No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure

Pedro Carmona [a], Aladdin Dwekat [b,c], Zeena Mardawi [b,c]

[a] Department of Accounting, Universitat de Val`encia, Val`encia, Spain
[b] Department of Accounting, An-Najah National University, Nablus, Palestine
[c] Universitat Polit`ecnica de Val`encia, Val`encia, Spain

ARTICLEINFO

JEL classification:
G32
G33
C 45 C46

Keywords:
Business failure
Machine learning
XGBoost
Model interpretability

ABSTRACT

This study opens the black boxes and fills the literature gap by showing how it is possible to fit a very precise Machine Learning model that is highly interpretable, by using a novel ML technique, Extreme Gradient Boosting (XGBoost), and applying new model interpretability improvements. In addition, we identify several significant indicators that could assist in predicting business financial distress. The data were collected from the Eikon database from a sample of 1760 French firms (1585 healthy and 175 failing) in 2018. Identifying the leading indicators of business failure is critical in assisting regulators, and for business managers to act expeditiously before a distressed business reaches crisis point. Our results reveal that higher levels of equity per employee, solvency, the current ratio, net profitability, and a sustainable return on investment are associated with a lower risk of business failure. In contrast, a higher number of employees leads to business failure.

1. Introduction

Estimation of business bankruptcy is one of the most challenging issues of modern economic and financial research (Mousavi et al., 2015). Recently, with globalisation, rapidly changing economies and increased financial risk in the context of the global economic crisis of 2008 (Boubaker et al., 2016), much more attention has been paid to increasing the reliability of prediction models and to extending prediction periods before the announcement of bankruptcy. Recent studies suggest that group frameworks are a promising bankruptcy prediction method that could reduce the weaknesses of traditional models and provide an effective business failure prediction model for business institutions. Among these frameworks is gradient boosting, which has been explored in business failure prediction and credit scoring. It is a valuable machine learning method that can achieve superior classification performance.

Machine learning (ML) and artificial intelligence (AI) have become crucial parts of the Fourth Industrial Revolution (Duan et al., 2021; Goodell et al., 2021; Faccia et al., 2021; Faccia and Petratos, 2021; Mosteanu and Faccia, 2020). ML includes the process of utilising big data analytics and algorithms to obtain information and predict complex phenomena (Oyewo et al., 2020; Rapanyane and Sethole, 2020). The advantages of using ML and AL include their dynamic nature techniques, in that they can be run constantly in the background and offer real-time decisions for banks and financial institutions (Syam and Sharma, 2018). Extreme Gradient Boosting (XGBoost) is one of the most innovative ML forecasting approaches; it is an upgrading of gradient boosting, and has attracted great attention in recent years due to its high speed and accuracy. In addition, it has the ability to form and add decision trees sequentially in an attempt to fix the mistakes of previous models. Recently, XGBoost has been employed by several scholars in economics and business research. For example, Xia et al. (2017) proposed a novel boosted tree for a credit scoring model based on XGBoost; they point out that XGBoost for credit scoring operates in a superior fashion to previous credit scoring systems, which reflects its trustworthiness in finance estimates. Moreover, Santhanam et al. (2017) argue that XGBoost is the most appropriate model to analyse the aspects that influence clients' retaining for software-as-a-service firms. In addition, according to Climent et al. (2019), such a technique performs better than logistic regression and other statistical techniques in issues related to corporate insolvency. However, few studies have applied XGBoost in business failure research; in one example, Climent et al. (2019) used it to pinpoint the financial indicators that could forecast bank failure in the Eurozone. Another study conducted by Carmona et al. (2019) used XGBoost to predict US bank failures.

On the other hand, ML approaches are usually criticised for being complex and challenging to interpret; they are frequently described as a "black box" architecture (Vega García and Aznarte, 2020). The comprehensiveness and trustfulness of a model's results are considered to be a hallmark of good science and are vital for the adoption of our model (Boehmke and Greenwell, 2020). Both analysts and business stakeholders need to understand and interpret the ML models results in order to trust the application of the model in business decisions. Interpretation of the results will also help users to estimate the importance of each financial ratio used and thus allow them to make decisions based on these. Nonetheless, while model interpretability is a rapidly expanding ML, and although there have been various efforts to analyse the different aspects of interpretations, none of these have been applied in the field of business failure.

Therefore, the purpose of this study is to fill the literature gap and open the black box by identifying several leading indicators that could help to predict business failure by using novel ML methods (i.e., XGBoost) and applying model interpretability. The research is based on a sample of 1850 French firms (1585 healthy and 175 failing) in 2018, with data taken from the Eikon database.

Consequently, our study makes several vital contributions in both the theoretical and practical areas to the rare literature in this important field. ***First***, thestudy identifies the vital indicators that may help to predict business failures using XGBoost. Observing these indicators detected by this cutting-edge algorithm could accelerate the timely identification of risk management control failures and eventually avoid business failure. ***Second***, to the best of our knowledge, only Carmona et al. (2019) and Climent et al. (2019) have utilised XGBoost to predict failures in the Eurozone and the US. However, these studies examined bank failure but did not apply model interpretability. Nevertheless, it is not suficient to identify a ML model that optimises predictive performance. Therefore, to the best of our knowledge, this is the first study that utilises model interpretability. Moreover, the use of such a model yielded very high predictive and explanatory power in terms of accuracy. ***Finally***, our study could be helpful for different interested parties such as scholars, policymakers, regulators, and professionals, as it offers new directions and insights for future research by using new methodological approaches (i.e., XGBoost) and model interpretability. Moreover, the high predictive and explanatory power of the model used in this study could help firm managers to track the most significant attributes detected in the one-year-before-failure model. Managers could also prevent business failure by taking timely decisions instead of waiting for government intervention. Regulators might also find the model helpful in recognising and warning about possibly distressed firms.

The findings of the study contribute to the common acceptance of ML interpretability methods, which could increase the implementation of ML and AI in business applications and our daily lives. The results also indicate that the most crucial factors in predicting business failure are the number of employees, equity per employee, the solvency ratio, current ratio, net profitability, and sustainable return on investment.

The remainder of this article is organised as follows. First, we carefully review the related literature and illustrate the study data in the following sections. The empirical and methodological sections are then presented. In the subsequent part, the study results are presented, and the final section presents the conclusions, future research recommendations and practical implications.

2. Literature review

The aim of business failure prediction should be the early warning of bankruptcy. The term 'bankruptcy' has various definitions in the literature. The most common refers to company insolvency, which is interpreted as its incapability to pay debts (du Jardin, 2017; Kalak et al., 2017; Stolbov and Shchepeleva, 2020). However, technical insolvency implies a liquidity shortage but does not yet necessarily mean bankruptcy. A lack of liquidity could be temporary and may be solved by suitable management decisions. Altman (1968) suggests that insolvency leads to bankruptcy when business total debt is more than the value of total assets held over the long term. On the other hand, a profitability criterion was suggested by Berryman (1993) to define the bankruptcy risk of a company. He proposes that bankruptcy risk is characterised by less return on equity in the long term compared to similar firms. However, a variety of broad interpretations of the term is proposed in the literature. Moreover, Davies (1997), in the context of collapsed firms in the UK and France, found that the majority had positive financial ratios and results in the period before bankruptcy.

Since Altman's pioneering model of multivariate discriminant analysis, numerous studies have developed a variety of statistical methods using logistic regression (LR) and multivariate discriminant analysis (DA) (Jabeur et al., 2021). As a result, statistical techniques have become widely used in predicting bankruptcy; however, many criticisms have been raised regarding statistical assumptions, including linearity, normality, and variable independence (Le and Viviani, 2018). In addition, significant concerns have been raised about the capability to modify the threshold of such models in order to boost the results. According to Jabeur et al. (2021), statistical methods do not offer reliable results due to the simplicity of the manual modification of the threshold to improve model efficacy. Furthermore, such modifications will not enhance the efficiency of the model in the overall population of firms after its implementation, but only in the inventor's test model.

One more concern of the traditional prediction models is their stationarity (Momparler et al., 2020). Typically, bankruptcy models use the financial ratios computed from financial statements for their predictive power, accessibility and standardisation (Du Jardin, 2016). In general, they allow for a useful distinction between failing and non-failing businesses. However, the approach on which they are constructed is one of their major weaknesses. Most of the models depend on the static values of ratios for a particular period, while bankruptcy has many causes and indicators. The capacity of a model to recognise the diverse negative conditions is a critical factor in its effectiveness and performance. Therefore, more recently most bankruptcy prediction methods have changed from statistical to intelligent methods, such as fuzzy logic (Momparler et al., 2020), neural networks (NN) (Hosaka, 2019), support vector machines (SVMs) (Le and Viviani, 2018), and gradient boosting machines (GBMs) (Campillo et al., 2018).

The development of information technology could help acquire an array of information regarding the various risk statuses of firms in several ways (Khoja et al., 2019; Mselmi et al., 2017). Statistical and AI techniques are gradually being utilised to enhance the effectiveness of insolvency prediction. In this vein, the DA approach was compared with other prediction models in terms of accuracy; i.e., probabilistic models, the NN approach and various ML algorithms. Previous research indicates that intelligence modelling methods deliver the best performance compared to traditional ones (du Jardin, 2017; Jabeur et al., 2021; Mosteanu and Faccia, 2020). Hosaka (2019) concludes that NN leads to better results in the process of identifying potentially failing firms. Similarly, Kim et al. (2016) tested the effectiveness of clustering techniques and genetic algorithms based on artificial neural networks (ANNs). Moreover, using a sample of Korean firms, the results of Lee and Choi (2013) suggest that back-propagation neural networks (BPNs) outperform multivariate DA in corporate bankruptcy.

On the other hand, several ML techniques have been successfully employed and their utility proven in the field of bankruptcy prediction. Le and Viviani (2018) compared two traditional statistical methods (DA and LR) and three ML approaches (ANN, SVM and k-nearest neighbours) in predicting financial distress, showing that ANN and k-nearest neighbours outperformed other traditional and ML models in terms of accuracy and generalisation performance. Tsai and Cheng (2012) compared the prediction power of SVM, DA, LR and NN in the presence of outliers. They concluded that, on average, SVM was the most efficient technique in predicting corporate analysis. In addition, Erdogan (2013) assessed bank

financial failure using SVM; his results reveal that SVM is efficient in obtaining valuable information from financial data and could be utilised as an early alert system for bankruptcy. Eling and Jia (2018) concluded that SVM is better than logistic regression, indicating superior overall predictability. Decision trees are also widely used in corporate failure (Zhou et al., 2017).

Different scholarly articles have successfully employed gradient boosting to predict financial distress. For instance, Alfaro et al. (2008) compared the prediction accuracy of AdaBoost and ANN. Their results show that AdaBoost reduced generalisation error by nearly 30% in relation to artificial neural network error. Cortes et al. (2008) ´ compared DA and AdaBoost; they suggest that AdaBoost is more accurate than DA and is also a useful technique to predict business failure. Zięba et al. (2016) and Jones (2017) applied XGBoost to business failure, indicating that XGBoost performed substantially better than traditional techniques. Momparler et al. (2016) identify the main four variables that could assist in predicting and avoiding EU bank failure using the boosted classification tree. They conclude that larger banks with the higher income from non-operating items, net loans to deposits, and a higher interbank ratio are more likely to face bankruptcy. On the other hand, using a sample of US banks, Carmona et al. (2019) compared the capability of XGBoost, random forest, and LR in predicting bank failure; they indicate that XGBoost outperformed other models in terms of accuracy and generalisation performance. Recently, Jabeur et al. (2021) applied CatBoost to financial distress prediction, showing that itsignificantly enhanced performance prediction quality.

Table 1
Features considered to fit XGBoost models.

Ratio	Definition	Ratio	Definition
R1	Solvency ratio	R19	Added value / Turnover
R2	Equity per employee	R20	Inventory turnover
R3	Total assets per employee	R21	Financial interest rate
R4	Global performance	R22	Interest / Turnover
R5	Sustainable return on investment	R23	Leverage
R6	Earnings before taxes (EBT) / Total assets	R24	Repayment capacity
R7	Accounts receivable	R25	Cash flow
R8	Accounts payable	R26	Working capital / Turnover
R9	Total net assets turnover	R27	Current ratio
R10	Staff costs/turnover	R28	Quick ratio
R11	Turnover per employee	R29	Total assets
R12	Annual average wage	R30	EBITDA
R13	Earnings before interest and taxes/Total assets	R31	EBT
R14	Earnings before interest, taxes, depreciation, and amortisation (EBITDA) / Total assets	R32	Working capital
R15	EBITDA/Number of employees	R33	Return on net assets
R16	Added value/Equity	R34	Return on equity
R17	Net profitability	R35	Number of employees
R18	Return on net equity	R36	Unemployment rate of firm's department

2.1. Sample and data

The data were collected randomly from the Eikon database for small and medium sized French firms operating in several activity sectors, which were divided into two groups of non-failing and failing firms. We excluded financial institutions because of the unique characteristics of this type of company (Dwekat et al., 2020, 2021); indeed, banking companies are typically barred from similar business failure studies. We obtained a sample of failed firms that filed for insolvency in 2018. The data were structured in such a way that the accounting year is available for one year before failure (2018). The final sample included 1760 firms, distributed into 1585 non-failing and 175 failings ones. We selected the study variables (financial ratios) in line with the previous business failure literature (Carmona et al., 2019; Climent et al., 2019; du Jardin, 2017; Khoja et al., 2019; Mselmi et al., 2017). Therefore, a series of 36 variables (ratios) (R1 to R36) were chosen (see Table 1) from those generally applied in the literature, representing considerable informational content for the analysis of firms' financial condition.

3. Methodology

We aimed to train an XGBoost to identify the most critical indicators that predict firms' financial distress. The fundamental goal was to identify a model that prevents overfitting and makes generalisable predictions. We gathered financial and accounting data of failed firms that filed for bankruptcy during 2018 and then formulated predictions for the settings one year before failure. Chen and Guestrin (2016) indicate that XGBoost is an effective and scalable implementation of the GBM of Friedman (2001) and Friedman et al. (2000) The algorithm upgrades ML gradient boosting tree-based models that are becoming vital in many areas. One of the features of gradient boosting is the creation and addition of decision trees sequentially, in an attempt to correct the mistakes of previous models.

Climent et al. (2019) and Carmona et al. (2019) applied XGBoost to predict bank failure in the Eurozone and the US. Both studies concentrated on predictive rather than explanatory modelling, the use of such methodology yielding a very high predictive power in terms of accuracy.

XGBoost is a tree ensemble model, which consists of a set of classification or regression trees. Adding the prediction of multiple trees, the model overcomes the limitations or weaknesses of a single tree (relatively low predictive performance). The final model was a linear combination of many trees (usually hundreds to thousands) that can be thought of as a regression model in which each term is a tree (Elith et al., 2008). Fig. 1

summarises the procedure presented in the stochastic gradient tree-boosting algorithm,[1] showing that the new improved model's prediction depends on its prediction error. Instead of stopping after one iteration of improvement *(k = 1)*, the process is repeated multiple times (*K*), with the learning of new error predictors for newly formed improved models until the accuracy or error is satisfactory. Therefore, after every iteration, the model learns and adds new information into the ensemble, accounting for the previous model's errors in a fraction λ.

XGBoost is a regularised gradient boosting method because it allows regularisation or the constraining of variable weights. This new feature distinguishes it from other boosting techniques. It helps reduce overfitting compared to usual GBM execution with no regularisation, employing penalising variable importance in high-dimensional scenarios and thus performing the variable selection. Penalty regularisation decreases overfitting, without generating a loss in predictive power and as a result fitted models generalise well to new or unseeing data. The combination of all trees provides the final model. The prediction output is given as follows:

$$Z_i = G(X_i) = \sum_{j=1}^{K} g_k(X_i)$$

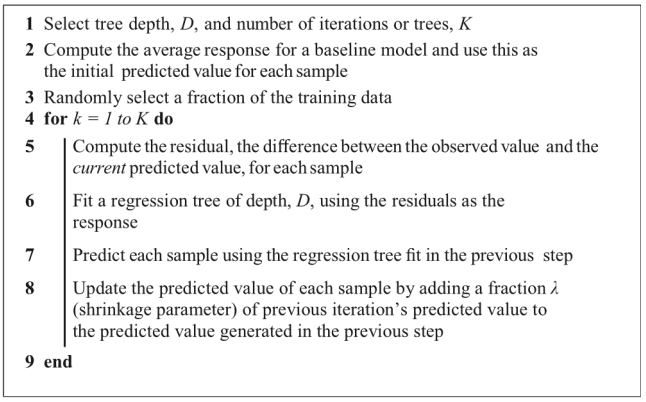
where X_i are the considered features, and $g_k(X_i)$ is the output function of each tree.

As in many ML algorithms, XGBoost requires hyper-parameter optimisation or tuning. ML is an iterative procedure in which the accuracy of predictions made by the models is continuously improved. In each of these iterations, specific parameters are also fine- tuned continuously. The best values of these parameters cannot be estimated from data, and must be tuned independently for a given predictive modelling problem. Therefore, hyper-parameter tuning identifies the set of the best hyper-parameters for a learning problem or discovers the parameters that yield a model with the highest prediction accuracy. In XGBoost models, the number of parameters is huge, but the highly critical ones are as follows.[2]

The number of rounds or trees: Boosting iterations or the number of trees to be built. More trees beyond a certain limit do not assure an improvement in model performance. Boosted tree models are built sequentially, and every new tree attempts to fix the errors made by prior trees, but the model rapidly reaches a point of diminishing returns.

Maximum individual tree depth: This refers to the size of the decision trees, also known as the depth. Shallow trees generally have bad performance because they capture few details of the data, while complex models of deeper trees are usually more likely to be overfitting because they capture too many details of the data, and the results cannot be generalised. The range of this parameter is from 0 to infinity.

Learning rate or shrinkage: This specifies the learning rate by which to shrink the feature weights. Shrinking feature weights after



NOTE: Taken and adapted from Kuhn and Johnson (2013)

Fig. 1. Stochastic gradient boosting algorithm. The trees in boosting depend on past trees, and the final prediction is based on an ensemble of tree models. NOTE: Taken and adapted from Kuhn and Johnson (2013).

each boosting step makes the boosting process more conservative and prevents overfitting. In addition, it determines how quickly the algorithm adapts, or the contribution of each tree to the growing model, in each iteration. As the learning rate falls, more steps or rounds must get to the optimum. The range is 0.0–1.0.

Gamma or minimum loss reduction (regularisation) for a further partition: The value of this parameter specifies the minimum relative improvement in squared error reduction in order for a split to happen. When properly tuned, this option can help reduce overfitting. The higher this parameter, the higher the regularisation. The default value is 0 (no regularisation), and the range of the parameter is from 0 to infinity.

Subsample ratio of features or columns: This allows specification of the column sampling rate for each split in each level or tree. The value ranges from 0 to 1.

[1] Detailed mathematical explanations of the XGBoost algorithm can be found at the following URL: https://xgboost.readthedocs.io/en/latest/ tutorials/model.html (visited on September 2021)

[2] Further information and detailed explanations about general and learning XGBoost parameters can be found on the web page: https://xgboost.readthedocs.io/en/latest/parameter.html

Subsample ratio of observations or rows: This allows specification of the observation sampling rate for each split in each level or tree. The value ranges from 0 to 1.

Minimum sum of instance weights needed in a child (minimum child weight): This defines the minimum number of observations for a leaf. The larger this parameter, the more conservative the algorithm, but very high values can produce underfitting.

Hyper-parameter tuning for model optimisation is an exploration problem and consists of identifying the ideal combination of these parameters. One of the methods most widely used for optimisation using hyper-parameters is random search. In order to ascertain the best configuration, a list of all possible values for each hyper-parameter in a specified range is constructed, known as the parameter grid, and only a random subset of all possible parameter combinations or configurations is trained. This means that the grid search moves through a space of all considered hyperparameters and builds the respective models, which are generally quite a few. The build process is constrained by time, although the more models that are built, the better.

Employing cross-validation methods, it is possible to determine an optimal combination of the aforementioned XGBoost parameters. In k-fold cross-validation, data are randomly partitioned into k sets or folds of equal size. Consequently, each set is successively held out and fitted with an XGBoost model using the remaining data. This process yields k estimates of prediction accuracy, which are combined into an overall estimation of the test error for each model as a measure of global model performance. It is quite common to take k = 10 (10-fold cross-validation), which was the choice for this study.

We divided the study sample of French companies into two different sets: 80% of the instances to train the XGBoost model to find the best model parameter configuration, with the remaining 20% forming the test data or hold-out sample to measure the performance of the best-fitted model.

All the models were fitted in R (R Core Team, 2019) version 3.6.2, and for XGBoost *xgboost package version 0.90.02* was used (Chen et al., 2019), together with *caret package version 6.0–85* (Kuhn, 2020). Additionally, for model interpretability, we used *DALEX package version 1.0* (Biecek, 2018; Biecek and Burzykowski, 2021).

4. Results and discussion

4.1. XGBoost model with all variables

In the first step, we tuned and fitted a model considering all the available ratios shown in Table 1. The purpose was to identify the most critical variables; that is, those with the highest predictive content in business failure. The aim was also to facilitate the exclusion of features that were just noise. This was important to control and find the optimal values of the hyper-parameters mentioned above to find the structure in the data in the best way possible. The AUC was taken to measure model performance; it is one single metric that combines the false positive rate and the true positive rate, and is widely used in ML when dealing with imbalanced data. In our study, this area was comparable to the probability that an actual randomly chosen failed company had a higher probability of being classified as a failure by the model than a non-failed company. The AUC values ranged between 0 and 1; and a value of one denoted a perfect model, one which provided the best fit to the data. We built a model that could be generalised on unseen data, avoiding overfitting, which required tuning and selecting the optimal XGBoost model parameters values using cross-validation.

Using the R caret package (Kuhn, 2020) and cross-validation (k = 10), we identified the best model hyper-parameter combination in the training data (80% of all data). In particular, we fitted 100 different XGBoost models, each with a different parameter configuration, with the best model yielding an AUC cross-validation resampling result of 0.953, which is exceptionally high, and near to the maximum value of one. Regarding the testing data (the remaining 20% of all the data) or the hold-out sample, the AUC was 0.946, and global accuracy rose to 90.62%. Because the AUCs for the testing data and cross-validation resampling are quite similar, this indicates that the resulting model is not overfitting. Fig. 2 shows the ten most important variables, which have a higher contribution to this XGBoost model for business failure prediction; the most significant percentage means a fundamental ratio for the prediction. For the boosted tree models, each gain of each tree feature was taken into account, then the average per feature, to give an overview of the entire model (Chen and Guestrin, 2016).

4.2. XGBoost model with the most important variables

As shown previously, the XGBoost using all the available features produced an excellent predictive model for business failure, with global accuracy and AUC very high in the hold-out sample (testing data). Therefore, taking a parsimonious approach, we selected the six most essential variables identified, with the most significant percentage of contribution, namely:

- R35: Number of employees
- R2: Equity per employee
- R1: Solvency ratio
- R27: Current ratio
- R17: Net profitability
- R5: Sustainable return on investment

Determining which input variables impacted a specific prediction is crucial to explanation. These most critical variables are the only selected features to build the next XGBoost model to predict business failure. The idea of finding simpler data models is a concept whose origin dates back a long time, and is referred to as Occam's razor because of William of Occam (an English Franciscan friar, scholastic philosopher, and theologian),

who proposed this concept in the 13th Century. Given two models of similar generalisation errors, the simpler model one be preferred over the more complex one; in that sense, Occam's razor advises us to use the simplest model.

Before fitting the new model, and as an exploratory analysis, we completed a correlation analysis of the data to identify the importance of the features and their sign-in relation to business failure probability, utilising a correlation funnel. This visualisation method uncovers insights by graphically elevating high-correlation features and lowering low-correlation ones (Fig. 3). It is a three- step process: transforming the data into a binary format; performing a correlation analysis; and visualising the feature-target relationships (failure or non-failure). This procedure is very efficient in understanding which features are most likely to provide understanding of business failure. The shape looks like a funnel (Cran.r-project.org, 2020).

Regarding the three most correlated variables, and according to the results displayed in Fig. 2, higher values of R35 (number of

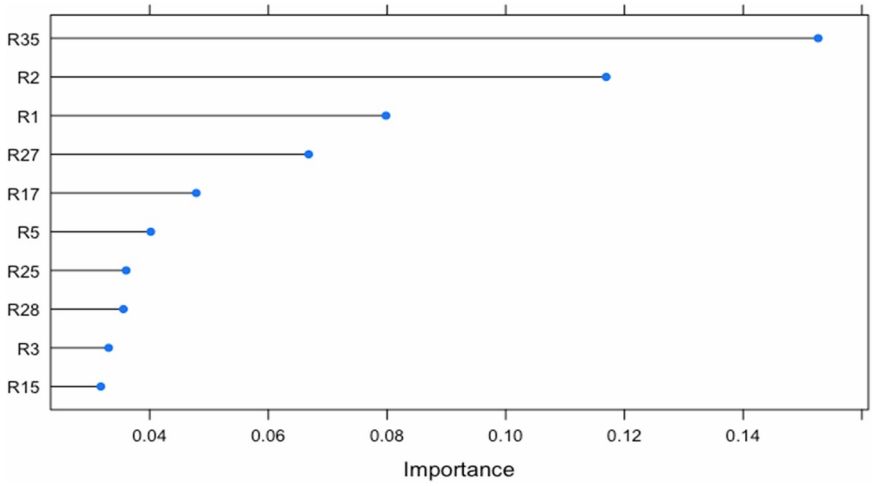


Fig. 2. 10 most important features of fitted XGBoost model using all features.

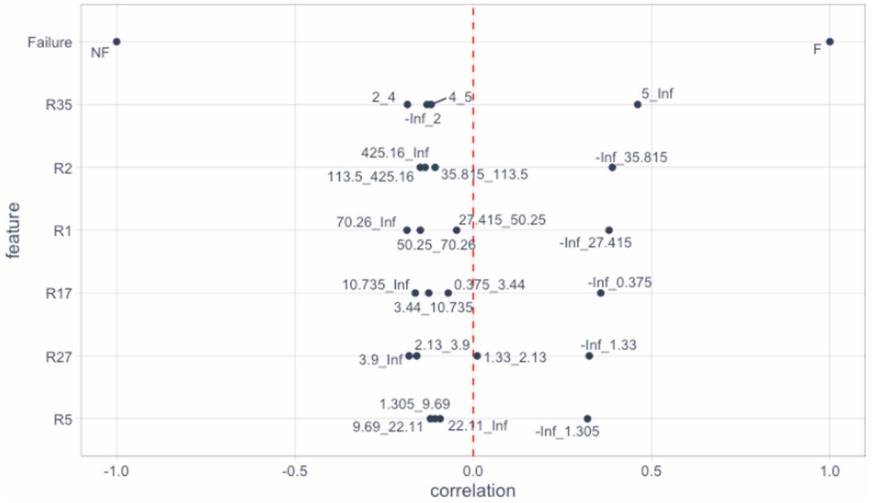


Fig. 3. Correlation funnel for most important selected features. Visualisation for understanding which features have higher relationships to response variable. First, numeric features are binned into ranges and then converted to binary features via one-hot encoding. Next, is performed a correlation analysis between the response (F represents business failure and NF non-business failure) and all the binned features. The plot lists the highest correlation features (based on absolute magnitude) at the top of the and the lowest correlation features at the bottom.

employees) are correlated with failed companies (from 5 to infinity); lower values of R2 (equity per employee) are correlated with failed companies (from minus infinity to 35.815); and lower values of R1 (solvency ratio) are associated with failed companies (from minus infinity to 27.415). These results mean that larger companies with bad solvency are more likely to face financial distress. This is in line with Climent et al. (2019), who concluded that larger banks are more susceptible to financial distress. With reference to R17, the next most correlated variable, the higher values of net profitability are correlated with a lower risk of business failure (from 10.735 to infinity). Additionally, Figs. 4 and 5 show the median and box plots of the features respectively, broken down into the two groups of interest: failed and non-failed companies.

After the exploratory analysis of these six selected features, the XGBoost was fitted with 100 different permutations of hyperparameters on the training data (80% of all data); cross-validation with $k = 10$ helped in identifying the optimum configuration of these parameters using the AUC as the performance measure. The optimal fitted model had an AUC cross-validation resampling result of 0.947, which was a little lower than

that of the previous model of 0.953 when considering all the variables. Therefore, the parsimonious XGBoost model with the most critical variables has predictive power very close to the original model that incorporates all the features. Table 2 displays the best hyper-parameter combination identified conducting to this new model. Fig. 6 plots the training and

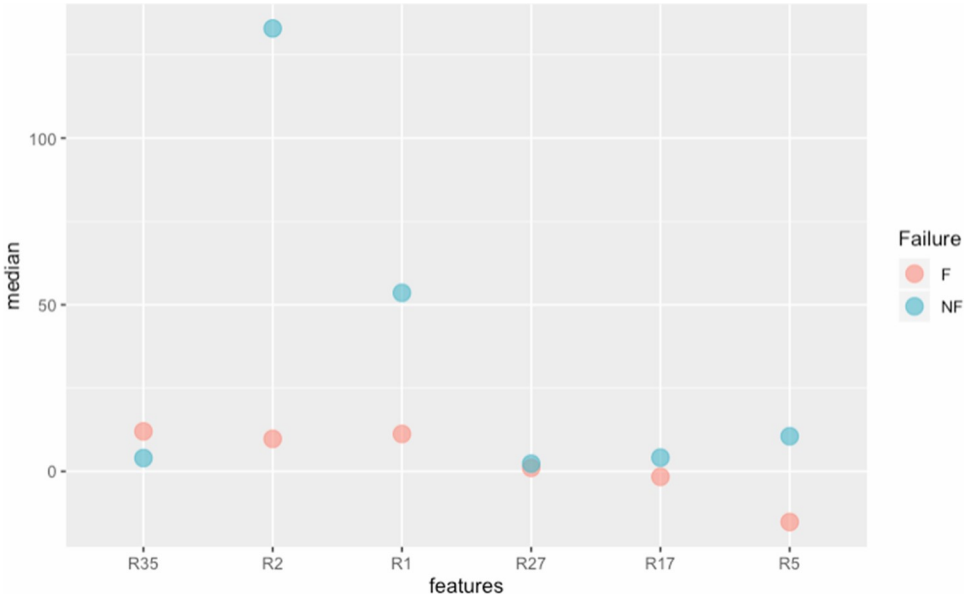


Fig. 4. Median of most important features (F represents business failure and NF non-business failure).

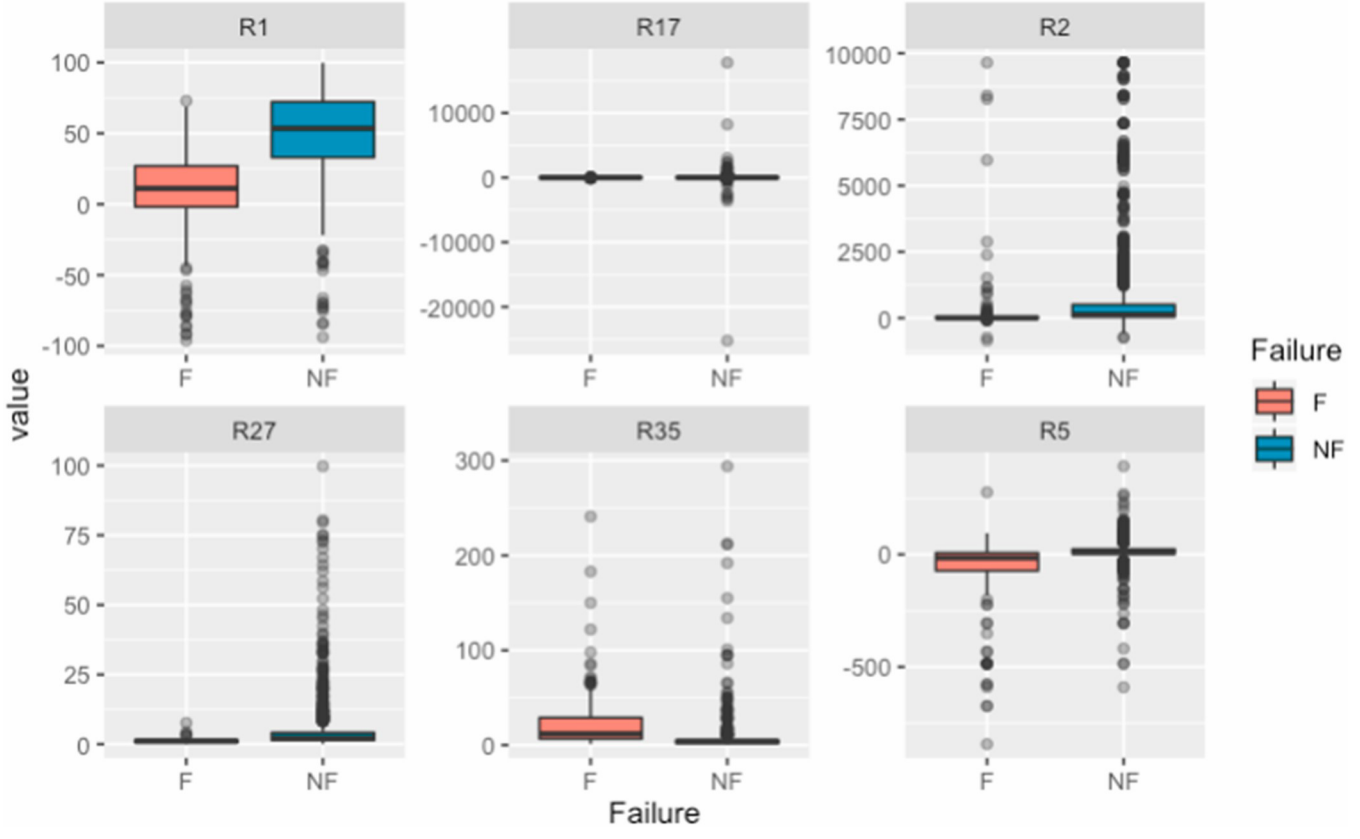


Fig. 5. Boxplots of most important features.

cross-validation values of the AUC, as a performance measure for the fitted XGBoost model. The cross-validation curve is stable and does not change as the number of trees or iterations grows, which may indicate a possible overfitting problem.

In the test data or hold-out sample (the remaining 20% of the data), the AUC yielded a value of 0.964, quite near to the highest value of one. This means that most of the business failure predictions are correct (Fig. 7). A confusion matrix or table of the prediction results for the best-identified threshold of 39% is presented in the right-hand panel of Fig. 7. The confusion matrix compares the XGBoost fitted model failure predictions to the actual outcomes (failed or non-failed businesses). Failure probability in the test data obtained from the XGBoost fitted model was converted into a dichotomous prediction: a value of one if it exceeds the best threshold of 39% and zero otherwise. We obtained 12.57% (100 – 87.43) as the number of misclassifications or global percentage of errors in the test data. The model sensitivity or true positive rate was 93.55% (failed companies classified correctly) and specificity or the true negative rate was 86.80% (non-failed companies classified correctly).

Furthermore, to illustrate that the XGBoost-fitted classifier differentiated most failed and non-failed companies correctly, Fig. 8 shows the density plot of the predicted probabilities of failure in" the hold-out data or test set for the two groups of companies. Probabilities are much higher and close to 1 for the failed company cases and lower and close to 0 for the non-failed ones. The best- identified cut-off or threshold point was 0.39 (see Fig. 7), which means that a vertical line on the x-axis" with this value would separate most of both types of company correctly.

4.3. Model interpretability

As we apply and embed increasingly complex predictive modelling and ML algorithms, both analysts and business stakeholders need methods to interpret and understand the modelling results so we can trust ML models application to business decisions (Doshi-Velez and Kim, 2017). Generally, greater accuracy often comes at the expense of interpretability; however, this section illustrates how it is possible to extract essential insights from complex ML algorithms, such as XGBoost.

ML model complexity is frequently linked to its interpretability; typically, the more complex the model, the more difficult it is to explain and interpret. For example, traditional regression procedures are among the most interpretable models; a change in any independent variable produces changes at a defined rate to the dependent variable, in only one direction, and a magnitude represented by a coefficient. Nevertheless, in modern ML algorithms, no coefficients represent the variation in the response variable motivated by a change in a given feature; these variations do not always change in one direction. Often, advanced ML models are considered "black boxes" due to their complex inner workings. However, because of their complexity, they are typically more accurate for predicting nonlinear, faint, or rare phenomena (Boehmke and Greenwell, 2020).

Linear models have long been the preferred tools for predictive modelling because they are effortless to interpret. ML models —based on sophisticated nonlinear algorithms— usually make more accurate predictions than conventional regression models, improving the results, but at the expense of reducing interpretability, which produces doubts or suspicions of ML. Nevertheless, nowadays the trade-off between the interpretability and accuracy of predictive models is continually decreasing. Therefore, it is possible to build accurate ML models to be used in place of previously linear ones, which are interpretable at the same time.

Interpretability refers to directly transparent modelling mechanisms; a good explanation is when you can no longer keep asking 'why' (Gilpin et al., 2018). Explainable ML typically refers to post hoc analysis and techniques to understand a previously trained model or its predictions. ML can enhance established analytical practices by increasing prediction accuracy over conventional but highly interpretable linear models. It produces quick, accurate, and unbiased decision making. Computers can theoretically use ML to make objective, data-driven decisions in critical situations such as criminal convictions, medical diagnoses, and college admissions, but interpretability, among other technological advances, is needed to guarantee the promises of correctness and objectivity (Hall and Gill, 2019).

XGBoost is a novel machine-learning algorithm belonging to the family of gradient boosting machines that is nonlinear, highly accurate, and challenging to interpret. It is often associated as black-box ML models, such as multilayer perceptron or neural networks. However, understanding an algorithm is mainly related to the transparency of ML systems; it is crucial to know how an automated decision or classification is made. Nevertheless, we will show that it is explainable by applying explanation techniques after model training.

The interpretability of an ML model is associated with the degree of comprehensiveness of why a particular decision or prediction has been made. For example, in the case of XGBoost, inaccurate automated classification of a company as a failure cannot be interpreted, since the internal mechanisms of the black box decision-making process are unknown. Therefore, predicting failure only solves

Table 2
Best XGBoost hyper-parameter combination identified via 10 cross-validation using all features.

Parameter	Value
Number of rounds or trees (boosting iterations)	98
Maximum individual tree depth	5
Learning rate or shrinkage	0.139
Gamma or minimum loss reduction (regularisation)	9.107
Subsample ratio of features or columns	35.56%
Subsample ratio of observations or rows	31.43%
Minimum sum of instance weight needed in a child	3

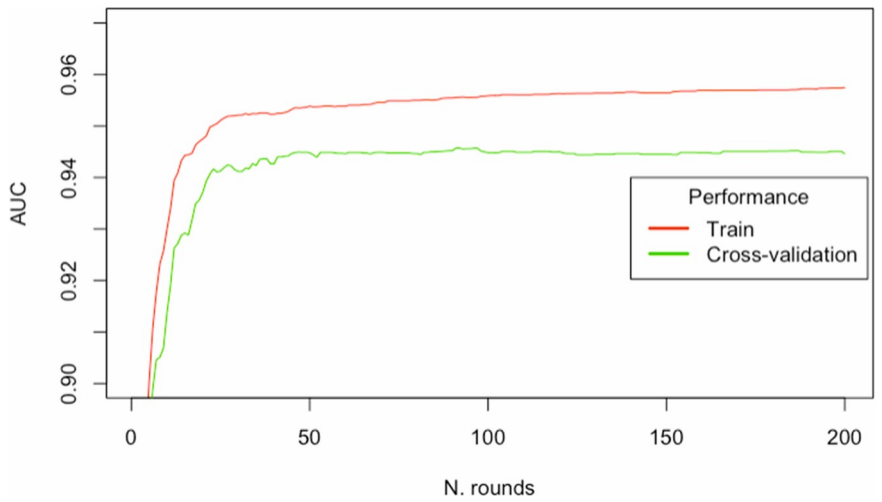
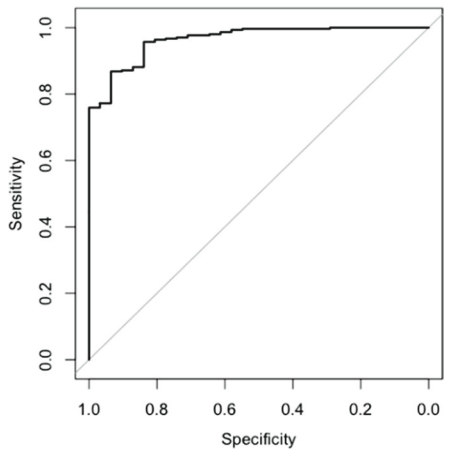


Fig. 6. Training (red) and cross-validation (green) values of AUC as the XGBoost algorithm builds more decision trees (number of rounds), using only the most identified important features. The cross-validation curve is stable and does not change as the number of trees or iterations grows, which may indicate a possible overfitting problem.

AUC (area under the curve)



NOTES: AUC = 0.9644. Accuracy = 87.43%

Confusion matrix

Actual		Predicted: Failed: Yes	Predicted: Failed: No	Accuracy
Failed	Yes	**29**	2	93.55%
	No	40	**263**	86.80%
Total				87.43%

NOTES:
Results for best threshold value of 0.39
N = 334

Fig. 7. XGBoost model performance on test set, using the most important identified features.

the problem partially; however, the process of incorporating machines and algorithms into our lives needs interpretability to achieve confidence and social acceptance (Carvalho et al., 2019). This paper contributes to explaining and understanding a very complex decision model -XGBoost- which could help raise the trust in such algorithms by providing insight into the mechanisms they use and the functions they create.

Approaches to model explanation can be generally classified as offering global or local explanations or interpretability. It is essential to understand the complete model on a global scale and analyse local areas of the data and predictions to derive explanations. Providing both levels of interpretability is critical for the acceptance and adoption of any ML model.

4.4. Global interpretability

Global measures help understand inputs and their overall modelled relationship with the prediction target; in some cases global interpretations can be highly approximate (Hall and Gill, 2019). Global interpretability shows the machine-learned relationships between the output or response variable and the input variables or features across the whole data. Global explanations are about understanding how a model makes predictions based on a complete view of its features and how they impact the underlying model

Fig. 8. Density plot of XGBoost probability predictions by company failure on test set (using most important features). Probabilities are much higher and close to 1 for the failed cases of companies and lower and close to 0 for the non-failed companies' cases (F represents business failure and NF non-business failure). structure. They answer questions regarding which features are more important, how they affect the response variable, and what types of possible interactions arise (Boehmke and Greenwell, 2020).

Model visualisation techniques provide graphical insights into the prediction behaviour of most ML algorithms. One of the first and the most popular tools for the inspection of black-box models at the global level are partial dependence plots, which focus on the general idea of showing how the expected model response behaves as a function of a selected feature, which is an average profile for all observations (Biecek and Burzykowski, 2021). Therefore, such plots provide a global interpretation of how a model's predictions fluctuate based on certain features, showing the average manner in which machine-learned response functions change based on the values of one or two input variables of interest, while averaging out the effects of all other input variables (Hall and Gill, 2019). Although these representations are not a perfect representation of the captured effects, they can provide a sound basis for interpretation, are easy to explain, and are relatively intuitive (Natekin and Knoll, 2013).

Fig. 9 presents the partial dependence plots of the six most critical identified predictors or variables used to build the final XGBoost model. They illustrate the marginal effect of each of the predictors on business failure probability after accounting for the average joint effect of the other predictors. Although this may not provide a complete description, it can reveal general trends and provide a good interpretation basis (Friedman, 2002).

The partial response plots for business failure and the influential variables indicate a positive association between business failure and ratio R35. However, there is a negative relationship for the remaining five ratios (R2, R1, R27, R17 and R5), meaning that on the one hand, the higher the "equity per employee ratio", the "solvency ratio", the "current ratio", the "net profitability" and the "sustainable return on investment ratio," the lower the probability of business failure; on the other hand, the higher the "number of employees", the higher the chance of business failure. These results are consistent with previous studies, indicating that companies with high liquidity, solvency and profitability are less likely to be subject to financial distress risk (Carmona et al., 2019; Climent et al., 2019; Jabeur et al., 2021; Tian and Yu, 2017).

As noted, these results are not contradictory to the economic interpretation of the financial ratios above. For instance, our results indicate that healthy companies have a higher level of liquidity as measured by the current ratio (R27) than companies in default. This may seem obvious, since companies in financial difficulty have difficulties in meeting their financial commitments. Our results also suggest that the lower the firm solvency ratio (R1), the higher the chance the firm will default on its current obligations. This is because firm solvency evaluates its ability to meet the deadlines of debt repayment. In addition, this result is consistent with Carmona et al. (2019), who indicate that larger banks are more vulnerable to financial distress.

Moreover, the importance of the model variable is useful for the evaluation of the explanatory variable significance. The main idea is to measure the extent to which the model fit decreases if the effect of a selected explanatory variable or a group of variables is removed. The effect is removed using perturbations such as resampling from an empirical distribution of the just permutation of the values of the variable (Fisher et al., 2018). The idea is in some sense borrowed from the necessary variable measures proposed for a random forest. If a variable is necessary, then after its permutation, we would expect that model performance will be lower. The greater the drop in performance, the more important the variable. Variable importance is one of the most central aspects of explaining ML models from a global perspective. It is a way of determining how input variables impact predictions, which is decisive to a global explanation.

XGBoost is an ensemble tree-based model, and as Hall and Gill (2019) state, a simple heuristic rule for variable importance in a decision tree is related to the depth and frequency at which a variable is split in a tree, with variables used higher and more frequently in the tree being more important.

The relative weights of the most critical variables are shown in Fig. 10. R35 is the weightiest one and accounts for more than 30% of the model explanations. The density distribution plots of this variable confirm a significant distribution difference between failed and non-failed companies (Fig. 11). Following the importance, we identified variables R1 and R27, both accounting for around 40% of the model explanation. These results are in line with the exploratory analysis discussed above (see Figs. 3, 4 and 5); in particular, the

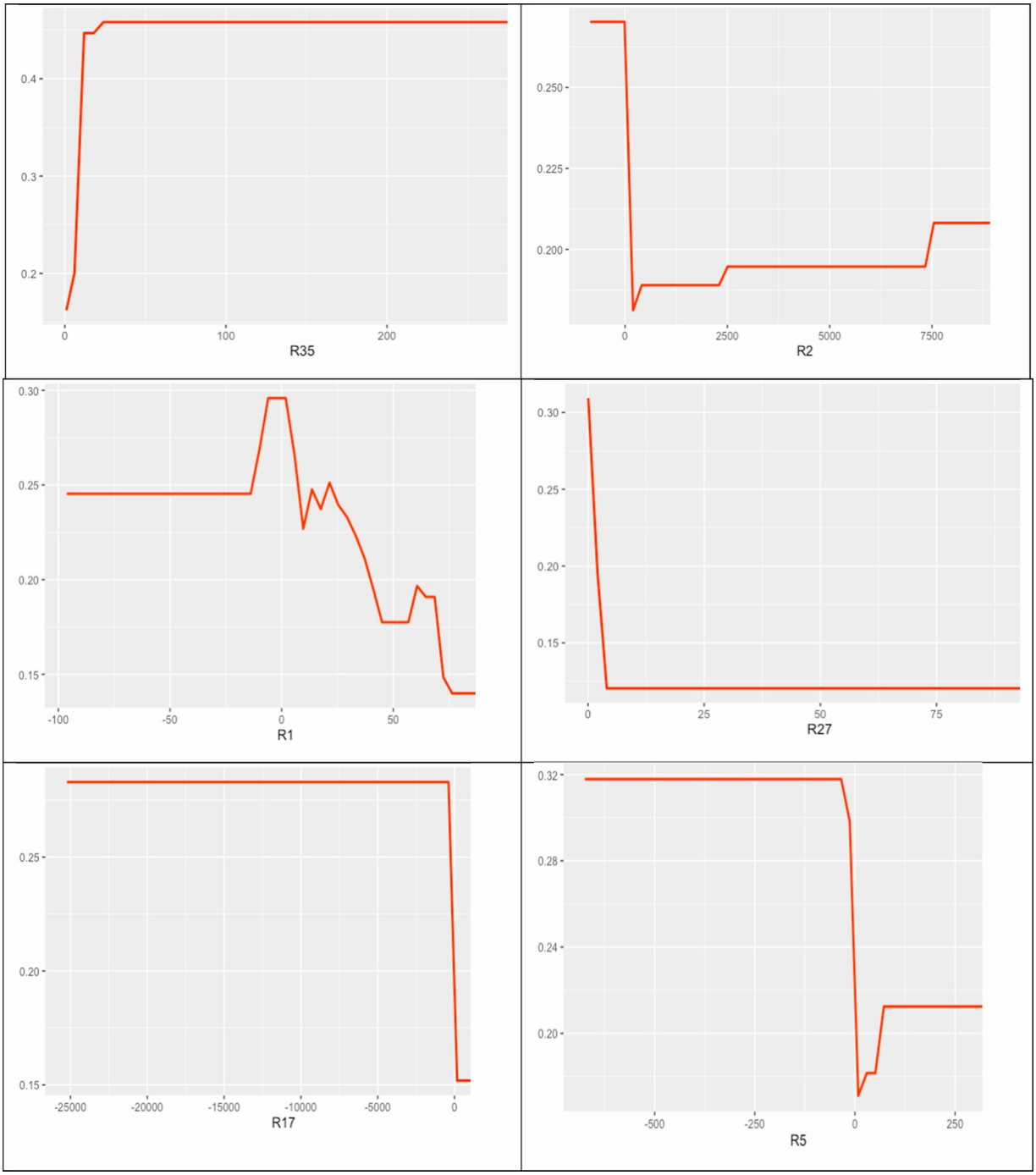


Fig. 9. Partial dependence plots of most important features used to build the final XGBoost model. They illustrate the marginal effect of each one of the predictors on the business failure probability (y axis), after accounting for the average joint effect of the other predictors.

correlational funnel reveals that R35 and R1 are among the highest correlated features with business failure.

In summary, we have demonstrated how global model interpretability helps understand the relationship between business failure and individual ratios. The model is easy to understand because it has been built using only a few features or ratios. On the contrary, a model based on more features would be difficult to fully understand, as it would be difficult to immediately interpret the whole model structure. Furthermore, a global view provides insight into how a model's predictions change based on specific input variables or features.

4.5. Local interpretability

Local information helps understand a model or predictions for a single row of data or a group of parallel rows. Because small parts of a machine-learned response function are more likely to be linear, local information can be more accurate than global information. It is also very possible that the best analysis of an ML model will come from combining the results of global and local interpretation practices (Hall and Gill, 2019). Instance-level or local approaches help to understand how a model produces a prediction for an individual observation.

Sometimes, it is very interesting or necessary to evaluate the effects of explanatory variables on model predictions for a single observation or instance. For example, in this particular study, the researcher or financial analyst might be interested in knowing why XGBoost classifies a given business as a failure based on the values of the predictors. They also might want to understand how model predictions would change if the values of some of the explanatory variables changed. For instance, what would be the predicted risk of business failure if a particular predictor increased or decreased?

The most common question related to the explanation of model prediction for a single instance is 'which variables contributed to this result the most?' The underlying idea is to calculate the contribution of an explanatory variable to the model's prediction as a shift in the expected model response after conditioning on other variables (Biecek and Burzykowski, 2021). We are interested in the chances of the non-failure of a specific business with a particular profile of values for the predictors.

Break-down plots provide a precise summary of the effects of each particular variable on the expected model response and are very easy to understand. They clarify the way an ML model yields a prediction for an individual instance. XGBoost business failure prediction is explained locally in Fig. 12 for four different cases, using break-down plots. The cases at the top show two correctly predicted failed businesses, while those at the bottom show two correctly predicted non-failed businesses. As can be noted, the average model response is 0.215. This is the probability of a company's failure of the test data set averaged over all companies in the set. It is not the fraction of failed companies, but the average model response, so a different model will produce different averages. For example, for the company in the top left of Fig. 12 —an actual failed company— the model prediction of failure is 0.984, which is much higher than the average prediction. A break-down plot shows the contributions of each feature to the prediction and computes them step by step; specifically, it breaks down the probability of failure prediction. Step by step, the probability changes are as follows for this observation:

+ 0.215: Intercept (Baseline)

+ 0.231: R35 = 24 (Number of employees) [prediction is now 0.446]

+ 0.159: R5 = − 96·17 (Sustainable return on investment) [prediction is now 0.605]

+ 0.171: R17 = − 2·22 (Net profitability) [prediction is now 0.776]

+ 0.133: R2 = 2·72 (Equity per employee) [prediction is now 0.909]

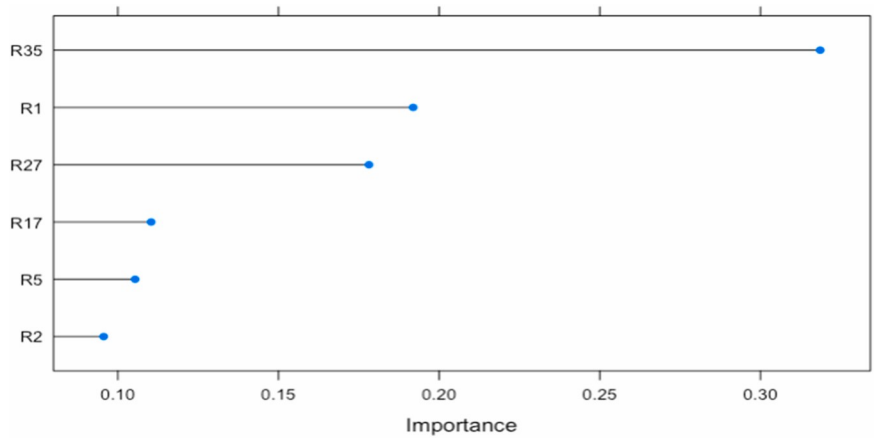


Fig. 10. Variable importance of final XGBoost fitted model using the most relevant identified features.

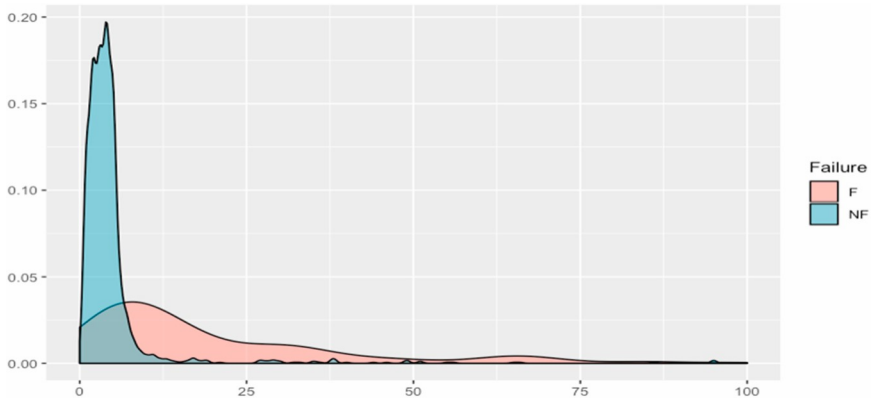


Fig. 11. Density distribution plot of variable R35 by failure, as the most important identified feature in the final XGBoosxt fitted model. Probabilities are much higher for the failed cases of companies and lower for the non-failed companies' cases (F represents business failure and NF non- business failure).

+ 0.064: R27 = 1·51 (Current ratio) [prediction is now 0.973]

+ 0.011: R1 = 6·44 (Solvency ratio) [final prediction is 0.984]

If we consider the company at the bottom right of Fig. 12, which has been classified correctly by the developed XGBoost model as non-failed, the model prediction of failure is very close to 0, which is much lower than the average prediction. For this example, step by step, the probability breaks as follows:

+ 0.215: Intercept (Baseline)

- 0.090: R27 = 4·35 (Current ratio) [prediction is now 0.125]

- 0.053: R1 = 83·22 (Solvency ratio [prediction is now 0.072]

- 0.040: R35 = 3 (Number of employees) [prediction is now 0.032]

- 0.017: R5 = 0.51 (Sustainable return on investment) [prediction is now 0.015]

- 0.002: R2 = 6677 (Equity per employee) [prediction is now 0.013]

- 0.003: R17= 2·47 (Net profitability) [final prediction is 0.010]

Based on these explanation results and comparing the two companies, it can be observed that the model classifies the first company as a failure because the current ratio, the solvency ratio, the sustainable return on investment and net profitability are relatively low, while the number of employees is quite high. On the other hand, XGBoost yields a low business failure probability for the second company because the current ratio, the solvency ratio, the sustainable return on investment and net profitability are quite high, but the number of employees is relatively low.

This disintegration process could be employed for every single company in the test data set to explain the XGBoost predictions thoroughly. According to Foster (2017), decomposition involves adding up the contributions of each feature for every tree in the ensemble, in precisely the same way as for a single decision tree. This means every prediction is stated as the sum of the feature impacts; these impacts are not static coefficients, as in logistic regression. The impact of a feature is dependent on the specific path that the observation took through the ensemble of trees.

In summary, utilising break-down plots, we have attempted to identify which features contribute the most to the XGBoost results and therefore provide an explanation of model prediction for a single instance. Local explanations enhance understanding by creating accurate explanations of each observation in a dataset.

5. Conclusion

The study has offered an example of how to train a black box advanced ML model (i.e., XGBoost) and how it is possible at the same time to interpret the results. Such models are labelled as black boxes due to their hidden internal mechanisms. Frequently, more complex black boxes are more accurate for top predictive performance, while on the other hand, more accuracy is associated with more challenging interpretability. Fortunately, some improvements have been suggested to assess the interpretation of ML models over the years. This study explains how to utilise such improvements to obtain essential insights. Therefore, this study opens up the black boxes and fills the literature gap by identifying the most critical indicators that could assist in anticipating business failure using novel ML techniques, XGBoost, and applying model interpretability. The data were gathered from the Eikon database for 1760 French firms (1585 healthy and 175 failing) in 2018.

Fig. 12. Four different break-down plots. Upper plots belong to two failed companies (probability of failure 0.984 and 0.864) and lower plots belong to two non-failed companies (probability of failure 0.005 and 0.011). They show the contributions of each feature to the prediction, breaking down in the x-axes the prediction probability of failure that returns final fitted XGBoost model. The average model response is 0.215 (intercept).

Our results reveal that the number of employees, equity per employee, solvency ratio, current ratio, net profitability, and sustainable return on investment are the most critical factors that affect business failure. Moreover, our study indicates that higher equity values per employee, solvency, the current ratio, net profitability, and a sustainable return on investment are associated with a lower risk of business failure. In contrast, a higher number of employees increases the chance of business failure.

In the study, we have attempted to identify which features contribute most to the XGBoost results and therefore provide an explanation of the model predictions from a global perspective. The results indicate that company size measured by the number of employees is the most critical factor contributing to business failures. On the other hand, the current ratio and solvency are the most essential measures that reduce the likelihood of business failure; therefore, firm managers should sensitively track these ratios. In addition, net profitability and the sustainable return on investment ratio present critical performance measures that firm managers should track carefully and compare with competitors and the industry average. In addition, underperformance is a crucial signal of possible business failure, and needs a quick reaction from managers, who must be capable of recognising the fundamental reasons for long-term underperformance to address such issues.

Moreover, our local interpretability results enhance the understanding of the financial ratios that contribute to business failures by creating accurate explanations for each observation in our data on failed and non-failed companies. Local explanations concern the chances of the non-failure of a specific business with a particular value profile for the predictors. This fills the gap of evaluating the effects of explanatory variables on business failure model predictions for a single observation or instance.

Our study contributes to the widespread acceptance of ML interpretability techniques, which could help to increase the implementation of ML and AI in business applications and our daily lives. The high predictive and explanatory power of the model developed in this paper could help firm managers to track the most significant attributes detected in the one-year-before-failure model. Moreover, using this approach allows creditors to compare and diagnose their own models of bankruptcy prediction. In addition, our work could help decision-makers and managers consider a large number of ratios to predict bankruptcy, allowing them to achieve higher performance and prevent business failure by taking more accurate and timely decisions instead of waiting for government intervention. Furthermore, our findings promote the significance of XGBoost and the use of AI and ML techniques in the banking sector, as they could provide them with critical information about the financial situation of firms, thus enabling them to grant more loans to the most viable companies. Finally, explaining the results in terms of the indicators is valuable for regulators in pinpointing and warning about potentially distressed businesses and the taking of proactive actions to avoid business failure.

Despite the strengths and importance of the methodological contributions discussed above, the study has certain limitations. First, the sample is based on annual data for the 36 chosen ratios. Future research could improve the accuracy of the predictions by using different sets of financial variables sampled at different time frequencies. Second, this study only uses financial characteristics to predict business failure. Therefore, it could be more beneficial if future investigations considered qualitative information such as managers' experience, the loss aversion feelings of firms, and more strategic variables in thier insolvency prediction models. In addition, it would be interesting for future research to employ interpretable ML models in predicting other industries, such as bank or financial institution failures. Finally, future studies could also replicate this one using a different sample.

Funding

This work was supported by the MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe" (Project PGC2018- 093645-B-I00).

References

Alfaro, E., García, N., Gamez, M., Elizondo, D., 2008. Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. Decis. Support Syst. 45 (1), ´ 110–122. https://doi.org/10.1016/j.dss.2007.12.002.

Altman, E.I.J.T. j o f, 1968. Financ. Ratios, Discrim. Anal. Predict. Corp. Bankruptcy 23 (4), 589–609.

Berryman, J.E., 1993. Small Business Failure and Bankruptcy: What Progress Has Been Made in a Decade? Small Enterp. Res. 2 (1–2), 5–27. https://doi.org/10.5172/ ser.2.1-2.5.

Biecek, P., 2018. Dalex: Explainers for complex predictive models in R. J. Mach. Learn. Res. 19 (84), 1–5, 19. 〈http://jmlr.org/papers/v19/18-416.html〉 .

Biecek, P., Burzykowski, T., 2021. Explan. Model Anal. (Pbiecek.github.io. Available at:) 〈https://pbiecek.github.io/ema/preface.html〉 .

Boehmke, B., Greenwell, B., 2020. Hands-On Machine Learning with R. CRC Press. Taylor & Francis Group,, Florida (USA).

Boubaker, S., Buchanan, B., Nguyen, D.K., 2016. Risk Management in Emerging Markets: Issues, Framework, and Modeling. Emerald Group Publishing,.

Campillo, J.P., Vargas, J.M., Iba´nez, P.C., 2018. Analysis of the algorithm Gradient Boosting Machine (GBM) in business failure prediction. Rev. Esp. De. Financ. Y. ˜ Contab. 47 (4), 507–532. https://doi.org/10.1080/02102412.2018.1442039.

Carmona, P., Climent, F., Momparler, A., 2019. Predicting failure in the US banking sector: An extreme gradient boosting approach. Int. Rev. Econ. Financ. 61, 304–323. https://doi.org/10.1016/j.iref.2018.03.008.

Carvalho, D.V., Pereira, E.M., Cardoso, J.S., 2019. Mach. Learn. Interpret.: A Surv. Methods Metr. Vol. 8.

Chen, T., & Guestrin, C., 2016, XGBoost: A scalable tree boosting system.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H.,. Zhou, T. (2019). xgboost: Extreme gradient boosting. R package version 0.90. 0.2. In.

Climent, F., Momparler, A., Carmona, P., 2019. Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. J. Bus. Res. 101, 885–896. https://doi.org/10.1016/j.jbusres.2018.11.015.

Cort´es, E.A., Martínez, M.G., Rubio, N.G., 2008. FIAMM return persistence analysis and the determinants of the fees charged. Span. J. Financ. Account. / Rev. Esp. De. Financ. Y. Contab. 37 (137), 13–32. https://doi.org/10.1080/02102412.2008.10779637.

Cran.r-project.org. (2020). Introducing Correlation Funnel - Customer Churn Example. Retrieved from 〈https://cran.r-project.org/web/packages/correlationfunnel/vignettes/introducing_correlation_funnel.html〉 .

Davies, D., 1997. Art of Managing Finance, Third ed..,. McGraw-Hill,, NEW YORK.

Doshi-Velez, F., Kim, B., 2017. A Roadmap for a Rigorous Science of Interpretability. arXiv Prepr. arXiv:1702. 08608v1 1–13.

Duan, Y., Goodell, J.W., Li, H., Li, X., 2021. Assessing machine learning for forecasting economic risk: Evidence from an expanded Chinese financial information set. Financ. Res. Lett. https://doi.org/10.1016/j.frl.2021.102273.

Du Jardin, P., 2016. A two-stage classification technique for bankruptcy prediction. Eur. J. Oper. Res. 254 (1), 236–252. https://doi.org/10.1016/j.ejor.2016.03.008. Du Jardin, P., 2017. Dynamics of firm financial evolution and bankruptcy prediction. Expert Syst. Appl. 75, 25–43. https://doi.org/10.1016/j.eswa.2017.01.016.
Dwekat, A., Seguí-Mas, E., Tormo-Carbo, G., 2020. The effect of the board on corporate social responsibility: bibliometric and social network analysis. Econ. Res. ´ -Ekon. Istraz. 33 (1), 3580–3603. https://doi.org/10.1080/1331677X.2020.1776139.
Dwekat, A., Seguí-Mas, E., Zaid, M.A.A., Tormo-Carbo, G., 2021. Corporate governance and corporate social responsibility: mapping the most critical drivers in the ´ board academic literature. Meditari Account. Res. https://doi.org/10.1108/MEDAR-01-2021-1155.
Eling, M., Jia, R., 2018. Business failure, efficiency, and volatility: Evidence from the European insurance industry. Int. Rev. Financ. Anal. 59, 58–76. https://doi.org/10.1016/j.irfa.2018.07.007.
Elith, J., Leathwick, J.R., & Hastie, T., 2008, A working guide to boosted regression trees. In (Vol. 77, pp. 802–813).
Erdogan, B.E., 2013. Prediction of bankruptcy using support vector machines: An application to bank bankruptcy. J. Stat. Comput. Simul. 83 (8) https://doi.org/10.1080/00949655.2012.666550.
Faccia, A., Manni, F., Capitanio, F., 2021. Mandatory esg reporting and xbrl taxonomies combination: Esg ratings and income statement, a sustainable value-added disclosure. Sustain. (Switz.) 13 (16). https://doi.org/10.3390/su13168876.
Faccia, A., Petratos, P., 2021. Blockchain, enterprise resource planning (ERP) and accounting information systems (AIS): Research on e-procurement and system integration. Appl. Sci. (Switz.) 11 (15). https://doi.org/10.3390/app11156792.
Fisher, A., Rudin, C., Dominici, F., 2018. Model Class Reliance: Variable Importance Measures for any Machine Learning Model Class, from the "Rashomon" Perspective. J. Mach. Learn. Res. 20.
Foster, D., 2017, NEW R package that makes XGBoost interpretable. Retrieved from 〈https://medium.com/applied-data-science/new-r-package-the-xgboost- explainer-51dd7d1aa211〉 .
Friedman, J., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. https://doi.org/10.1214/aos/1013203451.
Friedman, J., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2.
Friedman, J., Hastie, T., & Tibshirani, R., 2000, Additive logistic regression: A statistical view of boosting. In (Vol. 28, pp. 337–407).
Goodell, J.W., Kumar, S., Lim, W.M., Pattnaik, D., 2021. Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. J. Behav. Exp. Financ. https://doi.org/10.1016/j.jbef.2021.100577.
Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L., 2018. Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning. arXiv:1806. 00069 1–11.
Hall, P., Gill, N., 2019. In: Sebastopol, C.A. (Ed.), An Introduction to Machine Learning Interpretability, 2nd edition.,. O'Reilly Media, Incorporated, USA.
Hosaka, T., 2019. Bankruptcy prediction using imaged financial ratios and convolutional neural networks. Expert Syst. Appl. 117. https://doi.org/10.1016/j. eswa.2018.09.039.
Jabeur, S.B., Gharib, C., Mefteh-Wali, S., Arfi, W.B., 2021. CatBoost model and artificial intelligence techniques for corporate failure prediction. Technol. Forecast. Soc. Change 166. https://doi.org/10.1016/j.techfore.2021.120658.
Jones, S., 2017. Corporate bankruptcy prediction: a high dimensional analysis. Rev. Account. Stud. 22 (3), 1366–1422. https://doi.org/10.1007/s11142-017-9407-1. Kalak, I.E., Azevedo, A., Hudson, R., Karim, M.A., 2017. Stock liquidity and SMEs' likelihood of bankruptcy: Evidence from the US market. Res. Int. Bus. Financ. 42. https://doi.org/10.1016/j.ribaf.2017.07.077.
Khoja, L., Chipulu, M., Jayasekera, R., 2019. Analysis of financial distress cross countries: Using macroeconomic, industrial indicators and accounting data. Int. Rev. Financ. Anal. 66. https://doi.org/10.1016/j.irfa.2019.101379.
Kim, H.J., Jo, N.O., Shin, K.S., 2016. Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. Expert Syst. Appl. 59, 226–234. https://doi.org/10.1016/j.eswa.2016.04.027.
Le, H.H., Viviani, J.L., 2018. Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. Res. Int. Bus. Financ. 44. https://doi.org/10.1016/j.ribaf.2017.07.104.
Lee, S., Choi, W.S., 2013. A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. Expert Syst. Appl. 40 (8), 2941–2946. https://doi.org/10.1016/j.eswa.2012.12.009.
Momparler, A., Carmona, P., Climent, F., 2016. La Prediccion Del Fracaso Bancario Con La Metodología ´ "Boosting Classification Tree". Rev. Esp. De. Financ. Y. Contab. 45 (1), 63–91. https://doi.org/10.1080/02102412.2015.1118903.
Momparler, A., Carmona, P., Climent, F., 2020. Revisiting bank failure in the United States: a fuzzy-set analysis. Econ. Res. -Ekon. Istraz. 33 (1), 3017 –3033. https://doi.org/10.1080/1331677X.2019.1689838.
Mosteanu, N.R., Faccia, A., 2020. Digital systems and new challenges of financial management – fintech, XBRL, blockchain and cryptocurrencies. Qual. - Access Success 21 (174).
Mousavi, M.M., Ouenniche, J., Xu, B., 2015. Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. Int. Rev. Financ. Anal. 42, 64–75. https://doi.org/10.1016/j.irfa.2015.01.006.
Mselmi, N., Lahiani, A., Hamza, T., 2017. Financial distress prediction: The case of French small and medium-sized firms. Int. Rev. Financ. Anal. 50, 67–80. https://doi.org/10.1016/j.irfa.2017.02.004.
Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Front. Neurorobotics 7, 21.
Oyewo, B., Ajibola, O., Ajape, M., 2020. Characteristics of consulting firms associated with the diffusion of big data analytics. J. Asian Bus. Econ. Stud., Ahead–Print. (Ahead–Print.). https://doi.org/10.1108/jabes-03-2020-0018.
Rapanyane, M.B., Sethole, F.R., 2020. The rise of artificial intelligence and robots in the 4th Industrial Revolution: implications for future South African job creation. Contemp. Soc. Sci. 15 (4), 489–501. https://doi.org/10.1080/21582041.2020.1806346.
Santhanam, G.R., Holland, B., Kothari, S., & Ranade, N., 2017, Human-on-the-loop automation for detecting software side-channel vulnerabilities.
Stolbov, M., Shchepeleva, M., 2020. Systemic risk, economic policy uncertainty and firm bankruptcies: Evidence from multivariate causal inference. Res. Int. Bus. Financ. 52. https://doi.org/10.1016/j.ribaf.2019.101172.
Syam, N., Sharma, A., 2018. Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. Ind. Mark. Manag. 69, 135–146. https://doi.org/10.1016/j.indmarman.2017.12.019.
Tian, S., Yu, Y., 2017. Financial ratios and bankruptcy predictions: An international evidence. Int. Rev. Econ. Financ. 51. https://doi.org/10.1016/j.iref.2017.07.025. Tsai, C.F., Cheng, K.C., 2012. Simple instance selection for bankruptcy prediction. Knowl. -Based Syst. 27, 333–342. https://doi.org/10.1016/j.knosys.2011.09.017.
Vega García, M., Aznarte, J.L., 2020. Shapley additive explanations for NO2 forecasting. Ecol. Inform. 56. https://doi.org/10.1016/j.ecoinf.2019.101039.
Xia, Y., Liu, C., Li, Y.Y., Liu, N., 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. Expert Syst. Appl. 78, 225–241. https://doi.org/10.1016/j.eswa.2017.02.017.
Zhou, L., Si, Y.W., Fujita, H., 2017. Predicting the listing statuses of Chinese-listed companies using decision trees combined with an improved filter feature selection method. Knowl. -Based Syst. 128, 93–101. https://doi.org/10.1016/j.knosys.2017.05.003.
Zięba, M., Tomczak, S.K., Tomczak, J.M., 2016. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. Expert Syst. Appl. 58, 93–101. https://doi.org/10.1016/j.eswa.2016.04.001.