# Applying Machine Learning Techniques to Enhance the Predictive Power of the Altman Z-Score Model in European Union Companies: An Empirical Study

Alice Leitão

Dissertation written under the supervision of professor Nicolò Bertani.

Dissertation submitted in partial fulfilment of requirements for the MSc in Business Analytics, at the Universidade Católica Portuguesa, April 5th, 2023.

## Abstract

Predictive bankruptcy refers to using statistical models and financial analysis techniques to determine the likelihood of a company or organization going bankrupt. This can help investors, lenders, and other stakeholders make informed decisions about their exposure to financial risk.

Altman's Z-Score model is considered an effective tool for predicting bankruptcy and is widely used. However, the model relies only on a few financial ratios as covariates and does not include other variables that may be relevant to predicting bankruptcy.

This thesis starts by analyzing the predictive capabilities of the Altman Z-Score model in a set of European Union countries for private non-manufacturing companies. This is followed by testing if modern machine learning technics can achieve a better prediction when predicting bankruptcy. In the end, which features between a company's financial statement and the ones that Altman used in his model are the most important to consider while predicting bankruptcy.

Altman Z-Score model showed poor prediction capability for bankrupt companies. In contrast, the machine learning models showed increased predictive capabilities by conjugating variables from the Altman Z-Score model and company financial statements. It is possible to build a model with higher precision.

- **Title**: Applying Machine Learning Techniques to Enhance the Predictive Power of the Altman Z-Score Model in European Union Companies: An Empirical Study
- **Author**: Alice Teodoro Cachola Silveira Duarte Leitão
- **Keywords**: Bankruptcy; Altman Z-Score Model; Machine Learning

## Resumo

Previsão de falência refere-se ao uso de modelos estatísticos e técnicas de análise financeira, para determinar a probabilidade de uma empresa entrar em colapso. A utilização deste conjunto de procedimentos ajudar investidores e outras partes interessadas, a tomar decisões informadas sobre o risco financeiro e o grau de exposição dos investimentos associados.

O modelo Z-Score de Altman é considerado uma ferramenta eficaz para prever a falência, sendo amplamente utilizado. No entanto, o modelo conta apenas com alguns índices financeiros e não inclui outras variáveis que possam ser relevantes para a previsão de falência das organizações.

Esta tese começa por analisar as capacidades preditivas do modelo Altman Z-Score num conjunto de países da União Europeia, para empresas privadas não manufatureiras. Em seguida, testamos se as técnicas modernas de aprendizagem automática podem constituir instrumento para obtenção de uma melhor previsão de falência e, quais as características, quer de uma demonstração financeira da empresa, quer do modelo de Altman, se revelam as mais importantes a serem consideradas ao prever a falência.

O modelo Altman Z-Score mostrou uma capacidade preditiva fraca, enquanto os modelos de aprendizagem automática mostraram capacidades preditivas superiores ao conjugar variáveis do modelo Altman Z-Score e demonstrações financeiras da empresa. Concluindo que é possível construir um modelo com maior precisão.

- **Título**: Aplicação de Técnicas de Aprendizagem Automática para Melhorar o Poder Preditivo do Modelo Z-Score de Altman em Empresas da União Europeia: Um Estudo Empírico
- **Autor**: Alice Teodoro Cachola Silveira Duarte Leitão
- **Palavras-chave**: Falência; Altman Z-Score Model; Aprendizagem automática

## Acknowledgements

I want to express my gratitude to my family for allowing me to pursue my master's degree and for their continuous encouragement and support throughout my academic journey.

My sincerest appreciation also goes to my thesis supervisor, Professor Nicolò Bertani, for his invaluable guidance, support, and constant availability.

Additionally, I am grateful to my friends for their motivation and inspiration, which made my journey easier.

Finally, I extend my thanks to all my colleagues at KPMG for their support and encouragement in helping me reach the end of this journey.

## Table of Contents

**List of Figures**



**List of Tables**



**List of Equations**

# 1. Introduction

Predicting bankruptcy is essential for companies. Bankruptcy can have significant consequences for individuals and businesses, so identifying and mitigating the risk of bankruptcy is critical. Bankruptcy or financial distress in companies refers to a situation in which a company struggles to meet its financial obligations. This can occur for various reasons, such as poor management, a decline in the industry or economy, or excessive debt. Companies in financial distress often have difficulty obtaining financing and may struggle to meet debt payments or other financial obligations. They may also experience a decline in their stock price and a loss of investor confidence. (Team, Corporate Finance Institute 2022)

Bankruptcy prediction in organizations involves using financial analysis and modelling techniques to determine whether a company will become insolvent or file for bankruptcy. This can be useful for investors, creditors, and other stakeholders interested in assessing a company's financial health.

There are several methods used for bankruptcy prediction in organizations, such as financial ratio analysis, which is ratios below average that may indicate financial distress, and the Altman Z-Score model, which is a statistical model that calculates a score based on a company's financial ratios, which can be used to predict the likelihood of bankruptcy. A score below a certain threshold is a warning sign of financial distress. (Team, Corporate Finance Institute 2023)

Since the Altman Z-Score model is one of the most used financial models for predicting bankruptcy, this research aims to answer the following questions:

1: Is the predictive bankruptcy accuracy of the original Altman Z-Score model for European Union countries the same as the predictive bankruptcy accuracy shown by Altman in his paper?

2: Can a better model be created using machine learning techniques and more variables?

3: Are Altman variables the essential variables to consider while predicting bankruptcy?

The above research questions were elaborated in three stages. In the first stage, Altman Z-Score model predictive capabilities will be evaluated for private non-manufacturing companies in the European Union countries. In the second stage, the predictive models Linear Regression, Random Forest, Gradient Boosting, and Extreme Gradient Boosting will be tried to determine if it is possible to create a uniform model to predict the bankruptcy of those companies better than the Altman Z-Score model. In the final stage, the research is focused on finding which variables from a company's financial statement are the most important to consider while predicting bankruptcy.

The thesis is divided into sections: Section 2 presents the literature review with bankruptcy context and previous techniques for predicting bankruptcy. Section 3 covers this study's machine learning models and classification metrics. Section 4 includes the dataset exploration, preparation, and metrics to handle imbalanced datasets. Section 5 presents the results and discussion from testing the Altman Z-Score model and the feature importance from the machine learning models evaluated with the corresponding hyperparameters and threshold. The last section, 6, provides a conclusion on this study's findings and limitations.

# 2. Literature Review

## 2.1. Altman Z-Score model

Professor Edward Altman developed the Altman Z-Score financial model in the late 1960s, and it has been widely used for decades. It predicts the possibility of a company going bankrupt in the next two years. (E. I. Altman 1968)

Altman (1968) estimated the model using multiple discriminant analysis (MDA) to derive a linear equation with five financial ratios that classify companies into two groups: those that are likely to go bankrupt and those that are not. (E. I. Altman 1968)

The Altman Z-Score model is calculated using different formulas depending on the type of company being analyzed. After the scores are calculated, companies are categorized into one of the three grades: Safe Zone, Grey Zone, or Red Zone. (E. I. Altman 1968)

The original Altman Z-Score formula intended for public manufacturing companies is calculated using the following formula:

$$Z = 1.2A + 1.4B + 3.3C + 0.6D + 0.99E$$

*Equation 1: Altman Z-Score Formula for Public Manufacturing Companies*

A score above 2.99 indicates that the company is unlikely to go bankrupt, while a score below 1.81 indicates that the company is likely to go bankrupt. (E. I. Altman 1968)

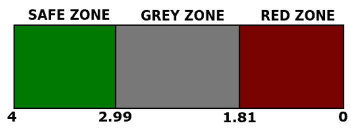



*Figure 1: Public manufacturing companies scale*

For private non-manufacturing companies, the Z-score is calculated using the following formula:

$$Z = 0.717A + 0.847B + 3.107C + 0.42D + 0.998E$$

*Equation 2: Altman Z-Score Formula for Private Non-Manufacturing Companies*

A score above 2.99 indicates that the company is unlikely to go bankrupt, while a score below 1.23 indicates that the company is likely to go bankrupt. (E. I. Altman 1968)

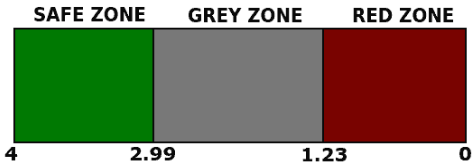



*Figure 2: Private non-manufacturing companies scale*

Where A, B, C, D, and E are the values of the five financial ratios used in the model.

The five ratios used in the Altman Z-Score model are:

A. $\frac{Working\ Capital}{Total\ Assets}$: This ratio measures a company's ability to meet its short-term obligations. A higher ratio indicates a better ability to meet short-term obligations.

B. $\frac{Retained\ Earnings}{Total\ Assets}$: This ratio measures a company's ability to generate and retain profits for future growth. A higher ratio indicates a better ability to create and maintain profits.

C. $\frac{Earnings\ Before\ Interest\ and\ Taxes\ (EBIT)}{Total\ Assets}$: This ratio measures a company's profitability. A higher ratio indicates a more profitable company.

D. $\frac{Market\ Capitalization}{Total\ Liabilities}$: This ratio measures a company's market value relative to its liabilities. A higher ratio indicates a stronger market position.

E. $\frac{Sales}{Total\ Assets}$: This ratio measures a company's ability to use its assets to generate sales. A higher ratio indicates a better ability to create sales.

One of the key strengths of the Altman Z-Score model is its ability to predict bankruptcy accurately. Altman (1968) reports that the model has a high level of accuracy in predicting bankruptcy, with an accuracy rate of 72% in predicting bankruptcy two years prior to the event.

However, the Altman Z-Score model also has some limitations.

1. Lack of flexibility: The model uses a fixed set of financial ratios and weighting factors, which may only be optimal for some companies or industries. This could lead to inaccurate predictions if a different set of financial ratios or weighting factors were more appropriate for a particular company or industry. (Stephen A. Hillegeist 2004)

2. Does not consider qualitative factors: The model is based solely on quantitative financial data and does not consider qualitative factors, such as management quality. (Szmalec 2023)

Overall, the literature on the Altman Z-Score model supports its effectiveness as a tool for predicting bankruptcy. It also highlights the need for further research and refinement of the model to its limitations.

## 2.2. Classical statistical models for bankruptcy

Before the 1980s, bankruptcy prediction models were mostly statistical models that used univariate, multivariate, and logit & probit methods. Perhaps the most recognized methods and authors are:

Discriminant Analysis (DA): A univariate model, first presented by William Beaver in 1966, is a statistical method that analyzes 30 financial ratios to predict the potential of business failure. Beaver analyzed one at a time to predict bankruptcy in 79 companies from 1954 to 1964. (Beaver 1966)

Multiple Discriminant Analysis (MDA): As previously noted, Altman (1968) uses multiple discriminant analyses to predict bankruptcy. In 1972, Deakin developed an MDA model with 14 financial ratios. (Deakin 1972)

Blum, in 1974, employed the MDA model for predicting bankruptcy and reported that it could successfully predict 94% of bankruptcy cases one year before the event. (Blum 1974)

In addition, Dambolena and Khoury built MDA models in 1980 and reached predictive accuracy rates of 87%, 85%, and 78% for one, three, and five years before the bankruptcy, respectively. (Dambolena 1980)

Logit Analysis: Since the early 1980s, Ohlson estimated a logit model using nine independent variables and reported that he could adequately predict around 92% of the bankrupt companies two years earlier. (Ohlson 1980)

Zavgren applied logit analysis to predict bankruptcy 1 to 5 years before the event. While his logit model's accuracy rate for one-year predictions was around the same as Ohlson's 92%, the error rates for long-term predictions were close to or relatively lower than those reported in prior MDA-based bankruptcy prediction studies. (Zavgren 1985)

## 2.3. Machine Learning and Artificial Intelligence models for bankruptcy

The second stage of bankruptcy model evolution began in the 1980s when many machine learning algorithms outperformed the older statistical models. Machine learning models such as Artificial Neural Networks, Decision Trees, Support Vector Machines and Random Forests were particularly effective for bankruptcy prediction.

Artificial Neural Networks (ANN): ANN are a machine learning technique that creates an analogy with human neural processing, and it is one of the most popular artificial intelligence techniques (SheetalSharma 2017). Sharda (1990) conducted the first neural network study in 1990 and found a greater prediction accuracy in bankruptcy prediction, followed by Kim (2010).

Decision Trees: Decision trees are a machine learning technique that uses a tree-like model of decisions and their possible consequences to predict the likelihood of business failure. (Taylor 2023)

Chen (2011) proposed in 2011 a model of financial distress prediction that compares decision tree classification to the logistic regression technique. The decision tree classification approach had better prediction accuracy than the linear regression approach. The same happened to Irimia-Dieguez in 2015, confirming that the performance of the decision tree prediction had outperformed logistic regression prediction (Irimia-Dieguez 2015).

Support Vector Machines (SVM): SVM works by mapping data to a high-dimensional feature space to categorize data points that are otherwise not linearly separable. A separator between the categories is established, and the data are processed so that the separator may be drawn as a hyperplane (IBM 2021). SVM was also found to be a very effective machine learning algorithm by Erdogan in 2012 (Erdogan 2012).

Recent studies have shown the XGBoost algorithm's capabilities in predicting bankruptcies. (Climent 2019)

## 2.4. Different bankruptcy definitions

There are many bankruptcy definitions because bankruptcy laws and practices differ across countries, jurisdictions, and cultures, which may change the definition and application of bankruptcy. Furthermore, specialists and stakeholders, such as lawyers, accountants, economists, and policymakers, may have different opinions and criteria for defining bankruptcy based on their professional knowledge and experience. (Dimitras 1996)

Several authors provide different definitions of bankruptcy. Here are some examples:

According to Altman and Hotchkiss, failure is a realized rate of return on invested capital that is significantly and consistently lower than rates on similar investments taking risk into account. Insolvency is when a company's obligations exceed its assets, meaning it cannot satisfy its present obligations. On the other hand, default happens when a company fails to fulfil a contract, particularly to pay a debt or appear in a lower court. There are two kinds of bankruptcy, according to Altman and Hotchkiss. The first relates to an enterprise's net worth position. The second to the company's formal statement in a federal district court, followed by a petition to liquidate its assets or attempt a recovery program. (E. &. Altman 2006)

Ross, Westerfield, and Jaffe concluded by summarizing earlier studies that there are three kinds of bankruptcy: legal bankruptcy, which means that the company goes to court for a declaration of bankruptcy. Technical bankruptcy, which describes a condition in which a company cannot repay debt and taxes on time, and accounting bankruptcy, which refers to the situation in which a company has negative book net assets. (Ross n.d.)

## 2.5. Crises in European Union

This study uses data from financial statements from 2015 to 2020. During this period European Union (EU) has faced crises that have significantly impacted member states' economies and their ability to avoid bankruptcy. Some of the most notable EU crises include:

Migration crisis: The migration crisis that began in 2015 created significant economic and social challenges for many EU member states. The influx of migrants strained public services and infrastructure and created tensions between member states over how to address the crisis. (Kancs 2018)

Brexit: The UK's decision to leave the EU in 2016 created significant uncertainty for the EU and its member states. The UK was one of the largest economies in the EU, and its departure created challenges for trade and economic cooperation between the UK and the EU. (Felbermayr, et al. 2017)

These crises have significantly impacted member states' economies and have increased the risk of bankruptcy in some cases. For example, the migration crisis also created economic challenges for some member states, particularly those struggling with high debt levels. The uncertainty created by Brexit has also had significant economic impacts, with some sectors, such as financial services, facing particular challenges.

## 3. Machine Learning models

Machine learning is an artificial intelligence field that involves creating and developing algorithms that allow computers to automatically improve their performance in a particular task through training. In other words, machine learning algorithms allow the computer to learn from data, identify patterns, and make decisions without being explicitly programmed.

Machine learning has become increasingly important in the past few years due to the explosion of digital data and the increasing need for systems that can automatically analyze and interpret data. It has been used in various applications, including image and audio recognition, natural language processing, predictive modelling, and virtual assistants.

In summary, machine learning is a crucial component of artificial intelligence that enables computer systems to learn and improve their performance automatically without being explicitly programmed. It has numerous applications in a variety of fields.

Predictive classification is a machine learning type that uses historical data to predict future events. Predictive classification aims to build a model that can automatically assign a new data point to one of several predefined categories or classes based on its similarity to previously observed data.

The data is usually divided into training and testing sets in a predictive classification task. The training set is used to build the model, while the testing set is used to evaluate its performance.

The model is trained by providing it with the features (i.e., the attributes or variables) of each data point and its corresponding class and adjusting its parameters to minimize the classification error.

Once the model has been trained, it can be used to make predictions on new, unseen data. Given a unique data point, the model will compute its features, compare them to the features of the training data, and assign it to the class that is most like it based on some similarity. It is a type of supervised learning, meaning that the algorithm is trained using labelled data and makes predictions based on that training.

In machine learning, several predictive models are commonly used, including Logistic Regression, Random Forests, Gradient Boosting, and Extreme Gradient Boosting (XGBoost).

1. Logistic Regression: Logistic Regression is a predictive model used for binary classification, predicting a binary outcome (e.g., yes or no, 0 or 1). It models the relationship between the dependent variable and one or more independent variables using a logistic function (sigmoid function) to calculate the probability of the dependent variable being 1.

*Equation 3: Logistic Function*

$$p = \frac{1}{1 + e^{-(B_0 + B_1 x_1 + B_2 x_2 + \cdots + B_n x_n)}}$$

This approach continuously analyzes multiple beta values to achieve the optimal log odds fit.

The logistic function is produced through all these iterations, and logistic regression seeks to maximize this function to determine the optimized parameter estimate. After the appropriate coefficients are determined, the conditional probability for each observation is determined, logged, and summed to provide a predicted probability. (IBM n.d.)

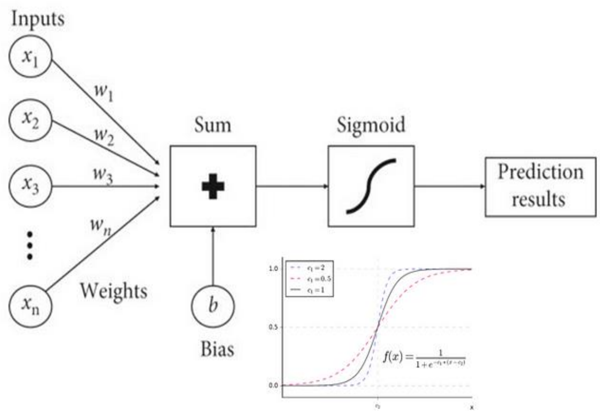



*Figure 3: Logistic Regression Flow*

2. Random Forest: Random Forest is a regression and classification prediction model. It is an ensemble model composed of several independent decision trees, each providing a prediction based on the data's features.

Each decision tree begins with a question. These questions operate as decision nodes in the tree, splitting the data - observations that meet the requirements will follow one path, while those that do not will take the opposite path. To diversify the trees, two processes occur:

The first process, bootstrapping, involves creating different data samples from the training set for each decision tree. Since these samples must be the same size as the original dataset, random data is injected through feature bagging, enabling the trees to develop differently. The next step is feature randomness, which involves randomly filtering the variable considered in each decision tree node, resulting in automatically distinct nodes.

The final prediction is obtained by averaging the predictions of the individual trees in regression or by taking the majority vote in the case of classification. Random forests are known for their robustness, ability to handle complex and non-linear relationships between features, and ability to handle large datasets with many features. (IBM n.d.)

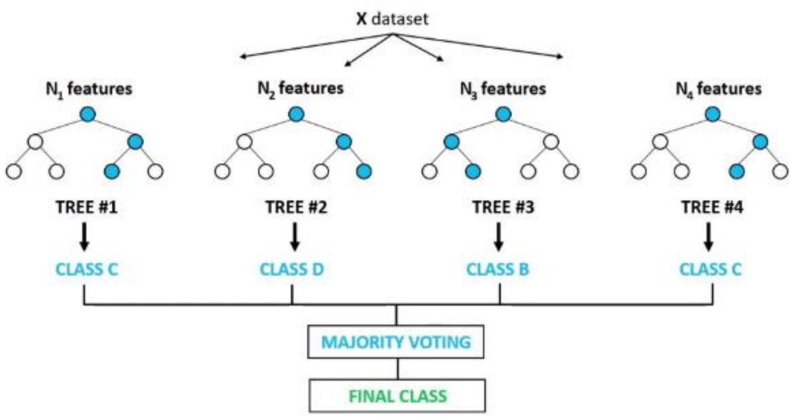



*Figure 4: Random Forest Architecture*

3. Gradient Boosting: Gradient boosting is a predictive model used for regression and classification tasks. It is called gradient boosting because it uses gradient descent to minimize the loss function. An ensemble model builds a series of weak decision trees and then combines them to form a robust model. At each iteration, the gradient boosting algorithm adjusts the weights of the features to focus on the examples misclassified by the current model. This iterative process continues until the model reaches a satisfactory level of performance. Gradient

boosting is known for its ability to handle complex relationships between features and its ability to produce highly accurate models, especially in the presence of large datasets. (Hastie 2009)

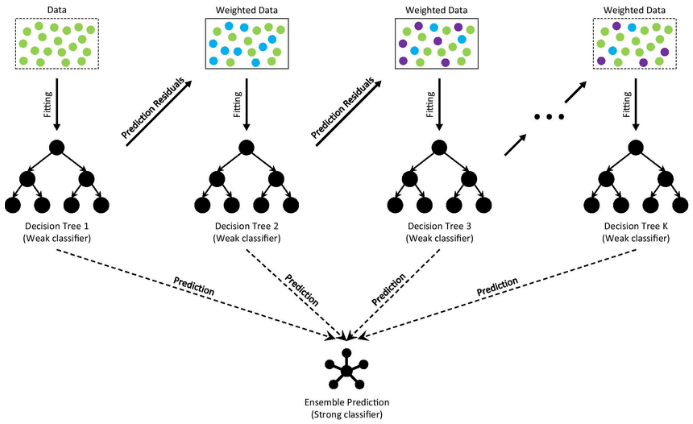



*Figure 5: Gradient Boosting Architecture*

4. Extreme Gradient Boosting (XGBoost): XGBoost is an optimized version of gradient boosting designed to handle large datasets and perform faster than traditional gradient boosting algorithms. It includes several additional features and optimizations that make it more powerful and flexible.

## 3.1. Evaluation metrics for classification models

A confusion matrix is a standard tool used in this classification task. It shows the number of correct and incorrect predictions made by the model for each class in the data.

- True positives (TP): the number of instances correctly predicted as positive (i.e., correctly classified as belonging to the target class).
- False positives (FP), also called Type I errors: is the number of instances incorrectly predicted as positive (i.e., incorrectly classified as belonging to the target class).
- True negatives (TN): the number of instances correctly predicted as negative (i.e., correctly classified as not belonging to the target class).
- False negatives (FN), also called Type II errors: is the number of instances incorrectly predicted as negative (i.e., incorrectly classified as not belonging to the target class).

| | | Predicted | |
|---|---|---|---|
| | | Negative (**N**)<br>- | Positive (**P**)<br>+ |
| **Actual** | Negative<br>- | True Negatives (T**N**) | False Positives (F**P**)<br>**Type I error** |
| | Positive<br>+ | False Negatives (F**N**)<br>**Type II error** | True Positives (T**P**) |

*Figure 6: Confusion Matrix*

In classification models, Type I and Type II errors are concepts used in statistical hypothesis testing.

An error is made when the bankruptcy prediction models incorrectly identify the class to which the company belongs. For example, Type I and Type II errors handle these incorrectly identified companies.

Type I error, a false positive, occurs when a test incorrectly rejects a true null hypothesis. In other words, it happens when a bankrupt or financially distressed company is incorrectly classified as healthy. Type II error, a false negative, occurs when a test fails to reject a false null hypothesis. In other words, it occurs when a healthy company is incorrectly classified as having financial distress.

Type I and Type II errors are essential when evaluating the performance of a classification model. The goal is to minimize the number of Type I and Type II errors so that the model accurately predicts the class of the observations.

To evaluate the models using a confusion matrix, the following metrics are used:

- Accuracy: the proportion of correct predictions, i.e., $\frac{TP+TN}{TP+TN+FP+FN}$
- Precision: the proportion of true positives among all positive predictions, i.e., $\frac{TP}{TP+FP}$
- Recall (also known as sensitivity or true positive rate): the proportion of true positives among all actual positives, i.e., $\frac{TP}{TP+FN}$
- F1 score: the harmonic mean of precision and recall, i.e., $2 \times \frac{precision \times recall}{precision+recall}$

A ROC (Receiver Operating Characteristic) is another metric used to evaluate the performance of a binary classification model. It is a graphical representation of the relationship between the true positive (TP) and the false positive (FP) for different classification thresholds.

A Precision Recall curve is another graphical representation of the performance of a binary classification model, which shows the trade-off between precision and recall for different classification thresholds.

The average precision (AP) is the area under the precision recall curve. An AP of 1 indicates perfect precision and recall, while an AP of 0 indicates that the model is no better than random guessing.

Like the ROC curve, the precision recall curve alone does not provide a complete picture of the model's performance. However, other evaluation metrics, such as the confusion matrix, should also be considered.

# 4. Data

## 4.1. Exploration

Orbis is a comprehensive global database of company information maintained by the business intelligence and data analytics company Bureau van Dijk. The database includes information on more than 400 million private and public companies worldwide. It provides a range of data points such as financial information, company hierarchies, ownership structures, news and media, and industry research. The Orbis database is widely used by financial institutions, government agencies, academic researchers, and other organizations for business intelligence, risk analysis, and other purposes. Access to the Orbis database is available through a subscription.

The study uses the Orbis database, and the population consists of all the private non-manufacturing companies in European Union, with data collected two years before the company status date. The raw dataset includes the year of analysis, company name, country, company financial statement variables and company status, with a total of 50 variables.

| | |
|---|---|
| Year of Analysis | Year when the information was collected: from 2015 to 2020 |
| Company | Company name |
| Country | Company country: only european union |
| Status | 1- Bankrupt, 0- Active |
| Status Date | Year when status was defined: from 2017 to 2022 |

*Figure 7: Some variables from the dataset*

Columns, such as Year of Analysis and Status Date, have been removed as they could not fit for this analysis. Since the aim of this study is to predict bankruptcy within a two-year, regardless of the reporting year or the year of bankruptcy, these columns were considered irrelevant and have been eliminated. The data is cleaned for those whose financial data needs to be completed. The final dataset comprises 10.972 companies, of which 96 are bankrupt from 11 countries. The final dataset ends up with 21.943 instances of active companies and 191 instances of bankrupt companies.

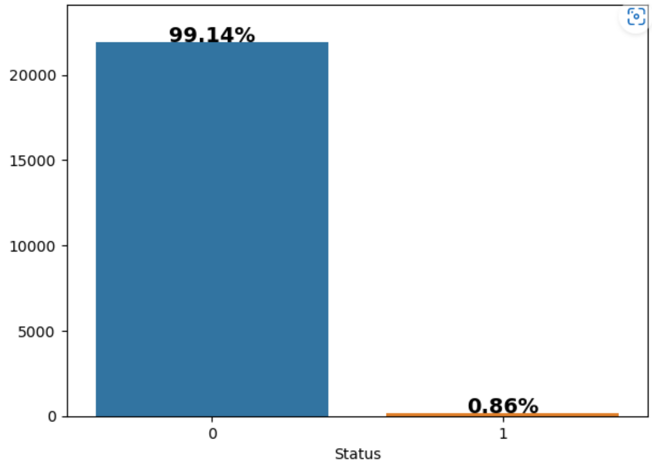



*Figure 8: The distribution shows a strong class imbalance*

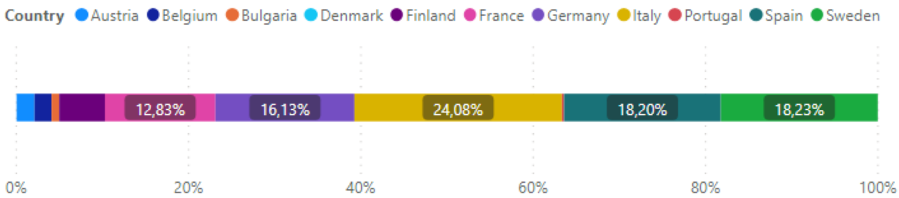



*Figure 9: Distribution of companies by country, Italy has the highest number of companies*

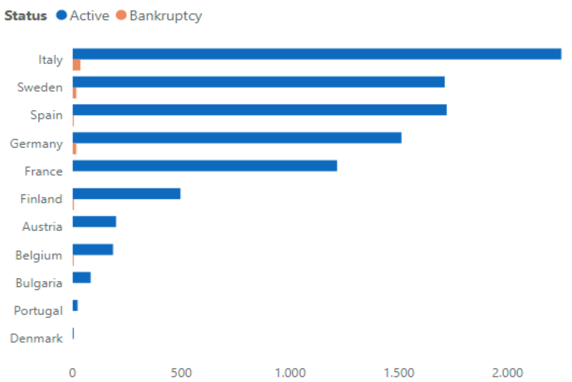



*Figure 10: Number of bankrupt and active companies by country*

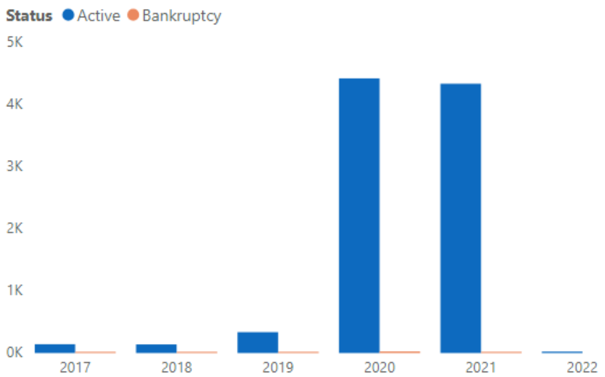



*Figure 11: Number of companies declared bankrupt or active per year, 2020 and 2021 are the predominant years in the dataset*

From the graphical descriptive analysis, it is possible to see that Italy, Sweden and Spain are the dominant countries in the dataset. Also, 2020 and 2021 are the years in which more companies were given as active and, thus, the years with more instances in the dataset.

## 4.2. Preparation

For this thesis's first stage, the Altman model variables were calculated.

X1= $\frac{Working\ Capital}{Total\ Assets}$

X2= $\frac{Retained\ Earnings}{Total\ Assets}$

X3= $\frac{Earnings\ Before\ Interest\ and\ Taxes\ (EBIT)}{Total\ Assets}$

X4= $\frac{Market\ Capitalization}{Total\ Liabilities}$

X5= $\frac{Sales}{Total\ Assets}$

For the second stage, the dataset was subject to a data cleaning and transformation process to ensure the quality of the model.

First, the categorical variable country was converted to a set of dummy variables. To use these variables as input to predictive models, they must first be transformed into a numerical format, with one variable for each country.

Second, an analysis was made to detect possible high correlations between the variables. In some cases, the input variables used in a predictive model may be correlated with each other. This can lead to problems such as overfitting or multicollinearity. Variance inflation factor (VIF) is a statistical measure that assesses the correlation between input variables in a predictive model. VIF can detect multicollinearity between input variables, which can cause problems such as unstable and unreliable estimates of model coefficients, overfitting, and decreased model performance. Calculating the VIF for each variable makes it possible to identify which ones are highly correlated with each other in the model. VIF scores between one and five are considered acceptable and indicate moderate correlation, however scores greater than five are usually considered to indicate high correlation. Since the dataset comprises variables from a balance sheet, several variables were expected to be related. These variables were removed following the Variance Inflation Factor approach. (Pulagam 2020)

### 4.2.1. Handle imbalanced data for Machine Learning classification problems.

Unbalanced datasets in machine learning can pose problems for the performance of a model. For example, when the number of instances belonging to one class is significantly higher than those belonging to the other class(es), the model may be biased towards the class with more instances. This can lead to poor performance in the minority class even if the model's overall accuracy is high.

To address this issue, several techniques can be used, such as:

- Oversampling the minority class by duplicating instances or generating synthetic examples. The idea behind oversampling is to balance the class distribution by making the minority class closer in size to the majority class.
- Undersampling the majority class by removing instances at random or using techniques such as Tomek links or cluster-based oversampling. The idea behind undersampling is to balance the class distribution by making the majority class closer in size to the minority class.
- Using different metrics, such as precision, recall, F1 or F2 score, or the area under the ROC curve, that consider the imbalance in the data.
- Resampling the dataset multiple times with replacement and aggregating the results.

Handling unbalanced datasets properly is essential to build a machine learning model that generalizes well to new unseen data. As well as choosing the proper method for the specific problem, oversampling can lead to overfitting and undersampling can result in loss of information. (Rocca 2019)

The Repeated Stratified K-Fold (RSK) cross-validation method is applied to all the models in this study. It is a technique that can help improve the performance of machine learning models when dealing with imbalanced data. During each iteration, the folds are stratified to ensure that each fold contains a representative sample of the different classes in the dataset. (Brownlee, Machine Learning Mastery 2020)

This study randomly split the data into five folds and repeated it four times. As a result, the models are trained and tested in a diverse set of samples, allowing them to interact with many different sets, which increases generalization to previously unknown data.

To keep addressing this unbalance data issue, an oversampling technic was applied after splitting the data for each model in this study. This technic was only applied to training data to prevent data leakage and avoid overfitting the model to the test data. Data leakage occurs when information from the test set is used to inform the training process, leading to overestimating the model's performance. This makes the model too closely tuned to the test data and needs to perform better on new, unseen data. By oversampling only, the training data, the model can learn from the synthetic samples generated during oversampling without seeing any information from the test set.

SMOTE helps with imbalanced data by generating synthetic examples of the minority class in the dataset, providing more training data for the minority class, which can help balance class distribution and improve the performance of machine learning models. This can reduce biases in the model and make it more accurate and generalizable.

The sampling_strategy parameter in SMOTE determines the desired ratio of the minority class after the SMOTE operation. In this study, sampling_strategy was set as 0.05, which means that the resulting dataset will have 5% of the instances of the minority class from the training dataset. Setting the sampling_strategy parameter to a low value, such as 0.05, in the SMOTE algorithm is a common approach to dealing with imbalanced datasets without over-sampling the minority class too much and potentially causing overfitting. (Brownlee, Machine Learning Mastery 2020)

Setting sampling_strategy to a low value can still result in a relatively imbalanced dataset, which may require other technics.

The synthetic examples are added to the dataset to create a new, less unbalanced dataset with 17.555 active instances and 877 bankrupt instances.

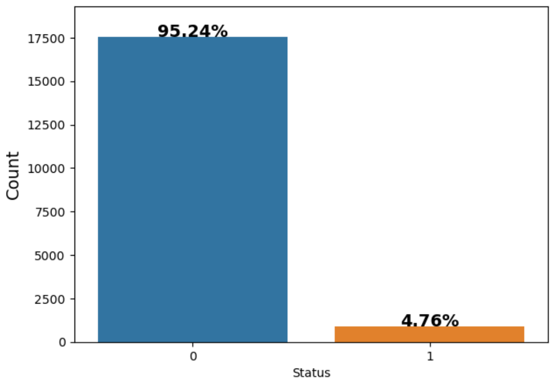



*Figure 12: Training dataset class distribution after SMOTE*

To keep maximizing the model's performance and since the dataset is still unbalanced, GridSearchCV is a popular technique for hyperparameter tuning. It is used to search over hyperparameters for the model. By doing so, GridSearchCV can identify the hyperparameters that optimize the chosen performance metric while considering the class imbalance. Finally, it uses cross-validation to estimate the performance of each hyperparameter combination and returns the best set of hyperparameters based on the performance metric specified. (Great Learning Team 2022)

Since the dataset is still unbalanced, accuracy may not be a good choice of performance metric, as it can be misleading and overly optimistic. This is because the classifier can achieve high accuracy by predicting the majority class for most instances without identifying the minority class instances.

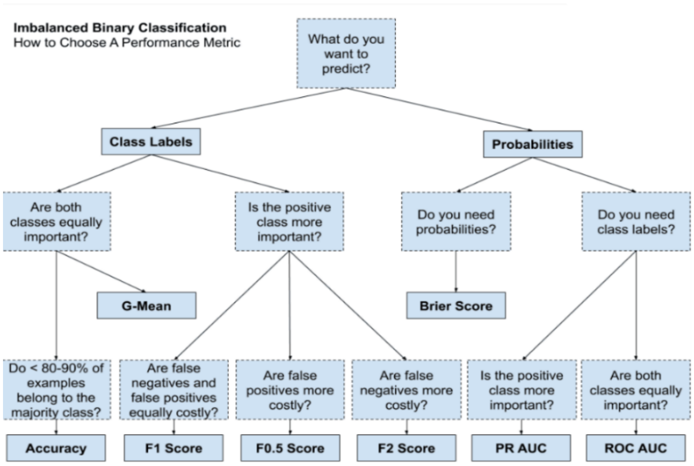



*Figure 13: How to choose the more appropriate performance metric for your problem*

In this study, false negatives are more important than false positives. In bankruptcy, a false negative result (i.e., failing to identify a bankruptcy when it will happen) could have serious consequences. In contrast, a false positive result (i.e., identifying a bankruptcy when it is not going to happen) may only result in more thoughtful decisions in the future.

In situations where false negatives are more important, it may be appropriate to use the F2 score as the performance metric in GridSearchCV. The F2 score places more emphasis on recall (i.e., the ability to identify positive cases correctly) than precision (i.e., the proportion of correct positive predictions), which can help to reduce the number of false negatives and improve the overall performance of the predictive model in situations where false negatives are more important. (Brownlee, Machine Learning Mastery 2020)

During the hyperparameter tuning process, GridSearchCV fits the model with each hyperparameter combination and computes the F2 score using cross-validation. The best set of hyperparameters is the hyperparameter combination that yields the highest average F2 score across all cross-validation folds.

The cv parameter was also used in all the models, and it specifies the cross-validation strategy to use during hyperparameter tuning. In this study, cv=5 specifies that 5-fold cross-validation should be used, which means that the data is split into five equally sized parts, and the model is trained and evaluated five times, with each part used as the test set once.

In classification problems with imbalanced data, thresholding can be a valuable technique to improve the model's performance. Threshold refers to the value that is used to determine whether a predicted outcome is classified as positive or negative. The optimal threshold will depend on the problem and the trade-off between false positives and false negatives. Thus by adjusting the threshold value for the final classification decision, we can balance the trade-off between precision and recall and optimize the F2 score to minimize the false positive rate. (Pocs 2020)

In general, if the class distribution is heavily imbalanced, the model may be biased towards predicting the majority class and have poor performance on the minority class. The default threshold value is usually 0.5, meaning a sample is classified as positive if the predicted probability is greater than or equal to 0.5.

To adjust the threshold value, we can calculate the predicted probabilities for the samples in the test set and then vary the threshold value to see how the performance of the model changes. The precision-recall curve was plotted in this case, and the threshold value that maximizes the F2 was selected for each model.

## 5. Results

This section begins by presenting the results from the Altman model for the companies in this study. The second part of this section presents the results obtained from the chosen models to perform the predictive analysis: Linear Regression, Random Forest, Gradient Boosting, and Extreme Gradient Boosting. At the end of this section, the results from each model with the feature importance technique are present, both for the models that used all variables and those that used only Altman variables.

## 5.1. Testing the Altman Z-Score model for private non-manufacturing companies

The first analysis was testing the Altman Z-Score model for the companies in this study.

For private non-manufacturing companies, the Altman Z-Score model predicted bankruptcy with an accuracy of 44% and a Type I error of 17%, meaning that 17% of actual bankruptcies were classified as healthy. Still, when predicting the active companies, the model showed 38% accuracy and a Type II error of 19%, meaning that 19% of actual active were classified as bankrupt.

| Private Non Manufacturing | Active | Bankruptcy |
|---|---|---|
| Distress Zone | 3785 | 84 |
| Grey Zone | 9809 | 70 |
| Safe Zone | 8349 | 37 |

| Private Non Manufacturing | Active | Bankruptcy |
|---|---|---|
| Distress Zone | 17,25% | 43,98% |
| Grey Zone | 44,70% | 36,65% |
| Safe Zone | 38,05% | 19,37% |

*Figure 14: Altman Z-Score Confusion Matrix for Private Non-Manufacturing Companies*

## 5.2. Machine Learning models results

This section is divided into two parts. In the first part, the hyperparameters selected via GridSearch and the optimal threshold are presented, along with the results and graphics obtained from the models using all financial statements and Altman variables. The second part highlights the same but with the models only using Altman variables.

In each subsection, the first table starts by presenting the models and the respective variables. The first column lists the model's names, the second column the respective hyperparameters used to tune and, the third column contains the threshold that improves the model capabilities. The instance is classified as bankruptcy if the probability score exceeds the threshold. Otherwise, it is classified as active.

To ensure that the models achieve the best possible performance, the following hyperparameters were considered by taking into account the nature of the data:

In logistic regression, the C hyperparameter was considered since it controls the strength of the regularization applied to the model and helps prevent overfitting. A higher value of C means less regularization and a more complex model. The Penalty hyperparameter determines the type of regularization used to prevent overfitting. It was set as l2 in both models meaning that regularization adds the squared values of the coefficients to the cost function.

For both Random Forest models, the class_weight hyperparameter is considered since it assigns weights to the classes in the dataset during model training. This hyperparameter is useful when the dataset is imbalanced. Setting class_weight as balance_subsample helps address the class imbalance issue in random forests. (Amy 2022).

For Random Forest, Gradient Boosting and XGBoost, the max_depth hyperparameter was used to decide the maximum depth the tree will be built, and n_estimators were used to control the number of decision trees.

Max_features, min_samples_leaf and min_samples_split hyperparameters were used in Random Forest and Gradient Boosting. The first one determines the maximum number of features considered for each node in the decision tree. The second controls the minimum number of samples required to be at a leaf node in each decision tree, and min_samples_split controls the minimum number of samples required to split an internal node in each decision tree.

In Gradient Boosting and XGBoost, the learning_rate hyperparameter was also used. This hyperparameter controls how fast the model learns. Min_child_weight is a hyperparameter from XGBoost. It controls the minimum weight required to create a new node in a decision tree during the tree building process.

In the second table of each subsection, the first column lists the name of each model and a brief analysis of the results. The second column includes the model classification report, confusion matrix, ROC AUC, and precision recall curve.

The classification report tells the precision, recall and F1-score for each class in the dataset. The confusion matrix clearly shows the number of true positives, false positives, true negatives, and false negatives for each class in the dataset. ROC AUC curve plots the true positive rate against the false positive rate for different probability thresholds of the classifier. The precision recall curve plots the precision and recall for different probability thresholds of the classifier.

### 5.2.1. Financial statement variables

| **Predictive Model** | **Hyperparameters** | **Threshold** |
|---|---|---|
| Logistic Regression | C: 1; penalty: 'l2' | 0.7 |

| | | |
|---|---|---|
| Random Forest | class_weight='balanced_subsample'; max_depth: 50; max_features: 3; min_samples_leaf: 1; min_samples_split: 2; n_estimators: 100 | 0.3 |
| Gradient Boosting | learning_rate=0.1; max_depth: 50; max_features: 3; min_samples_leaf: 1; min_samples_split: 2; n_estimators: 100 | 0.00469 |
| XGBoost | learning_rate: 0.1; max_depth: 10; min_child_weight: 1; n_estimators: 200 | 0.395191 |

*Table 1: Hyperparameters and threshold used in predictive models for Financial Statement Variables*

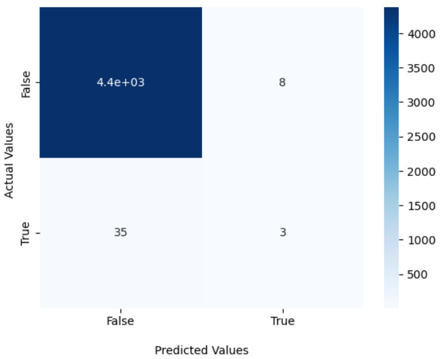
Logistic Regression

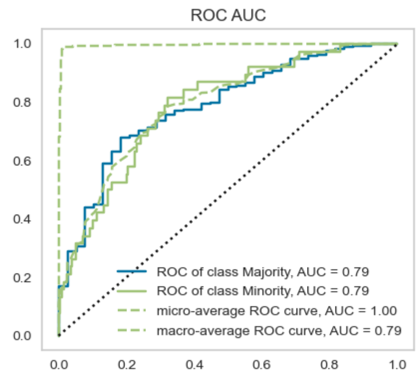
- The model can correctly classify all the observations in around 99%.
- From all the actual bankrupt, the model can correctly predict 27% of those observations.
- From all the active companies, the model can correctly predict 99% of those observations.
- F1 Score for bankruptcy is 12% since the low precision and recall.

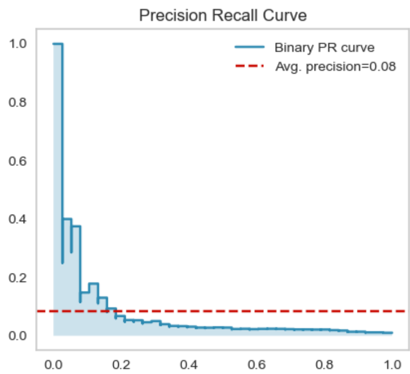
```
              precision    recall  f1-score   support

           0       0.99      1.00      1.00      4388
           1       0.27      0.08      0.12        38

    accuracy                           0.99      4426
   macro avg       0.63      0.54      0.56      4426
weighted avg       0.99      0.99      0.99      4426
```

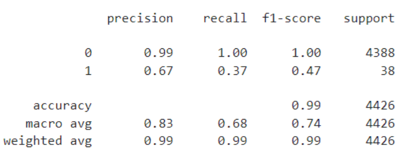


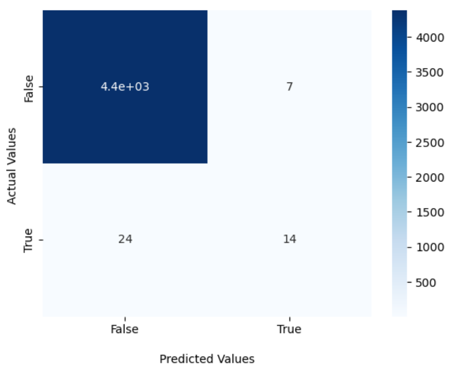



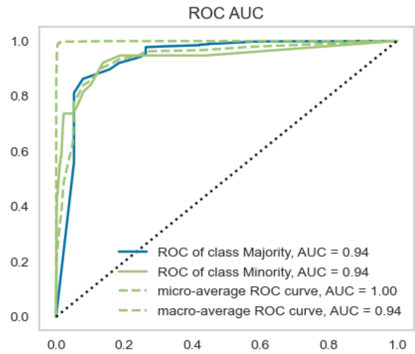
## Random Forest

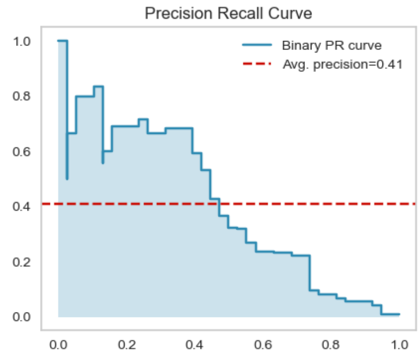
- The model can correctly classify all the observations in around 99%.
- From all the actual bankrupt, the model can correctly predict 67% of those observations.
- From all the active companies, the model can correctly predict 99% of those observations.
- F1 Score for bankruptcy is 47% since the reasonable precision and low recall.



| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.99 | 1.00 | 1.00 | 4388 |
| 1 | 0.67 | 0.37 | 0.47 | 38 |
| accuracy | | | 0.99 | 4426 |
| macro avg | 0.83 | 0.68 | 0.74 | 4426 |
| weighted avg | 0.99 | 0.99 | 0.99 | 4426 |

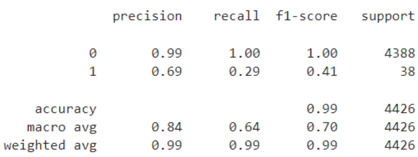


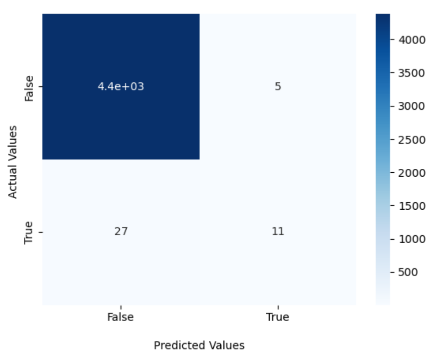



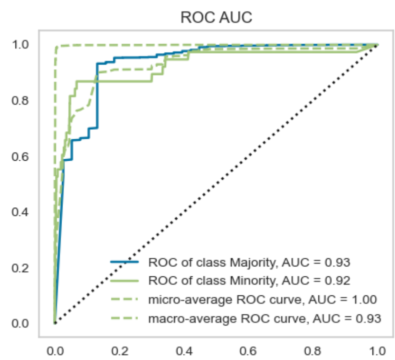

## Gradient Boosting

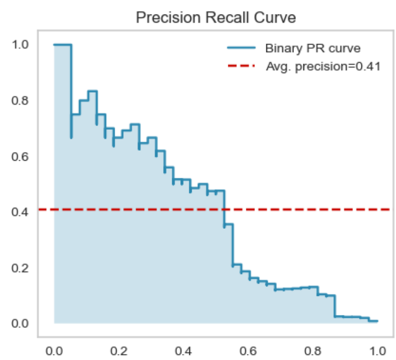
- The model can correctly classify all the observations in around 99%.
- From all the actual bankrupt, the model can correctly predict 69% of those observations.
- From all the active companies, the model can correctly predict 99% of those observations.
- F1 Score for bankruptcy is 41% since the reasonable precision and low recall.



| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.99 | 1.00 | 1.00 | 4388 |
| 1 | 0.69 | 0.29 | 0.41 | 38 |
| accuracy | | | 0.99 | 4426 |
| macro avg | 0.84 | 0.64 | 0.70 | 4426 |
| weighted avg | 0.99 | 0.99 | 0.99 | 4426 |



XGBoost

- The model can correctly classify all the observations in around 99%.
- From all the actual bankrupt, the model can correctly predict 77% of those observations.
- From all the active companies, the model can correctly predict 99% of those observations.
- F1 Score for bankruptcy is 63% since the good precision and reasonable recall.

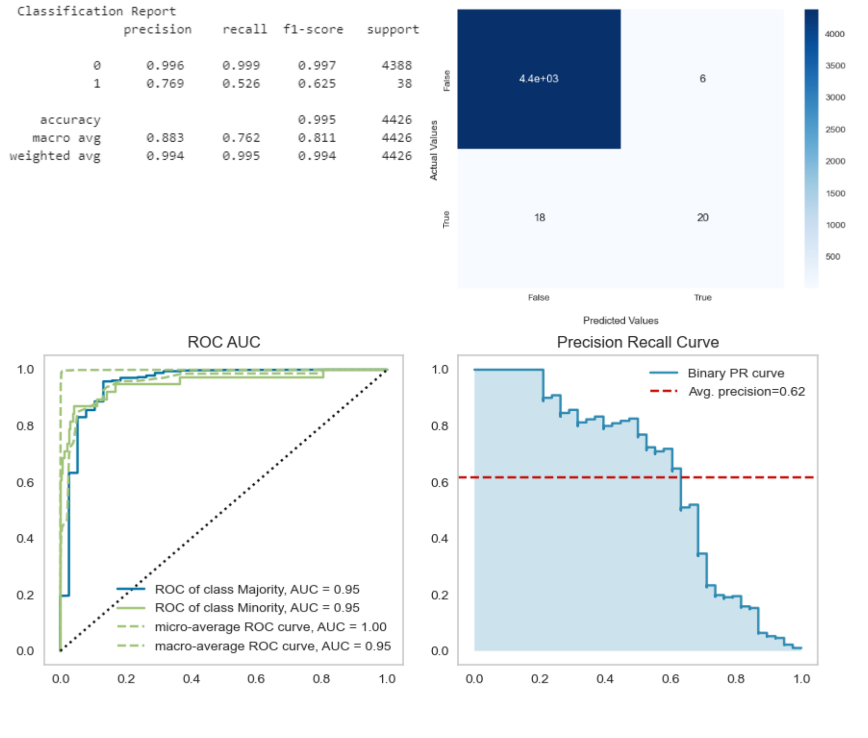



*Table 2: Models Evaluation for Financial Statement Variables*

Overall, the metrics indicate that XGBoost performs better than all the other models, with a capability to predict 77% of actual bankruptcy correctly and an F1 score of 63%.

Gradient Boosting and Random Forest present similar performance metrics capable of predicting actual bankruptcy correctly of 69% and 67% and an F1 Score of 41% and 47%.

Logistic regression comes last with a small capability in predicting bankruptcy of 27%.

### 5.2.2. Altman variables

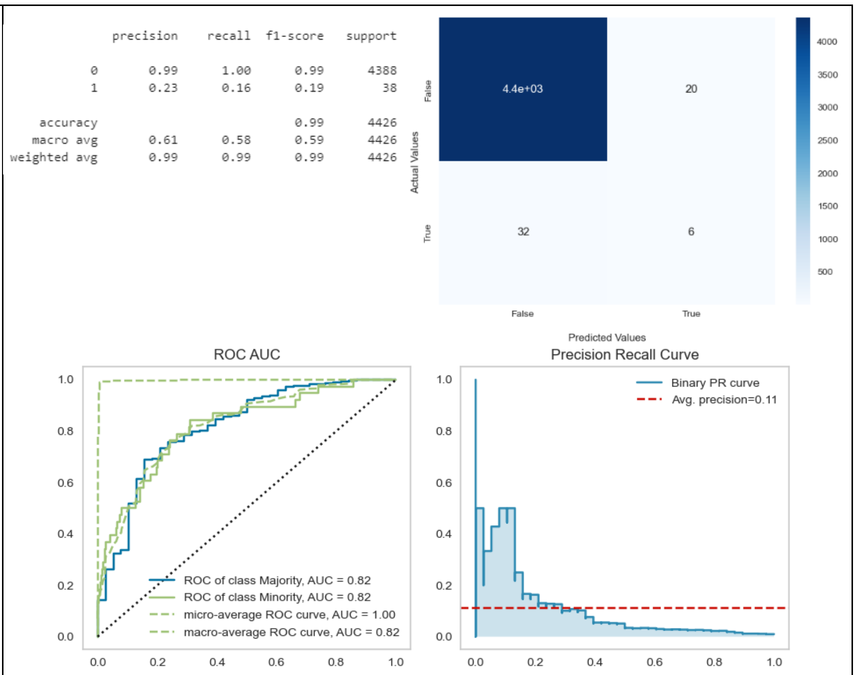
| Predictive Model | Hyperparameters | Threshold |
|---|---|---|
| Logistic Regression | C: 0.01; penalty: 'l2' | 0.115512 |
| Random Forest | class_weight='balanced_subsample'; max_depth: 100; max_features: 'sqrt'; min_samples_leaf: 1; min_samples_split: 2; n_estimators: 100 | 0.4 |

| | | |
|---|---|---|
| Gradient Boosting | learning_rate=0.1; max_depth: 100; max_features: 'sqrt'; min_samples_leaf: 1; min_samples_split: 2; n_estimators: 100 | 0.030881 |
| XGBoost | learning_rate: 0.1; max_depth: 10; min_child_weight: 1; n_estimators: 200 | 0.132161 |

*Table 3: Hyperparameters and threshold used in predictive models for Altman Variables*

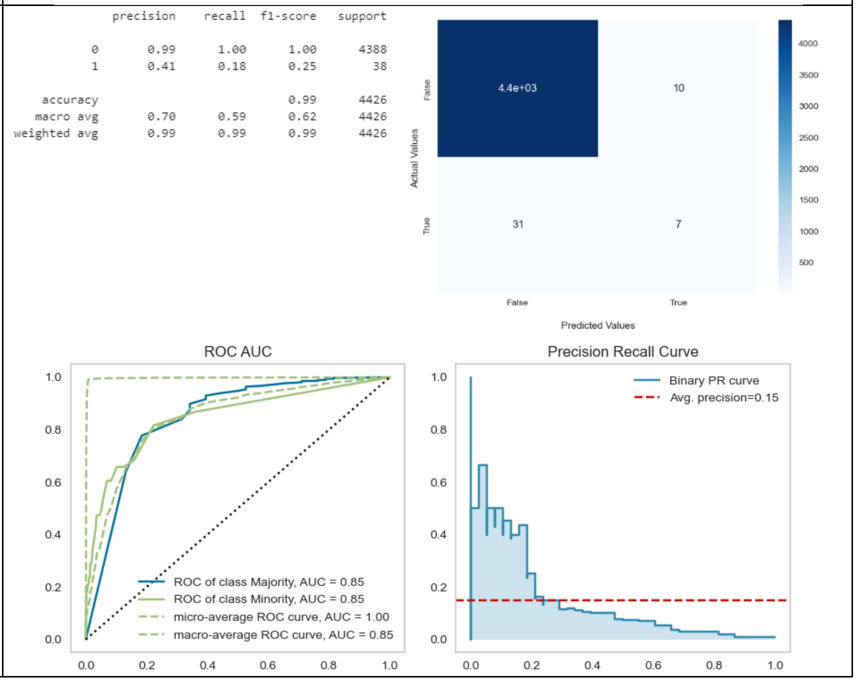
Logistic Regression

- The model can correctly classify all the observations in around 99%.
- From all the actual bankrupt, the model can correctly predict 23% of those observations.
- From all the active companies, the model can correctly predict 99% of those observations.
- F1 Score for bankruptcy is 12% since the low precision and recall.




Random Forest

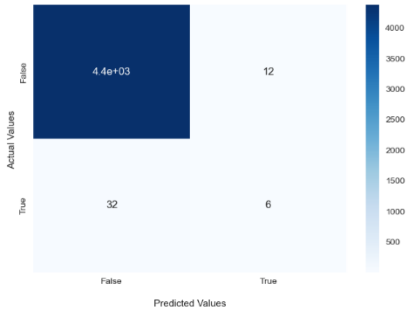
- The model can correctly classify all the observations in around 99%.
- From all the actual bankrupt, the model can correctly predict 41% of those observations.
- From all the active companies, the model can correctly predict 99% of those observations.
- F1 Score for bankruptcy is 25% since the reasonable precision and low recall.

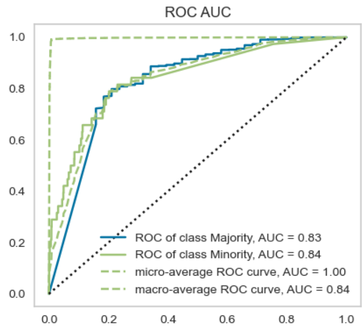


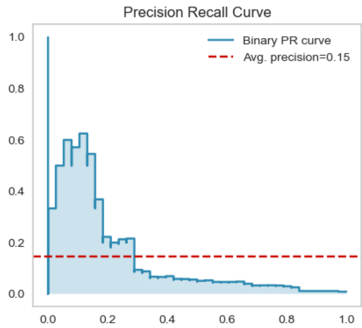
**Gradient Boosting**

- The model can correctly classify all the observations in around 99%.
- From all the actual bankrupt, the model can correctly predict 33% of those observations.
- From all the active companies, the model can correctly predict 99% of those observations.
- F1 Score for bankruptcy is 21% since the reasonable precision and low recall.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.99 | 1.00 | 0.99 | 4388 |
| 1 | 0.33 | 0.16 | 0.21 | 38 |
| accuracy | | | 0.99 | 4426 |
| macro avg | 0.66 | 0.58 | 0.60 | 4426 |
| weighted avg | 0.99 | 0.99 | 0.99 | 4426 |



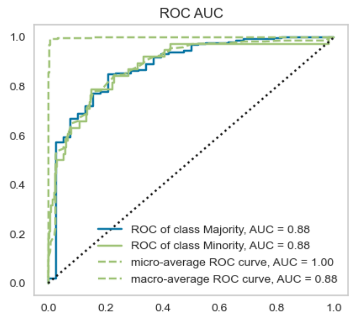




**XGBoost**

- The model can correctly classify all the observations in around 99%.
- From all the actual bankrupt, the model can correctly predict 44% of those observations.
- From all the active companies, the model can correctly predict 99% of those observations.
- F1 Score is 26% since the reasonable precision and low recall.

Classification Report

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.993 | 0.998 | 0.995 | 4388 |
| 1 | 0.438 | 0.184 | 0.259 | 38 |
| accuracy | | | 0.991 | 4426 |
| macro avg | 0.715 | 0.591 | 0.627 | 4426 |
| weighted avg | 0.988 | 0.991 | 0.989 | 4426 |

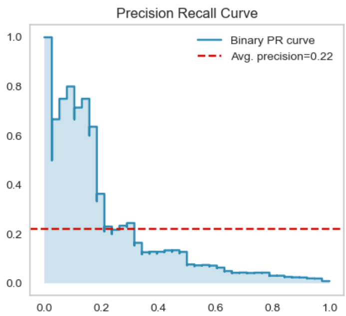




*Table 4: Models Evaluation for Altman Variables*

This subsection allows us to analyze how a machine-learning model works with the same variables as Altman to compare the results fairly.

Overall, there is a high decrease in performance metrics results, similar to what this study got when testing the Altman Z-Score model.

Every model showed an inferior result when predicting b8iankruptcy, with XGBoost predicting only 44% of actual bankruptcy, followed by Random Forest and Gradient Boosting with the capability of correctly predicting 41% and 33%. F1 score also decreased significantly, with results of 26% for XGBoost, 25% and 21% for Random Forest and Gradient Boosting.

Logistic regression comes last with a small capability in predicting bankruptcy of 23%.

## 5.3. Feature Importance

This section is divided into two subsections. The first subsection presents the results of the feature importance of the models that considered both financial statement variables and Altman variables. The second subsection presents the results of the feature importance of the models that only considered Altman variables.

In the tables, the first column lists the name of each model, followed by the second column that presents the highest scores, and the third column that displays the corresponding variable.

Feature importance measures how much each input feature contributes to a model's predictions. Feature importance can vary depending on the model and dataset.

Logistic models: These models can use coefficients to determine feature importance. The larger the absolute value of a coefficient, the more critical the corresponding feature.

Random Forests: These models calculate feature importance by measuring how much each feature reduces impurity in the data. The most valuable features for lowering impurity are considered the most important.

Gradient Boosting models: These models calculate feature importance based on how often a feature is used to construct decision trees during training. Features that are used more frequently are considered more important.

Extreme gradient Boosting models: These models calculate feature importance based on the contribution of each feature to the tree-based model. Specifically, it calculates the gain, or the improvement in accuracy, resulting from splitting a particular feature. The higher the gain, the more important the feature.

### 5.3.1. Financial statement variables

| Models | Score | Variables |
|---|---|---|
| Logistic Regression | 2.2<br>1.1 | Interest paid<br>PL for period Net income |
| Random Forest | 0.09908 | Other shareholders funds |

| | | |
|---|---|---|
| | 0.09878<br>0.09131<br>0.06435 | X2<br>X4<br>PL for period Net income |
| Gradient Boosting | 0.08276<br>0.08065<br>0.05982 | X4<br>Other shareholders funds<br>X2 |
| XGBoost | 0.08225<br>0.06459<br>0.06016<br>0.05199<br>0.05095 | X4<br>Country_Finland<br>Other shareholders funds<br>X2<br>Country_Sweden |

*Table 5: Feature Importance for Financial Statement Variables*

Based on the results, it is possible to conclude that X4 ($\frac{Market\ Capitalization}{Total\ Liabilities}$) has a significant presence in the best three models. Market capitalization is the total value of a company's outstanding shares of stock, while total liabilities are the sum of all the debts and obligations a company owes. Dividing market capitalization by total liabilities gives the market-to-liability ratio, which measures how well a company can meet its obligations to creditors.

The variable X2 ($\frac{Retained\ Earnings}{Total\ Assets}$) is also present in the best tree models, as the second more critical for both XGBoost and Random Forest and in third place for Gradient Boosting. Retained earnings are the portion of a company's profits that the company keeps instead of being paid out as dividends to shareholders. On the other hand, total assets are the sum of all the assets owned by a company, including both tangible and intangible assets. The ratio of retained earnings to total assets can be used to measure the earnings a company has reinvested in its business relative to its total asset base.

Another variable strongly present in the models' prediction is the Other shareholder's funds. Other shareholder funds refer to the residual interest in a company's assets after deducting liabilities. It represents the money left for shareholders if the company liquidates all its assets and pays off all its debts. When predicting bankruptcy, other shareholders' funds can provide insights into the company's financial health and shareholder support. If shareholders have confidence in the company's management and business strategy, they are more likely to invest in the company and increase its equity position. On the other hand, if shareholders lose confidence in the company, they may sell their shares, which can decrease the company's equity position.

The XGBoost model has Finland and Sweden as the second and fifth most significant variables, respectively, meaning that these features (that indicate the country of the company being analyzed) contribute significantly to effectively predict that a company would bankrupt within two years. However, it is essential to note that high feature importance does not necessarily mean that these features are causing bankruptcy directly. Instead, these features strongly relate to the dependent variable (bankruptcy) in the dataset used to train the model.

### 5.3.2. Altman variables

| **Models** | **Score** | **Variables Importance** |
|---|---|---|
| Logistic Regression | -1.05 | X4 |
| | -0.21 | X1 |
| | 0.19 | X3 |
| | -0.1 | X2 |
| | -0.09 | X5 |
| Random Forest | 0.30352 | X4 |
| | 0.23394 | X2 |
| | 0.15920 | X3 |
| | 0.15464 | X1 |
| | 0.14869 | X5 |
| Gradient Boosting | 0.31043 | X4 |
| | 0.21415 | X5 |
| | 0.17742 | X1 |
| | 0.15695 | X2 |
| | 0.14105 | X3 |
| XGBoost | 0.30483 | X4 |
| | 0.18895 | X5 |
| | 0.18479 | X2 |
| | 0.17189 | X1 |
| | 0.14953 | X3 |

*Table 6: Feature Importance for Altman Variables*

Just like the result of the models in the previous section, X4 ($\frac{Market\ Capitalization}{Total\ Liabilities}$) remains the most significant variable for all the models, followed by X5 ($\frac{Sales}{Total\ Assets}$) for XGBoost and Gradient Boosting. For Random Forest, X2 ($\frac{Retained\ Earnings}{Total\ Assets}$) is the second most significant variable.

This is different from Altman's model, where X3 ($\frac{Earnings\ Before\ Interest\ and\ Taxes\ (EBIT)}{Total\ Assets}$) is considered the most important variable for bankruptcy prediction, and X4 ($\frac{Market\ Capitalization}{Total\ Liabilities}$) is considered the least significant variable.

## 6. Conclusion

After conducting a thorough analysis of the Altman Z-Score model for private non-manufacturing companies in European Union countries, this research has shown that the model's performance is poor. However, the findings reveal that machine learning techniques combined with other variables outperform Altman's model. While Altman's variables remain significant in predicting bankruptcy, including additional variables from a company's financial statement enhances the model's accuracies.

In his paper, Altman claims 72% accuracy in predicting bankruptcy, but in this study, his model resulted in only 44% accuracy in predicting bankruptcy within a two-year.

The second hypothesis of this research aimed to investigate the possibility of developing a machine learning model that could outperform the Altman Z-Score model. The results show that the Random Forest, Gradient Boosting, and XGBoost models, incorporating additional variables beyond Altman's, all demonstrated better performance. Among these models, XGBoost achieved the highest accuracy, correctly predicting 77% of all bankruptcy observations, with an F1-Score of 63%. These findings suggest that machine learning techniques, combined with the inclusion of relevant financial statement variables, can improve the accuracy of bankruptcy prediction models.

The third hypothesis of this research aimed to investigate whether the Altman variables are the most appropriate when predicting bankruptcy. The findings indicate that X4, representing the Market Capitalization to Total Liabilities ratio, was consistently the most significant variable in all the best-performing models. This suggests that a company's ability to meet its financial obligations to creditors can serve as a signal for evaluating its overall health. The X2 variable, which represents the ratio of Retained Earnings to Total Assets, was also present in some models as the second most critical variable. This highlights the importance of a company's reinvestment of earnings in its business relative to its asset base.

Additionally, the Other Shareholder's Funds variable demonstrated a strong presence in the models. This variable represents the money left for shareholders if the company liquidated all its assets and paid off all its debts.

Based on the analysis of the machine learning models using only the Altman variables, it was found that the significance assigned to each variable varied between the machine learning model and the Altman Z-Score model. In other words, even though both models used the same variables, the importance of these variables differed between them.

These findings suggest that while the Altman variables remain relevant, incorporating additional variables can enhance the accuracy of bankruptcy prediction models.

The XGBoost models feature importance analysis identified Finland and Sweden's as significant variables, highlighting the crucial role of these countries in predicting bankruptcy. It indicates that these countries have a strong relationship with the dependent variable (bankruptcy) in the training dataset. Therefore, a comprehensive analysis, and a financial interpretation of the results are necessary to comprehend why these dummy variables are significant and what they reveal about the risk of bankruptcy. Further investigation could provide some insight into the underlying reasons for the association between these countries and bankruptcy risk and offer, if the case, insights into potential risk mitigation strategies for companies from these countries.

In conclusion, this research demonstrates that machine learning models with additional variables outperform traditional models like the Altman Z-Score in predicting bankruptcy for private non-manufacturing companies in European Union countries. The Random Forest, Gradient Boosting, and XGBoost models performed better, with XGBoost achieving the highest bankruptcy accuracy. These models also highlighted the significance of variables beyond Altman's model, such as the Market Capitalization to Total Liabilities ratio and the Retained Earnings to Total Assets ratio. Additionally, the XGBoost model identified Finland and Sweden as significant variables, indicating the potential usefulness of geographic analysis in bankruptcy prediction.

Machine learning models can process vast amounts of data and identify patterns that may not be apparent to human analysts. However, traditional models like the Altman Z-Score are still valuable due to their simplicity and accessibility to a broader audience. Therefore, it is essential to consider the strengths and weaknesses of each model when selecting a tool for bankruptcy prediction.

Overall, this research suggests that incorporating additional variables and using machine learning techniques can enhance the accuracy of bankruptcy prediction models for private non-manufacturing companies in the European Union.

## 6.1. Limitations of the study and future work

The study has several limitations that need to be considered. Firstly, the sample size is relatively small, with only 191 bankrupt companies and 21,943 active companies, which may lead to a lack of representativeness and missing meaningful connections among data. Furthermore, the class imbalance issue raises concerns regarding the reliability of the predictive models. Secondly, the predictive models do not consider external factors such as economic conditions, competition, industry changes, and regulatory issues, which can significantly impact a company's financial performance. Additionally, the study only focuses on private non-manufacturing companies, which limits its generalizability to other types of companies.

Future research could explore the unique characteristics of the business environment or market conditions in Scandinavian countries, such as Finland and Sweden, to better understand the association between this specific features and bankruptcy risk. A geographic analysis could also be conducted to determine whether regional factors affect bankruptcy risk. Ultimately, these findings could help central banks, such as European Central Bank to better monitor and evaluate risk management strategies to mitigate bankruptcy risk in certain regions.

## 7. References

Altman, E., & Hotchkiss, E. 2006. *Corporate financial distress and bankruptcy.* John Wiley & Sons.

Altman, Edward I. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *The Journal of Finance* 589–609.

Amy. 2022. *Medium.* https://medium.com/grabngoinfo/balanced-weights-for-imbalanced-classification-465f0e13c5ad.

Beaver, W. 1966. "Financial ratios as predictors of failure." *Journal of Accounting Research* 71–111.

Blum, M. 1974. " Failing company discriminant analysis." *Journal of Accounting Research* 1-25.

Brownlee, Jason. 2020. *Machine Learning Mastery.* https://machinelearningmastery.com/repeated-k-fold-cross-validation-with-python/.

—. 2020. *Machine Learning Mastery.* https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/.

—. 2020. *Machine Learning Mastery.* https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/.

Chen, M.Y. 2011. "Predicting corporate financial distress based on integration of decision tree." *ELSEVIER* 11261-11272.

Climent, F., Momparler, A., & Carmona, P. 2019. "Predicting failure in the U.S. banking sector: An extreme gradient boosting approach." *Journal of Business Research* 885–896.

Dambolena, I. G., & Khoury, S. J. 1980. "Ratio stability and corporate failure." *Journal of* 1017-1026.

Deakin, E. B. 1972. "A discriminant analysis of predictors of business failure." *Journal of Accounting Research* 167-179.

Dimitras, A., Zanakis, F., & Zopoudinis, C. 1996. "A survey of business failure with an emphasis on failure prediction methods and industrial applications." *European Journal of Operational Research* 487-513.

Erdogan, Birsen Eygi. 2012. "Prediction of bankruptcy using support vector machines: An application to bank bankruptcy." *Journal of Statistical Computation and Simulation.*

Felbermayr, Gabriel, Clemens Fuest, Jasmin Katrin Gröschl, and Daniel Stöhlker. 2017. "Economic Effects of Brexit on the European Economy." *EconPol Policy Report.*

Great Learning Team. 2022. *Great Learning.* https://www.mygreatlearning.com/blog/gridsearchcv/.

Hastie, T., Tibshirani, R., & Friedman, J. 2009. *The elements of statistical learning: Data mining, inference, and prediction (2nd ed.).* Springer.

2021. *IBM.* https://www.ibm.com/docs/en/spss-modeler/saas?topic=models-how-svm-works.

n.d. *IBM.* https://www.ibm.com/topics/logistic-regression.

n.d. *IBM.* https://www.ibm.com/topics/random-forest.

Irimia-Dieguez, A.I. and Blanco-Oliver Vazquez-Cueto. 2015. "A comparison of classification/ regression trees and logistic regression in failure models”." *Procedia Economics and Finance* 9 – 14.

Kancs, d'Artis, and Patrizio Lecca. 2018. "Long-term social, economic and fiscal effects of immigration into the EU: The role of the integration policy." *The World Economy* 2599-2630.

Kim, M.J., & Kang, D.K. 2010. " Ensemble with neural networks for bankruptcy prediction." *Expert Systems with* 3373-3379.

Ohlson, J. A. 1980. "Financial ratios and the probabilistic prediction of bankruptcy. ." *Journal of accounting research* 109-131.

Pocs, Mate. 2020. *Medium - Data Driven Investor.* https://medium.datadriveninvestor.com/handling-imbalanced-data-with-basic-classifier-models-5ce3d61874f1.

Pulagam, Sushmitha. 2020. *How to detect and deal with Multicollinearity.* https://towardsdatascience.com/how-to-detect-and-deal-with-multicollinearity-9e02b18695f1.

Rocca, Baptiste. 2019. *Towards Data Science.* https://towardsdatascience.com/handling-imbalanced-datasets-in-machine-learning-7a0e84220f28.

Ross, S., Westerfield, R., & Jaffe, J. n.d. *Corporate finance.* McGraw-Hill/Irwin.

Sharda, M. D. Odom and R. 1990. "A neural network model for bankruptcy prediction." *International Joint Conference on Neural Networks* 163-168 vol.2.

SheetalSharma. 2017. *Data Science Central.* https://www.datasciencecentral.com/artificial-neural-network-ann-in-machine-learning/.

Stephen A. Hillegeist, Elizabeth K. Keating, Donald P. Cram, Kyle G. Lundstedt. 2004. "Assessing the probability of bankruptcy." *Review of Accounting Studies.*

Szmalec, Izabela. 2023. *CEOpedia Management Online.* https://ceopedia.org/index.php/Altman_Z_score.

Taylor, Sebastian. 2023. *Corporate Finance Institute.* https://corporatefinanceinstitute.com/resources/data-science/decision-tree/.

Team, CFI. 2022. *Corporate Finance Institute.* https://corporatefinanceinstitute.com/resources/commercial-lending/financial-distress/.

—. 2023. *Corporate Finance Institute.* https://corporatefinanceinstitute.com/resources/commercial-lending/altmans-z-score-model/.

Zavgren, C. V. 1985. "Assessing the vulnerability to failure of American industrial firms." *Journal of Business and Accounting* 19-45.