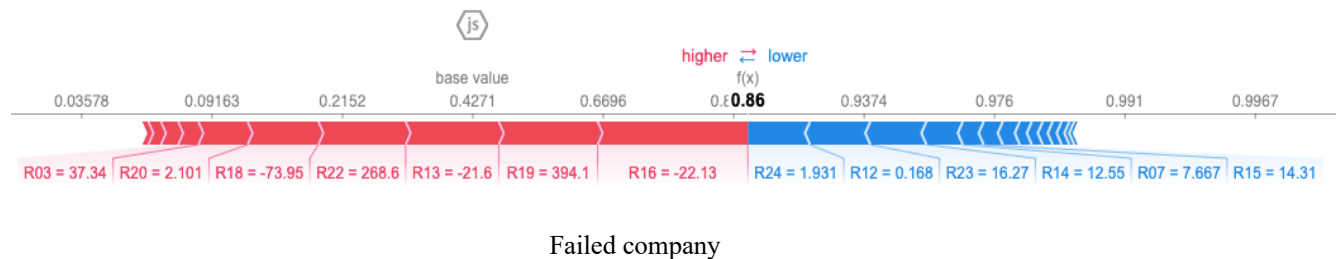
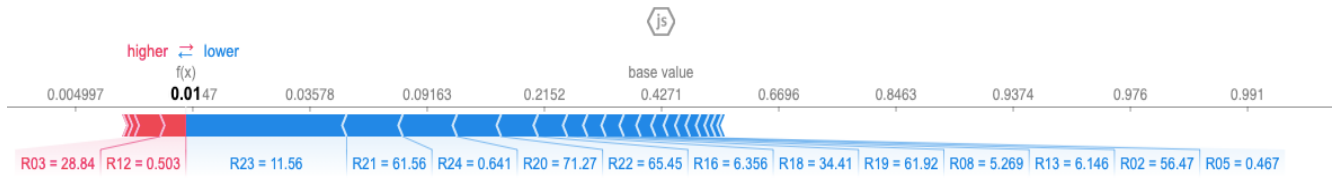


Fig. 5. SHAP dependence plot of R09 and R05 for a 1-year horizon.

Other SHAP dependence plots for the remaining important ratios are presented in Appendix F. We can see that there is considerable dispersion in the interaction between ratios, and the linear relationships are not always presented. It is difficult to confirm a relationship between these variables and their impacts on the prediction (Shapley value), but we do have an overview of feature values and their SHAP values. This finding is consistent with Sigrist and Leuenberger (2023) where the most important variables have non-linear effects. This explains why machine learning models outperform traditional models, which are based on linear regression.

5.4 SHAP force plot for local explanations





Non-failed company

Fig. 6. SHAP force plot for a non-failed company and a failed company one year before failure.

The most important application of SHAP is represented in Fig. 6 SHAP force plot. The figure answers the question “Why does the XGBoost model predict that a company fails?”. The SHAP force plot allows us to clearly explain the bankruptcy prediction for every firm in the dataset. This implementation therefore allows banks and credit institutions to explain their decisions using machine learning and comply with implementation of the right to explanation in the United States and Europe.

The SHAP force plot is used to visually represent the important features that influence the estimation of a company's bankruptcy probability. In Fig. 6, local explanations for two individual instances in the XGBoost model are presented. The prediction process begins with a base value of 0.4271, which represents the average of all predictions.

In the SHAP force plot, each feature is represented as an arrow. The color scheme is used to indicate the contribution of each feature to the bankruptcy risk. The color red signifies a positive contribution, suggesting that the feature value increases the probability of bankruptcy. Conversely, the color blue indicates a negative contribution, implying that the feature value reduces the bankruptcy risk. Features with a more substantial impact on the score are positioned closer to the dividing boundary between red and blue, and the size of the bar represents the magnitude of that impact. To calculate the final predicted bankruptcy probability, the sum of all attribute attribution values is computed and added to the base value.

By analyzing the SHAP force plot, valuable insights can be gained regarding the relative importance of different features in determining a company's bankruptcy probability. This plot aids in understanding how each feature contributes to the overall prediction, whether positively or negatively.

In the top of Fig. 6, the company has an 86% chance of going bankrupt. The most important features that push the failure probability from the base value 42.71% to 86% are as follows:

R16 (Profit or Loss before Tax/Operating Revenue)*100: -22.13

R19 (Stocks/Total sales)*360 (days): 394.1

R13 (EBIT/Operating Revenue)*100: -21.6

R22 (Cost of Labor/Value-Added)*100: 288.6

R18 (Profit or Loss/Capital Employed)*100: -73.95

R20 (Debtors/Operating Revenue)*360 (days): 2.1

With a SHAP force plot, we can see why XGBoost predicts that the company will be bankrupt. This company suffers losses and inefficiencies (R16, R13 and R18 are negative). A high cost of labor exceeding value-added by almost 2.9 times is another factor contributing to the failure prediction. In addition, this company has a very strict credit policy of only 2.1 days that might have a negative influence on sales and cause a very high DSI of 394.1 days compared with the mean of 31.2 days and 3.8 days for non-failed and failed firms, respectively (Appendix B, Table 6).

In the bottom of Fig. 5, the non-failed case has only 1% of failure prediction. The financial variables that have a strong impact on low failure probability are as follows:

R23 (Taxes/Value-Added)*100 : 11.56

R21 (Creditors/Operating Revenue)*360 (days): 61.56

R24 (Interest/Value-Added)*100 : 0.641

R20 (Debtors/Operating Revenue)*360 (days): 71.27

R22 (Cost of Labor/Value-Added)*100: 65.45

R16 (Profit or Loss before Tax/Operating Revenue)*100: 6.356

In our sample, the medians of R16 for failed and non-failed firms are 1.2% and 2.5% respectively (Appendix B, Table 6). This is therefore a profitable and efficient company which explains why the firm is able to contribute 11.56% of its value-added to the government compared with the median of only 5% for failed firms and 11% for non-failed firms (Appendix B, Table 6). Regarding labor cost management, 65.45% of value-added is contributed to employees in comparison with the median of 74.9% for non-failed firms and 88.4% for failed firms (Appendix B, Table 6). In terms of credit policy, the company allows 71.27 days for its clients to pay their bills, which helps the firm have a DSI of only 61.92 days (R19).

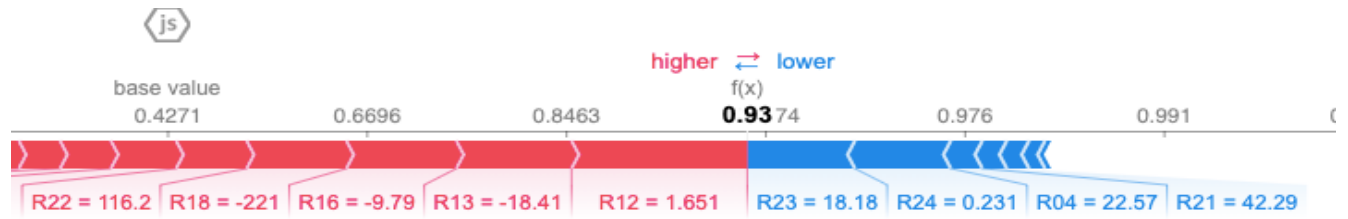


Fig. 7. SHAP force plots for another failed company one year before failure.

Fig. 7 illustrates another case of bankruptcy. The financial ratios that increase the probability of failure are presented below.

R12 (Total Sales/Total Assets) = 1.651

R13 (EBIT/Operating Revenue)*100 = -18.14

R16 (Profit or Loss before Tax/Operating Revenues)*100 = -9.79

R18 (Profit of Loss/Capital Employed)*100 = -221

R22 (Cost of Labor/Value-Added)*100 = 116.2

The use of financial ratios in bankruptcy prediction varies across different companies, as shown in Fig. 6, where the failed company is found to have a different set of important features than the company being analyzed in Fig. 7 (R16, R19, R13, R22, R18, R20 vs. R12, R13, R16, R18 and R22). Additionally, SHAP analysis reveals that certain ratios, such as R16 and R12, may not be the most important features globally, but may be crucial for predicting the failure of a subset of companies. The local explanation provided by SHAP, as demonstrated in Fig. 6 and 7, offers clear and compelling evidence of its benefits. Specifically, the SHAP local explanation identifies the ratios that increase the prediction of failure, as well as those ratios that decrease the bankruptcy risk for each company. Moreover, the Shapley values assigned to each ratio are unique to each company, enabling practitioners to understand why a particular company is failing by identifying the important features that contribute to the prediction of failure. This customization of important features is a significant advantage of using SHAP, as it provides practitioners with a clear and intuitive way to interpret and explain the behavior of their machine learning models in a local context. By identifying the specific features that contribute to a particular prediction, SHAP empowers practitioners to make more informed decisions and take appropriate actions to mitigate the risk of failure for individual companies.

6. Conclusions

In this research, we evaluated the bankruptcy prediction performance of eight binary classification models on French corporate data from 2013 to 2017. Ensemble models provide the highest accuracy and most stable results based on AUC-ROC and the correct average prediction accuracy. This finding supports previous research showing that ensemble-based techniques perform better than single classifiers. However, SVM still has a performance comparable to the ensemble models based on Type-I and Type-II errors. As XGBoost is the best performer among our ensemble models, we employed SHAP in order to illustrate the important features and their relation to the prediction of corporate failure. Furthermore, SHAP can also provide an analysis of important features in a single prediction. This application assists us in determining which variables contribute to a company's bankruptcy prediction.

Our study has some theoretical implications. Firstly, we show that using the SHAP explanation, we can examine what takes place inside a black box bankruptcy prediction model. This application increases the credibility of the black box model and allows us to gain insight into XGBoost's forecast. Secondly, we contribute to the literature by identifying a number of important financial ratios and visualizing their complex relationship with failure prediction. In terms of local interpretation, we found that a financial ratio can contribute differently to the bankruptcy prediction of each company. Finally, we confirm the effectiveness of several ensemble methods in bankruptcy prediction for companies from various industries. These methods are easy to implement and can model non-linear relationships between financial ratios and the prediction.

Our experimental results have several practical implications for credit management in banks and financial institutions. Ensemble classifiers allow credit managers to effectively detect financial distress up to 5 years earlier. These classifiers are easy to replicate because they do not require complicated data preparation steps. Moreover, practitioners using SHAP will be able to give a full explanation for each decision made. As a result, companies applying for a loan can fully understand the decision made by credit institutions.

Our study has several limitations that can be improved by future research. Firstly, our dataset consists of only 24 yearly financial ratios. Other non-financial variables could be used to improve prediction performance, such as textual data or corporate governance indicators. Secondly, our dataset includes almost equal numbers of failed and non-failed firms, which rarely occurs in practice. Finally, we use SHAP to create a transparent relationship between feature value and

failure prediction, but these relationships cannot be used to derive causal insights. It is appropriate to explore some useful tools to estimate the causal effects between financial ratios and bankruptcy.

Appendix A. Main studies dealing with explaining and interpreting ensemble-based models

Studies	Method	Important variables
Jones et al. (2017)	Relative variable importance	Annual Growth in Capital Expenditure, Annual Growth in Leverage Free Cash Flow, Earnings per Share, Annual Growth in Revenue, Total Asset Write Downs to Total Equity
Son et al. (2019)	Feature importance score Partial dependence plot	Cannot be disclosed due to confidentiality
Sigrist and Hirnschall (2019)	Feature importance score Partial dependence plot Local partial dependence plot	Rating no. 2, Retaining earnings to assets , repayment score, current ratios (the names of the second and third important variables cannot be disclosed)
Perboli and Arabnezhad (2021)	SHAP	The five most important variables belong to two feature types: Cost/Debt, Revenue/Profit
Jabeur et al. (2021)	Feature importance score Partial dependence plot	1-year horizon: Return on shareholder funds, Return on capital employed , Return on total assets 2-year horizon: Solvency ratio, Return on shareholder funds, Return on capital employed 3-year horizon: Stock turnover, Solvency-ratio, Credit period
Bussmann et al. (2021)	SHAP	total assets to total liabilities, EBIDTA to interest coverage ratio, profit before taxes plus interest paid then divided by total assets, trade receivables divided by operating revenues, profit or loss after tax divided by shareholder funds
Smith and Alvarez (2022)	Feature importance score Partial dependence plot	1-year horizon: TL.TA, logTA, logSALES, SALES.EBIT, CL.FinExp 2-year horizon: logTA, logSALES, TL.TA, TL.EQ, EBIT.FinExp 3-year horizon: logSALES, logTA, TL.TA, CL.FinExp, DEBTORS.SALES 4-year horizon: logSALES, TL.TA, logTA, CL.FinExp, DEBTORS.SALES
Yildirim et al. (2021)	SHAP	Total loans to Total Assets, Own funds to Total Assets, Cash Ratio, Net sales + Stock changes/Stocks, Gross Profit to Net Sales
Sigrist and Leuenberger (2023)	SHAP	Relative size, Excess return, Idiosyncratic stock volatility, Market value/Total assets, EBIT/Total Assets
Schalck and Yankol-Schalck (2021)	SHAP	Self-ownership, Capital (Equity capital in euros), Revenues (Turnover of the year in euros), Region, Employees (Number of staff members)

Appendix B : Descriptive statistic of variables

See table 6, 7, 8, 9, 10.

Table 6

Summary of descriptive statistic of variables for a 1-year horizon.

Variable	Mean		St. Dev.		Median		Min		Max	
	NF	F	NF	F	NF	F	NF	F	NF	F
R01	6.0	45.9	67.7	1890.6	1.8	1.8	0.0	0.0	3103.6	107904.7
R02	43.7	38.9	74.3	123.8	45.4	51.1	-3621.9	-3324.0	100.0	100.0
R03	18.7	25.8	33.9	91.0	11.5	9.5	0.0	0.0	1978.7	2853.3
R04	33.3	42.8	21.4	34.9	31.7	45.4	-92.3	-99.4	99.0	100.0
R05	0.4	0.4	0.3	1.2	0.4	0.5	-7.1	-46.5	1.9	1.8
R06	1.9	4.5	1.5	8.8	1.6	2.0	0.1	0.0	59.2	97.3
R07	0.6	2.7	80.1	117.8	0.7	0.0	-4431.4	-2006.6	1022.5	5025.0
R08	3.5	1.3	7.0	19.9	2.6	2.2	-94.3	-99.7	95.4	99.6
R09	-46.3	2645.4	7855.5	91443.7	44.5	77.9	-580696.0	-33466.3	1462.6	4760439.0
R10	51.5	91.2	54.5	129.5	38.1	49.3	0.0	0.0	744.7	983.8
R11	8.5	7.8	101.1	44.2	1.2	0.9	-60.4	-44.7	3513.4	1235.6
R12	0.4	0.4	0.3	0.5	0.3	0.3	-0.9	-1.5	4.8	7.3
R13	3.5	-0.4	6.7	19.7	2.5	1.3	-70.1	-94.7	94.5	90.1
R14	20.8	-39.1	271.1	3202.5	19.0	36.4	-19981.5	-172572.1	136.8	1753.4
R15	19.5	-11.7	651.6	3705.8	15.3	7.0	-10715.9	-130029.3	46700.1	141520.7
R16	3.7	-0.1	7.5	20.3	2.5	1.2	-96.1	-98.7	94.5	98.2
R17	19.2	21.8	1498.8	1258.1	15.9	4.4	-81628.5	-44055.9	11814.7	33863.1
R18	12.6	-33.7	155.8	978.1	12.8	3.1	-10976.9	-50944.9	737.7	556.1
R19	47.9	46.2	56.3	105.2	31.2	3.8	0.0	0.0	961.3	999.7
R20	51.6	70.1	41.8	107.6	46.2	42.1	0.0	0.0	607.2	983.3
R21	56.5	73.2	39.7	104.8	51.2	43.0	0.0	0.0	933.1	963.3
R22	91.0	151.8	996.6	1911.4	74.9	88.4	0.0	0.0	73198.3	97029.3
R23	12.7	12.9	108.8	85.9	9.5	5.0	0.0	0.0	7812.1	3728.0
R24	5.3	16.0	241.2	576.7	0.6	0.2	0.0	0.0	17718.0	28008.0

F: Failed company, *NF*: Non-failed company. Variables are defined in Table 3.

Table 7

Summary of descriptive statistic of variables for a 2-year horizon.

Variable	Mean		St. Dev.		Median		Min		Max	
	NF	F	NF	F	NF	F	NF	F	NF	F
R01	5.8	10.1	65.3	125.2	1.8	1.8	0.1	0.1	2923.2	6377.2
R02	44.6	39.0	47.8	310.7	45.3	51.9	-1613.1	-17173.1	100.0	135.3
R03	19.0	28.3	27.9	286.9	11.9	10.4	0.0	0.0	1306.0	16152.6
R04	33.4	43.2	21.3	32.0	31.9	43.7	-96.3	-91.7	98.8	100.0
R05	0.4	0.4	0.3	0.7	0.4	0.5	-10.8	-19.6	1.6	2.4
R06	2.0	4.2	1.9	7.7	1.6	2.0	0.1	0.1	86.4	98.7
R07	4.5	1.7	92.9	71.5	0.7	0.0	-1000.9	-1468.3	5952.5	2147.0
R08	3.5	3.2	6.6	18.5	2.7	2.8	-83.2	-96.6	92.9	99.6
R09	213.8	3764.6	11216.1	175044.9	44.1	74.4	-1315.6	-4290.0	823105.5	9602889.0
R10	51.5	81.2	57.8	118.8	37.1	43.1	0.0	0.0	895.5	974.9
R11	7.5	8.2	77.2	62.3	1.2	1.1	-65.8	-189.2	3067.1	2832.0
R12	0.4	0.5	0.3	0.6	0.3	0.3	-0.9	-5.4	4.1	7.7
R13	3.4	1.5	6.8	17.8	2.4	1.8	-81.1	-96.6	93.4	94.1
R14	24.8	3.6	18.9	1125.4	19.3	37.5	-75.4	-46761.8	143.4	237.0
R15	10.6	-23.8	213.6	649.2	14.8	8.9	-8944.9	-32078.0	9111.6	1121.6
R16	3.6	1.8	7.4	18.8	2.4	1.7	-81.3	-98.9	94.1	98.4
R17	42.8	78.0	652.3	1117.8	15.5	6.9	-39377.2	-12257.1	12219.1	31728.1
R18	14.1	4.1	67.9	172.8	12.8	5.5	-3912.3	-5099.3	483.6	5528.3
R19	48.0	46.6	59.2	100.3	31.1	5.5	0.0	0.0	969.6	973.0
R20	51.2	63.7	42.6	94.6	45.6	41.4	0.0	0.0	701.4	980.4
R21	56.5	64.4	39.6	87.9	50.9	41.2	0.0	0.0	895.1	978.3
R22	80.1	119.5	136.4	594.4	75.2	86.2	0.0	0.0	6872.1	28904.9
R23	12.1	11.1	62.7	72.8	9.4	4.8	0.0	0.0	4416.0	3300.0
R24	2.2	4.7	10.4	74.4	0.7	0.3	0.0	0.0	373.0	3237.4

F: Failed company, *NF*: Non-failed company. Variables are defined in Table 3.

Table 8

Summary of descriptive statistic of variables for a 3-year horizon.

Variable	Mean		St. Dev.		Median		Min		Max	
	NF	F	NF	F	NF	F	NF	F	NF	F
R01	5.7	7.5	93.7	54.0	1.8	1.9	0.1	0.0	6455.5	2420.6
R02	43.9	41.7	53.8	231.8	45.6	51.1	-2020.2	-12831.7	99.6	150.3
R03	19.3	27.0	34.1	189.8	11.5	11.2	0.0	0.0	1710.5	10664.9
R04	33.2	41.8	21.7	30.7	31.9	42.3	-98.6	-99.6	99.8	100.0
R05	0.4	0.3	0.3	2.4	0.4	0.5	-5.6	-111.8	1.4	3.3
R06	2.0	3.8	2.0	7.0	1.6	2.0	0.2	0.0	83.1	99.9
R07	1.8	-9.7	132.5	626.0	0.7	0.1	-8955.0	-36729.0	2735.7	814.8
R08	3.4	3.7	6.4	16.3	2.6	3.0	-73.4	-97.0	65.1	96.5
R09	63.3	426.1	221.5	9265.4	43.1	70.0	-688.3	-4344.6	13132.7	494634.7
R10	50.2	74.7	56.2	108.5	37.1	41.8	0.1	0.0	930.2	957.9
R11	6.7	7.8	81.1	34.0	1.2	1.5	-590.4	-580.6	4501.6	741.9
R12	0.4	0.5	0.3	0.5	0.3	0.4	-1.5	-2.2	4.1	6.7
R13	3.2	1.9	6.9	16.5	2.4	2.0	-85.5	-99.5	92.8	92.7
R14	17.9	32.0	502.6	259.2	19.2	37.6	-36881.9	-14243.3	149.0	396.0
R15	6.1	-23.2	367.2	869.9	14.9	9.1	-25278.9	-34771.8	666.3	9529.9
R16	3.4	2.3	7.5	17.0	2.4	1.8	-92.5	-99.5	92.8	99.4
R17	40.1	87.4	738.4	1792.0	15.8	7.5	-30964.8	-4207.3	16416.0	89343.2
R18	14.8	-9.8	50.1	658.9	13.0	6.3	-1819.3	-37629.3	632.2	2247.6
R19	47.4	48.4	58.8	103.4	30.0	6.9	0.0	0.0	953.3	903.1
R20	49.9	57.8	40.2	78.1	44.7	40.8	0.0	0.0	605.7	974.8
R21	55.2	57.9	38.1	73.1	49.1	40.7	0.0	0.0	737.8	965.0
R22	83.5	125.3	358.9	815.5	74.9	86.6	0.0	0.2	25042.0	35535.2
R23	11.6	10.3	32.5	66.4	9.4	4.8	0.0	0.0	2091.0	3123.1
R24	2.5	4.0	18.4	65.4	0.7	0.4	0.0	0.0	831.0	3458.4

F: Failed company, *NF*: Non-failed company. Variables are defined in Table 3.

Table 9

Summary of descriptive statistic of variables for a 4-year horizon.

Variable	Mean		St. Dev.		Median		Min		Max	
	NF	F	NF	F	NF	F	NF	F	NF	F
R01	4.4	7.8	39.7	103.2	1.7	1.9	0.0	0.1	2713.2	5936.5
R02	42.5	28.3	96.2	1108.3	45.6	51.3	-4952.9	-65350.0	100.0	185.7
R03	19.6	42.7	33.9	1108.3	11.5	11.9	0.0	0.0	1670.8	65450.0
R04	32.7	40.6	21.7	30.0	31.3	40.1	-93.1	-98.1	100.0	100.0
R05	0.3	0.4	1.6	1.1	0.4	0.5	-108.9	-39.3	1.5	24.9
R06	1.9	3.6	2.2	6.5	1.6	1.9	0.0	0.0	90.0	87.0
R07	1.1	2.7	79.4	40.6	0.6	0.2	-2767.5	-1287.5	1982.3	945.2
R08	3.2	4.0	7.1	15.9	2.5	3.0	-95.1	-99.9	68.5	92.2
R09	83.9	6271.6	1259.1	362450.9	41.7	66.0	-4070.6	-4698.4	77611.7	21710777.0
R10	50.4	73.1	65.3	107.8	36.2	40.4	0.0	0.0	999.9	978.8
R11	8.2	8.9	99.4	38.4	1.3	1.6	-679.9	-71.9	5288.4	1156.4
R12	0.4	0.6	0.3	0.5	0.3	0.4	-2.3	-5.3	5.3	6.7
R13	3.2	2.3	6.7	15.9	2.3	2.0	-67.2	-93.9	91.9	95.7
R14	-23.7	-0.5	3546.7	2314.0	19.5	38.2	-259441.5	-138578.6	140.6	173.2
R15	0.6	-5.8	826.6	211.5	13.7	9.4	-60105.3	-5803.4	1378.9	1111.1
R16	3.3	2.4	7.6	16.6	2.3	1.9	-75.7	-96.6	92.1	96.4
R17	206.9	90.6	10555.8	1259.1	15.2	8.8	-40945.9	-12752.3	749488.0	29357.5
R18	14.5	-76.3	193.1	4622.4	12.7	7.1	-10647.6	-275100.0	7481.4	831.1
R19	47.1	47.3	59.5	100.5	30.0	6.8	0.0	0.0	917.0	989.1
R20	51.1	56.8	45.8	78.1	45.1	39.4	0.0	0.0	979.2	999.9
R21	55.1	55.8	39.6	64.8	48.6	40.8	0.0	0.0	774.3	967.9
R22	90.0	100.8	902.4	204.9	75.8	85.9	0.0	0.0	65604.4	8387.4
R23	12.3	8.8	59.7	35.2	9.7	5.0	0.0	0.0	4281.2	1836.4
R24	3.2	3.0	35.1	47.9	0.9	0.4	0.0	0.0	1849.3	2663.3

F: Failed company, *NF*: Non-failed company. Variables are defined in Table 3.

Table 10

Summary of descriptive statistic of variables for a 5-year horizon.

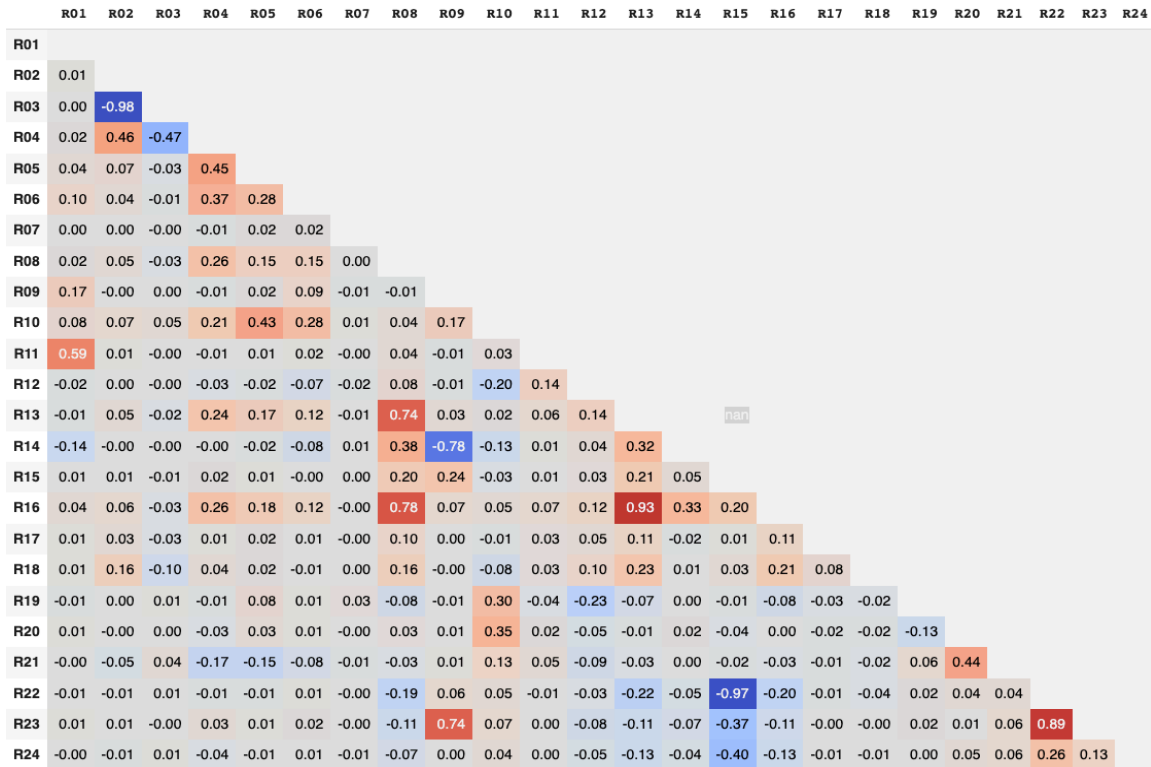
Variable	Mean		St. Dev.		Median		Min		Max	
	F	NF	F	NF	F	NF	F	NF	F	NF
R01	4.1	5.9	15.5	24.3	1.7	1.8	0.0	0.0	516.4	908.0
R02	43.4	42.9	76.8	233.5	44.7	51.1	-4039.9	-12602.8	129.7	333.7
R03	20.9	25.0	52.2	49.2	12.9	13.8	0.0	0.0	3388.0	2014.0
R04	32.0	39.9	21.5	28.9	30.2	38.9	-96.1	-97.5	100.0	99.8
R05	0.3	0.0	0.5	25.1	0.3	0.5	-22.6	-1545.0	1.6	5.5
R06	1.9	3.5	2.4	6.5	1.6	1.9	0.0	0.0	85.9	98.4
R07	0.5	3.3	980.6	48.7	0.7	0.3	-56148.5	-870.5	43736.0	1634.4
R08	3.2	4.9	6.8	15.3	2.4	3.4	-98.5	-89.4	93.3	98.4
R09	68.5	168.4	607.2	1172.1	40.0	64.0	-1172.6	-13177.2	42522.2	59386.0
R10	49.5	68.6	65.0	100.2	35.4	38.1	0.0	0.0	929.2	936.2
R11	6.6	10.7	50.5	62.9	1.3	1.7	-150.0	-126.3	2497.9	2543.6
R12	0.4	0.6	0.4	0.6	0.3	0.4	-1.6	-0.7	5.8	15.4
R13	3.0	3.3	7.0	15.3	2.1	2.3	-98.4	-94.6	99.3	90.3
R14	20.8	39.1	225.1	40.3	19.2	38.7	-15475.1	-1175.5	99.2	197.0
R15	12.3	-2.3	92.2	370.4	13.2	10.3	-2035.2	-20711.6	5006.9	1347.4
R16	3.2	3.5	7.6	16.1	2.2	2.2	-96.1	-96.2	92.6	92.2
R17	133.9	77.9	10543.9	864.2	14.1	9.6	-230368.7	-8419.1	715850.7	26860.3
R18	12.3	6.2	161.5	90.1	12.2	7.5	-10399.0	-2355.5	1349.0	2563.8
R19	45.6	46.8	56.0	98.2	29.4	6.8	0.0	0.0	736.9	998.3
R20	51.1	54.2	46.1	68.6	45.1	39.4	0.0	0.0	823.0	956.0
R21	56.4	54.4	44.1	63.1	49.1	39.8	0.0	0.0	752.1	798.9
R22	77.9	96.6	57.2	217.9	75.6	84.8	0.0	0.2	2035.3	10342.6
R23	12.3	10.1	15.0	112.4	10.3	5.1	0.0	0.0	835.4	6688.4
R24	2.6	3.9	11.5	55.1	0.9	0.4	0.0	0.0	574.5	3016.5

F: Failed company, *NF*: Non-failed company. Variables are defined in Table 3.

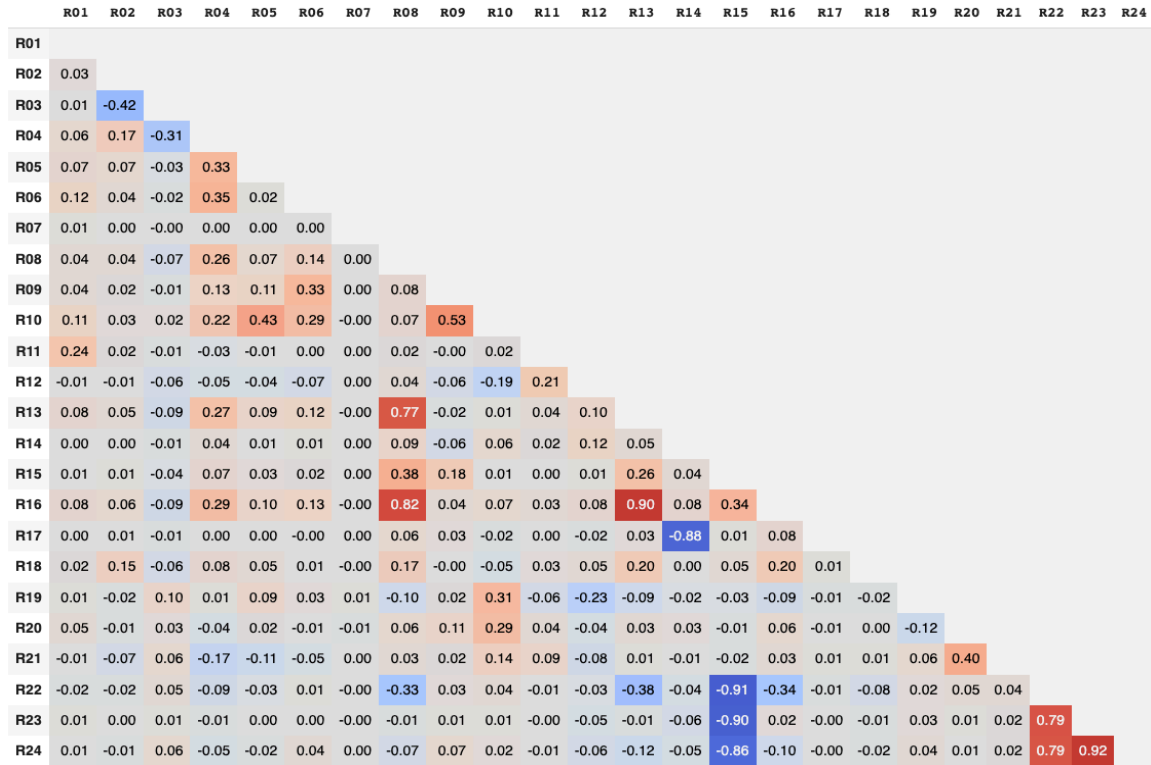
Appendix C: Correlation analysis



Correlation analysis 1-year horizon.



Correlation analysis 2-year horizon.



Correlation analysis 5-year horizon.

Appendix D : Tuning hyperparameters for models

Optimal parameters for Random Forest one year before the bankruptcy

imp_strategy: mean
learning_rate: 0.1
max_depth: 14
max_features: auto
n_estimators: 1000
random_state: 2

Optimal parameters for XGBoost one year before the bankruptcy

colsample_bytree: 0.5
learning_rate: 0.1
max_depth: 9
n_estimators: 400
objective: binary: logistic
reg_lambda: 2
subsample: 0.5

Optimal parameters for LightGBM one year before the bankruptcy

colsample_bytree: 0.5
learning_rate: 0.01
max_depth: 9
n_estimators: 400
objective: binary
reg_lambda: 0.3

Optimal parameters for NGBoost one year before the bankruptcy

Dist: Bernoulli
Score: LogScore
n_estimators: 2000
learning_rate: 0.005
minibatch_frac: 0.3
col_sample: 0.8

Optimal parameters for SVM one year before the bankruptcy

imp_strategy: mean
pca_whiten: true
C: 2
Class_weight: balanced
Gamma: auto
Kernel: rbf
Probability: true

Optimal parameters for Neural network one year before the bankruptcy

imp_strategy: mean
activation: tanh
alpha: 0.0001
hidden_layer_sizes: (100,)
learning_rate: adaptive
solver: adam

Optimal parameters for Logistic regression one year before the bankruptcy

imp_strategy: mean
n_components: 15
max_iter: 10000

tol: 0.1

C: 10000

**Optimal parameters for Discriminant analysis one year before the
bankruptcy**

imp_strategy: mean

n_components: none

solver: svd

tol: 0.0001

Appendix E: SHAP summary plot

See figure 8, 9, 10, 11, 12.

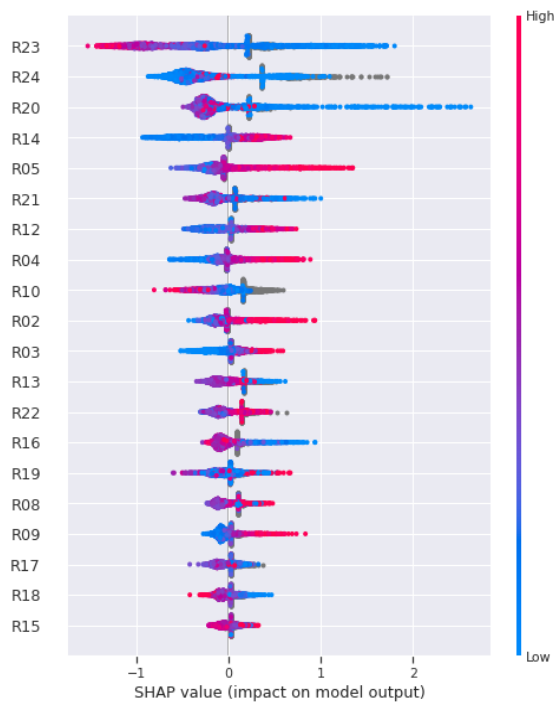


Fig. 8. SHAP summary plot for a 2-year horizon

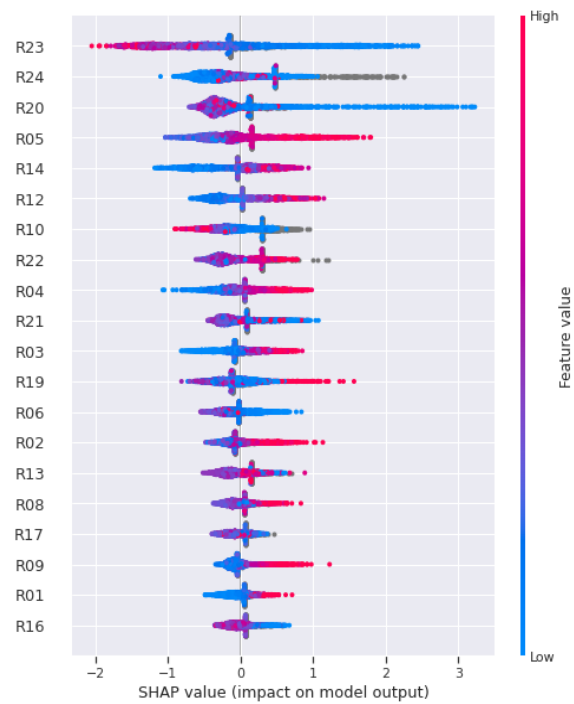


Fig. 9. SHAP summary plot for a 3-year horizon

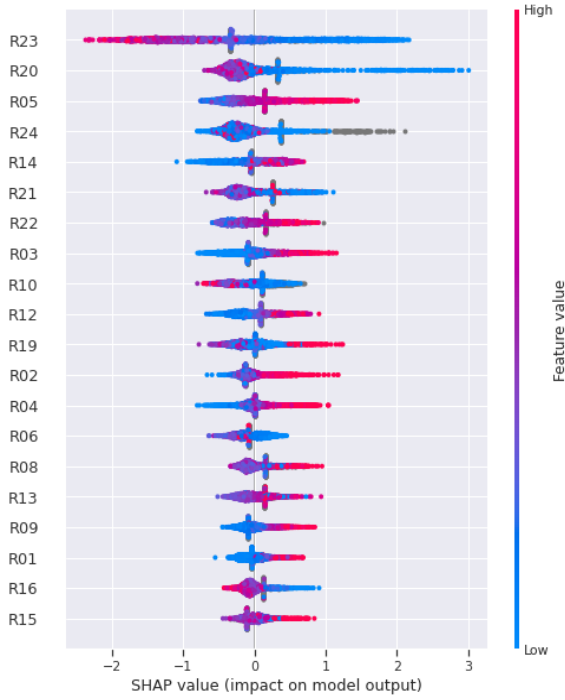


Fig. 10. SHAP summary plot for a 4-year horizon

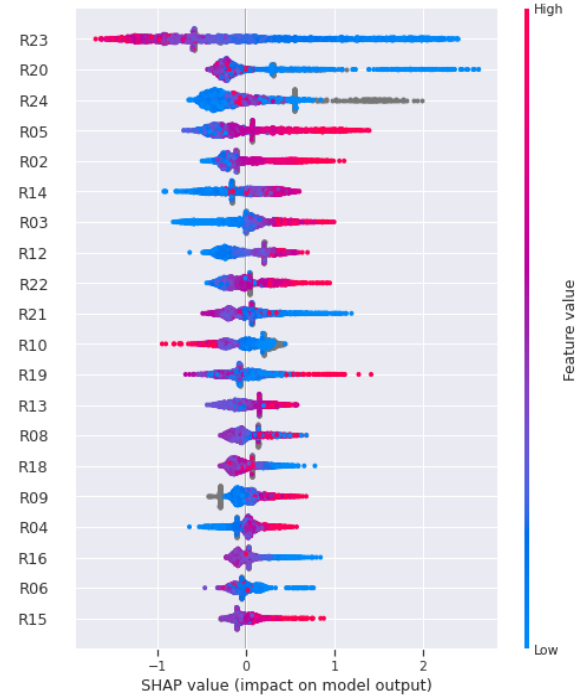


Fig. 11. SHAP summary plot for a 5-year horizon

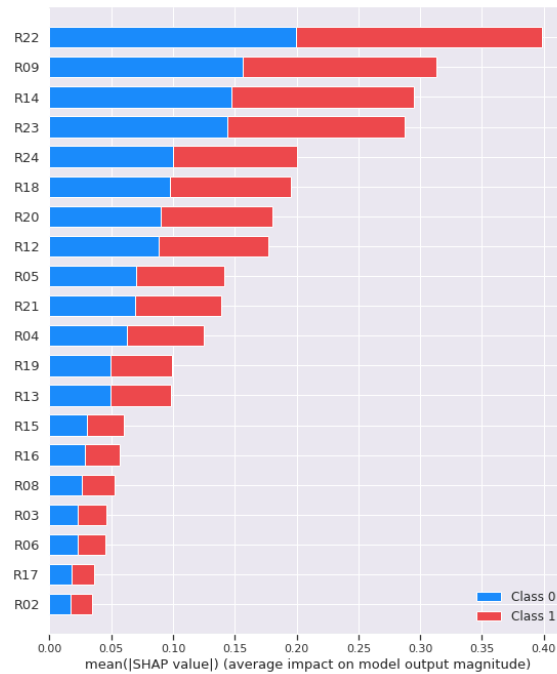
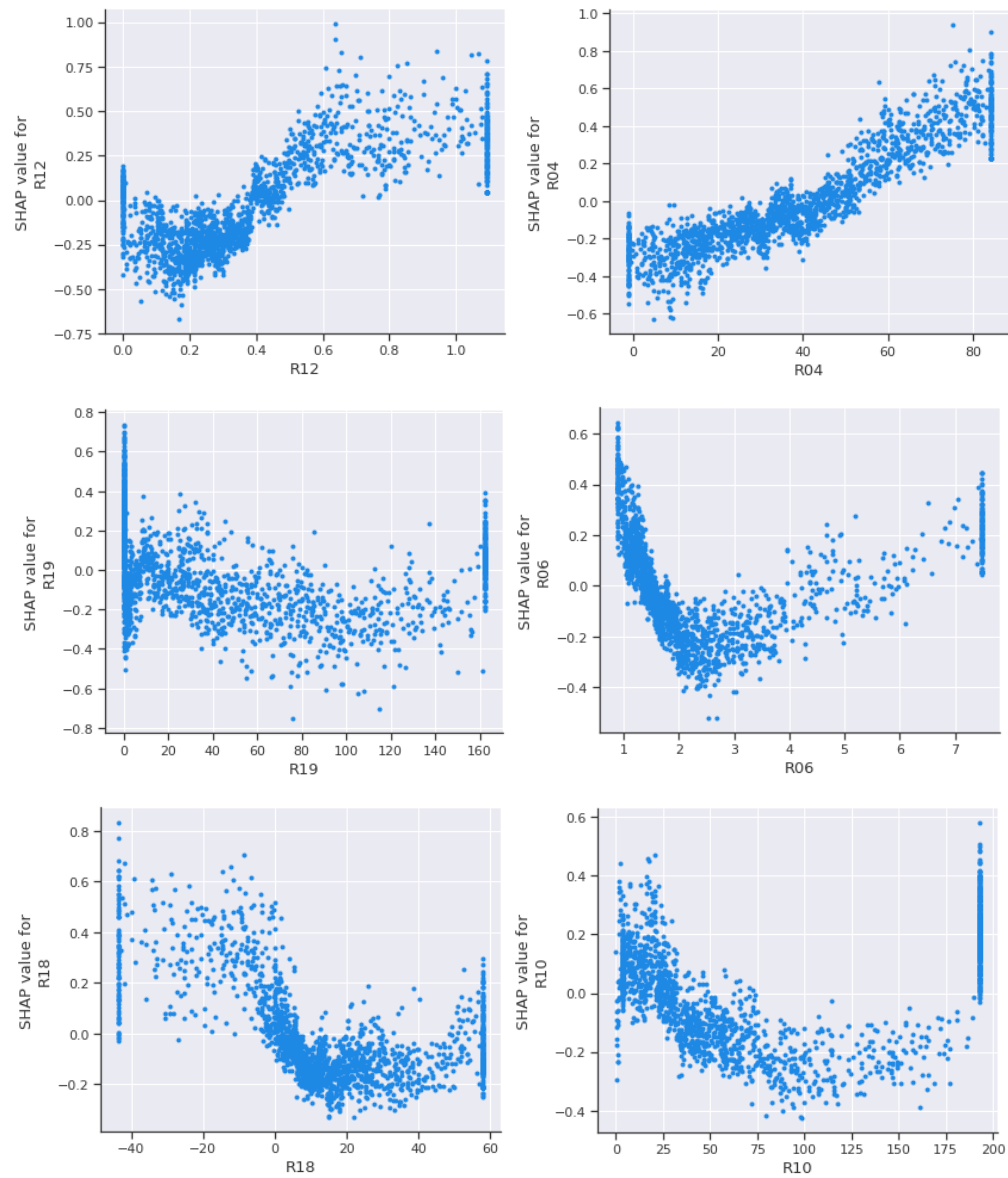


Fig. 12. SHAP summary plot for LightGBM for a 1-year horizon

Appendix F : SHAP dependence plot

See figure 13.



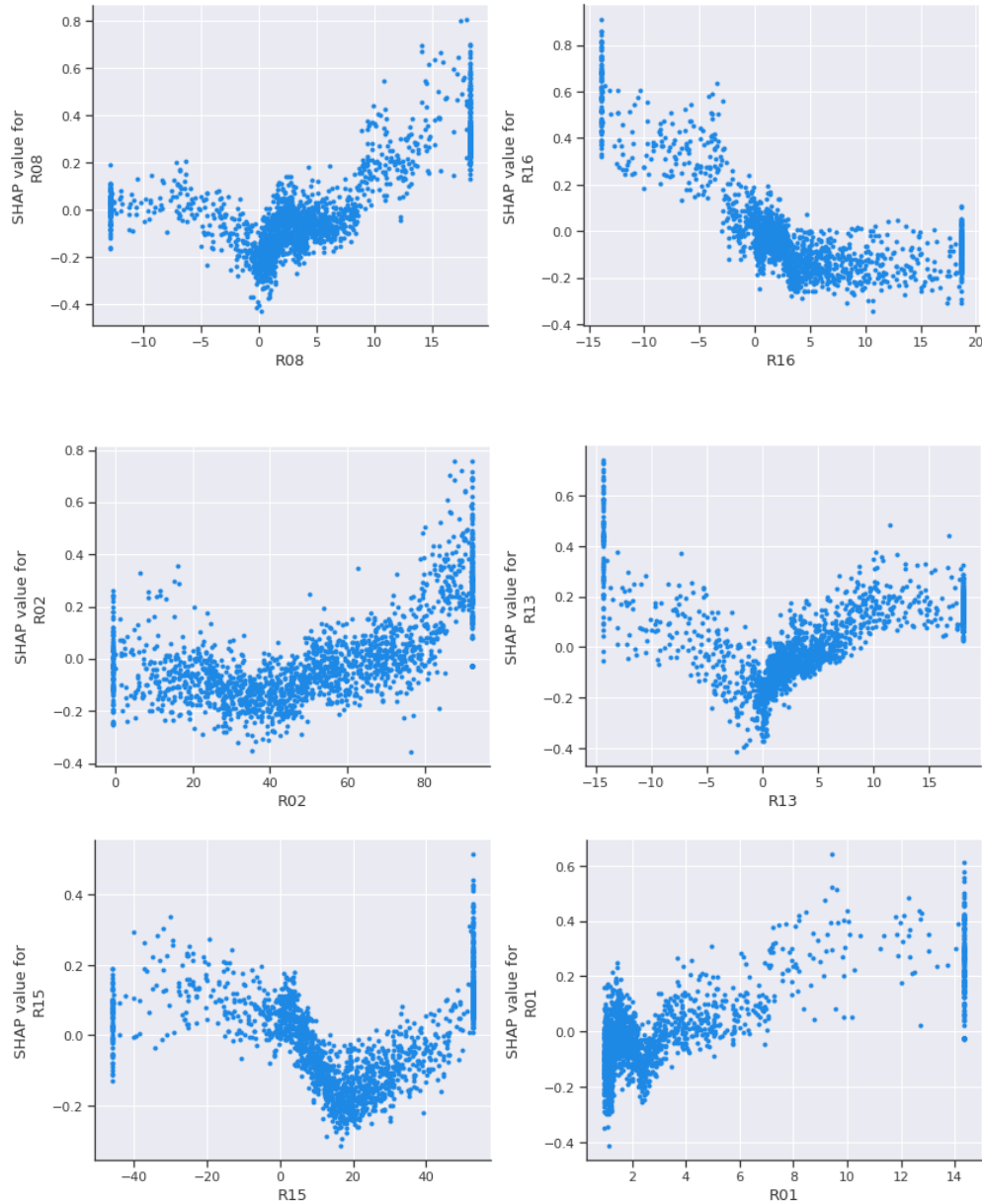


Fig. 13. SHAP dependence plots for important ratios one year before failure.

References

- Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, *94*, 164–184.
- Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems*, *45*(1), 110–122. <https://doi.org/10.1016/j.dss.2007.12.002>

- Almaskati, N., Bird, R., Yeung, D., & Lu, Y. (2021). A horse race of models and estimation methods for predicting bankruptcy. *Advances in Accounting*, 52, 100513. <https://doi.org/10.1016/j.adiac.2021.100513>
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589–609.
- Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETATM analysis A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1(1), 29–54.
- Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2020). A Race for Long Horizon Bankruptcy Prediction. *Applied Economics*, 52(37), 4092–4111. <https://doi.org/10.1080/00036846.2020.1730762>
- Babar, B., Luppino, L. T., Boström, T., & Anfinsen, S. N. (2020). Random forest regression for improved mapping of solar irradiance at high latitudes. *Solar Energy*, 198, 81–92. <https://doi.org/10.1016/j.solener.2020.01.034>
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417. <https://doi.org/10.1016/j.eswa.2017.04.006>
- Bardos, M. (1995). Détection précoce des défaillances d'entreprises à partir des documents comptables. *Bulletin de La Banque de France*, 3, 57–71.
- Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, 4, 71. <https://doi.org/10.2307/2490171>
- Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education*, 1–42.
- Ben Jabeur, S., Stef, N., & Carmona, P. (2022). Bankruptcy Prediction using the XGBoost Algorithm and Variable Importance Feature Engineering. *Computational Economics*. <https://doi.org/10.1007/s10614-021-10227-1>
- Ben Jabeur, S., Stef, N., & Carmona, P. (2023). Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. *Computational Economics*, 61(2), 715–741.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

- Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable Machine Learning in Credit Risk Management. *Computational Economics*, 57(1), 203–216. <https://doi.org/10.1007/s10614-020-10042-0>
- Carmona, P., Climent, F., & Momparler, A. (2019). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 61, 304–323. <https://doi.org/10.1016/j.iref.2018.03.008>
- Charalambous, C., Martzoukos, S. H., & Taoushianis, Z. (2022). Estimating corporate bankruptcy forecasting models by maximizing discriminatory power. *Review of Quantitative Finance and Accounting*, 58(1), 297–328.
- Charalambous, C., Martzoukos, S., & Taoushianis, Z. (2023). A Neuro-Structural Framework for Bankruptcy Prediction. *Quantitative Finance (Forthcoming)*.
- Chen, C., Zhang, Q., Ma, Q., & Yu, B. (2019). LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. *Chemometrics and Intelligent Laboratory Systems*, 191, 54–64. <https://doi.org/10.1016/j.chemolab.2019.06.003>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen, T.-K., Liao, H.-H., Chen, G.-D., Kang, W.-H., & Lin, Y.-C. (2023). Bankruptcy Prediction Using Machine Learning Models with the Text-based Communicative Value of Annual Reports. *Expert Systems with Applications*, 120714.
- Cho, S., Hong, H., & Ha, B.-C. (2010). A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: For bankruptcy prediction. *Expert Systems with Applications*, 37(4), 3482–3488. <https://doi.org/10.1016/j.eswa.2009.10.040>
- Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. *Journal of Business Research*, 101, 885–896. <https://doi.org/10.1016/j.jbusres.2018.11.015>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>

- Dikshit, A., & Pradhan, B. (2021). Interpretable and explainable AI (XAI) model for spatial drought prediction. *Science of the Total Environment*, *801*, 149797.
- du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing*, *73*(10–12), 2047–2060. <https://doi.org/10.1016/j.neucom.2009.11.034>
- du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, *242*(1), 286–303. <https://doi.org/10.1016/j.ejor.2014.09.059>
- du Jardin, P. (2016). A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research*, *254*(1), 236–252.
- du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, *75*, 25–43. <https://doi.org/10.1016/j.eswa.2017.01.016>
- du Jardin, P. (2019). Forecasting bankruptcy using biclustering and neural network-based ensembles. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-019-03283-2>
- du Jardin, P., & Séverin, E. (2011). Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model. *Decision Support Systems*, *51*(3), 701–711. <https://doi.org/10.1016/j.dss.2011.04.001>
- Duan, T., Avati, A., Ding, D. Y., Thai, K. K., Basu, S., Ng, A., & Schuler, A. (2020). *NGBoost: Natural Gradient Boosting for Probabilistic Prediction*. 11.
- Hernandez Tinoco, M., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, *30*, 394–419. <https://doi.org/10.1016/j.irfa.2013.02.013>
- Hung, C., & Chen, J.-H. (2009). A selective ensemble based on expected probabilities for bankruptcy prediction. *Expert Systems with Applications*, *36*(3), 5297–5303. <https://doi.org/10.1016/j.eswa.2008.06.068>
- 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*, 120658. <https://doi.org/10.1016/j.techfore.2021.120658>
- Jabeur, S. B., & Serret, V. (2023). Bankruptcy prediction using fuzzy convolutional neural networks. *Research in International Business and Finance*, *64*, 101844.

- Jones, S., Johnstone, D., & Wilson, R. (2017). Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting*, 44(1–2), 3–34.
- Ke G, Meng Q, Finley T et al (2017) LightGBM: A highly efficient gradient boosting decision tree. In: *Advances in neural information processing systems*, pp 3146–3154
- Kellner, R., Nagl, M., & Rösch, D. (2022). Opening the black box – Quantile neural networks for loss given default prediction. *Journal of Banking & Finance*, 134, 106334. <https://doi.org/10.1016/j.jbankfin.2021.106334>
- Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689–702. <https://doi.org/10.1016/j.ejor.2016.10.031>
- Loureiro, A. L. D., Miguéis, V. L., & da Silva, L. F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, 114, 81–93. <https://doi.org/10.1016/j.dss.2018.08.010>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- Lundberg, S. M., & Lee, S.-I. (2018). Consistent feature attribution for tree ensembles. *arXiv:1706.06060 [Cs, Stat]*. <http://arxiv.org/abs/1706.06060>
- Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., & Niu, X. (2018). Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. *Electronic Commerce Research and Applications*, 31, 24–39. <https://doi.org/10.1016/j.elerap.2018.08.002>
- Mercadier, M., & Lardy, J.-P. (2019). Credit spread approximation and improvement using random forest regression. *European Journal of Operational Research*, 277(1), 351–365. <https://doi.org/10.1016/j.ejor.2019.02.005>
- Moen, P. A. (2020). *Bankruptcy prediction for Norwegian enterprises using interpretable machine learning models with a novel timeseries problem formulation* [Master's Thesis]. NTNU.
- Molnar, C. (2020). *Interpretable machine learning. A guide for making black box models explainable*. Available at: <https://christophm.github.io/interpretable-ml-book/>
- Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., & Bischl, B. (2022). General pitfalls of model-agnostic interpretation

- methods for machine learning models. *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*, 39–68.
- Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 36(2), 3028–3033. <https://doi.org/10.1016/j.eswa.2008.01.018>
- Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *1990 IJCNN International Joint Conference on Neural Networks*, 163–168 vol.2. <https://doi.org/10.1109/IJCNN.1990.137710>
- Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109. <https://doi.org/10.2307/2490395>
- Perboli, G., & Arabnezhad, E. (2021a). A Machine Learning-based DSS for mid and long-term company crisis prediction. *Expert Systems with Applications*, 174, 114758.
- Perboli, G., & Arabnezhad, E. (2021b). A Machine Learning-based DSS for mid and long-term company crisis prediction. *Expert Systems with Applications*, 174, 114758. <https://doi.org/10.1016/j.eswa.2021.114758>
- Perboli, G., & Arabnezhad, E. (2021c). A Machine Learning-based DSS for mid and long-term company crisis prediction. *Expert Systems with Applications*, 174, 114758. <https://doi.org/10.1016/j.eswa.2021.114758>
- Schalck, C., & Yankol-Schalck, M. (2021). Predicting French SME failures: New evidence from machine learning techniques. *Applied Economics*, 53(51), 5948–5963. <https://doi.org/10.1080/00036846.2021.1934389>
- Séverin, E., & Veganzones, D. (2021). Can earnings management information improve bankruptcy prediction models? *Annals of Operations Research*, 306(1–2), 247–272. <https://doi.org/10.1007/s10479-021-04183-0>
- Sigrist, F., & Hirsenschall, C. (2019). Grabit: Gradient tree-boosted Tobit models for default prediction. *Journal of Banking & Finance*, 102, 177–192. <https://doi.org/10.1016/j.jbankfin.2019.03.004>
- Sigrist, F., & Leuenberger, N. (2023). Machine learning for corporate default risk: Multi-period prediction, frailty correlation, loan portfolios, and tail probabilities. *European Journal of Operational Research*, 305(3), 1390–1406.

- Smith, M., & Alvarez, F. (2021). Predicting Firm-Level Bankruptcy in the Spanish Economy Using Extreme Gradient Boosting. *Computational Economics*. <https://doi.org/10.1007/s10614-020-10078-2>
- Smith, M., & Alvarez, F. (2022). Predicting Firm-Level Bankruptcy in the Spanish Economy Using Extreme Gradient Boosting. *Computational Economics*, 59(1), 263–295. <https://doi.org/10.1007/s10614-020-10078-2>
- Son, H., Hyun, C., Phan, D., & Hwang, H. J. (2019). Data analytic approach for bankruptcy prediction. *Expert Systems with Applications*, 138, 112816. <https://doi.org/10.1016/j.eswa.2019.07.033>
- Staňková, M. (2023). Threshold Moving Approach with Logit Models for Bankruptcy Prediction. *Computational Economics*, 61(3), 1251–1272.
- Sun, X., Liu, M., & Sima, Z. (2020). A novel cryptocurrency price trend forecasting model based on LightGBM. *Finance Research Letters*, 32, 101084. <https://doi.org/10.1016/j.frl.2018.12.032>
- Tsai, C., & Wu, J. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34(4), 2639–2649. <https://doi.org/10.1016/j.eswa.2007.05.019>
- Tsai, C.-F., Hsu, Y.-F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing*, 24, 977–984. <https://doi.org/10.1016/j.asoc.2014.08.047>
- Vega García, M., & Aznarte, J. L. (2020). Shapley additive explanations for NO2 forecasting. *Ecological Informatics*, 56, 101039. <https://doi.org/10.1016/j.ecoinf.2019.101039>
- Veganzones, D., & Séverin, E. (2018). An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems*, 112, 111–124.
- Veganzones, D., Séverin, E., & Chlibi, S. (2023). Influence of earnings management on forecasting corporate failure. *International Journal of Forecasting*, 39(1), 123–143. <https://doi.org/10.1016/j.ijforecast.2021.09.006>
- Wang, G., Ma, J., & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications*, 41(5), 2353–2361. <https://doi.org/10.1016/j.eswa.2013.09.033>

- Yıldırım, M., Okay, F. Y., & Özdemir, S. (2021). Big data analytics for default prediction using graph theory. *Expert Systems with Applications*, 176, 114840. <https://doi.org/10.1016/j.eswa.2021.114840>
- Zhang, K., Xu, P., & Zhang, J. (2020). Explainable AI in deep reinforcement learning models: A shap method applied in power system emergency control. *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, 711–716.
- Zhao, S., Xu, K., Wang, Z., Liang, C., Lu, W., & Chen, B. (2022). Financial distress prediction by combining sentiment tone features. *Economic Modelling*, 106, 105709. <https://doi.org/10.1016/j.econmod.2021.105709>
- Zięba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, 58, 93–101. <https://doi.org/10.1016/j.eswa.2016.04.001>
- Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 59–82.