Article

Minimalistic Logit Model as an Effective Tool for Predicting the Risk of Financial Distress in the Visegrad Group

Michal Pavlicko * and **Jaroslav Mazanec**

Department of Quantitative Methods and Economic Informatics, The Faculty of Operation and Economics of Transport and Communications, University of Zilina, Univerzitna 8215/1, 010 26 Zilina, Slovakia; jaroslav.mazanec@uniza.sk

* Correspondence: michal.pavlicko@uniza.sk

Abstract: Predicting financial distress is one of the most well-known issues in corporate finance. Investors and other stakeholders often use prediction models as relevant tools for identifying weaknesses to eliminate potential threats to business partners. This paper aims to present an effective logistic regression model for a one-year-ahead prediction of financial distress with the minimum set of predictors as a part of risk management. The paper is motivated by various works dealing with the curse of dimensionality phenomenon and the observation that the increasing number of logit-model predictors does not improve the prediction—on the contrary. Monitoring the significance of improvement in the stepwise growth of the predictor set is used to identify the minimal set. Logistic regression with cross-validation is involved in the modelling process. The proposed model is compared with other logit-based models used regionally or globally on the same large dataset, which underlines the model validity and robustness. The proposed logit model contains only two significant predictors and achieves excellent performance metrics compared to other models. The added value of the article lies in a simple application for managers, investors, creditors, financial institutions, and others with a reliable classification of companies into healthy and unhealthy company groups.

Keywords: financial distress; logit model; prediction model; Visegrad group; curse of dimensionality; risk management

MSC: 62J12; 62M20; 62P20; 91B30; 91B82

1. Introduction

Nowadays, financial distress prediction is a very important issue in corporate finance especially for academics and practitioners such as owners, managers, investors, banks, financial institutions and the state. Potential economic and non-economic losses resulting from the company's bankruptcy can be minimized using prediction models.

The main purpose of the article is to propose a universal tool for Central European companies from various sectors with the minimal possible predictor set and high prediction performance. The focus on the minimal predictor set is due to the effort to avoid the negative effect of the so-called "dimensional curse", which has been monitored in various studies and also during our previous work. We emphasize the easy application of the proposed model for a one-year-ahead prediction. Moreover, we hope that the presented model meets all criteria to achieve the goal.

The approach is based on the following methodological steps. First, we determine the significant variables to monitor the magnitude of performance increments in stepwise predictor set growth. Second, we create several generalized linear models linked by logit function and select the best model based on two major metrics. Third, we verify the proposed model on a test sample. The model is compared to other universal or regional prediction logit models. We believe that this methodology leads to relevant findings in

the field of risk management for Central European companies. In addition, we gather a wide range of theoretical and empirical experiences from previous scientific researchers with a focus on top authors such as Geise, Kuczmarska and Pawlowski (2021) [1], Brozyna, Mentel and Pisula (2016) [2], Balina and Juszczyk (2014) [3], Pisula (2012) [4], and Jakubík and Teplý (2008) [5], dealing with issues of estimating the likelihood of business failure using logistic regression.

The paper is divided into five sections. The literature review summarizes a wide of spectrum theoretical and empirical results about the prediction of financial distress from the previous research. The methodology explains a way to design a prediction model based on logistic regression, data preparation, and performance metrics used to assess prediction power. The results present the final estimation of proposed model parameters with the performance metrics achieved on the testing sample. The discussion includes a comparison of the proposed model with other universal or regional models for Central European enterprises. Finally, we summarize the key findings of the proposed model.

2. Literature Review

Financial prediction is one of the most important areas in the field of corporate finance to identify potential threats and opportunities for improving effective corporate governance. Bankruptcy has various consequences for owners, managers, employees, suppliers, and others. Bankruptcies of major companies can threaten everyday life at the regional or national level. In other words, bankruptcy has microeconomic as well as macroeconomic consequences such as increased unemployment and insolvency in mortgage loans. Indriyanti (2019) [6] argues that financial distress is insolvency for high fixed costs, illiquid assets, or economic fluctuations. Sun et al. (2014) [7] summarizes the definitions of financial distress based on previous studies. Financial distress is the inability to pay its liabilities or pay preferred dividends. On the other hand, previous research defines financial distress using proposed models with recommended values for classifying companies in terms of financial distress. However, some models are not able to divide businesses into prosperous and non-prosperous companies for the grey area. This disadvantage is typical of multivariable discriminant models compared to logistics models. Taffler (1983) [8] explains that a company with a TM index of less than 0.2 is bankrupt. Fulmer et al. (1984) [9] argue that a company with a negative H-factor is bankrupt. Altman et al. (1984) [10] claim that a company with a Z-index lower than 1.1 is bankrupt. Finally, Springate (1978) [11] identifies a bankrupt company based on the SM-index. If this index is less than 0.862, the company is in bankruptcy. On other hand, Kovacova and Kliestikova (2017) [12] explain that the Slovak legislation determines a bankrupt company based on five criteria as a negative profit after tax, financial independence indicator not exceeding 0.04, current liquidity not exceeding 1, the company has at least two overdue liabilities from two different creditors, the total amount of payable and not payable liabilities is higher than the value of a company's assets.

Smith and Winakor (1935) [13], FitzPatrick (1932) [14], Ramser and Foster (1931) [15], and Merwin (1942) [16] were pioneers in financial prediction in the first half of the 20th century. However, Altman (1968) [17] is one of the most important researchers in predicting a company's financial distress. On other hand, Beaver (1966) [18] applied univariate analysis in comparison with other models based on multivariate discriminant analysis (MDA). Ohlson (1980) [19] developed O-score using logistic regression analysis in this area. Zmijewski (1984a) [20] was the first author to use a probit model to solve the prediction of financial distress. Their model contains three variables, namely return on assets, debt ratio, and current ratio. Probit and logit analysis plays a major role in predicting financial distress. Logistics regression provides an easily reproducible model compared to more sophisticated and complex approaches such as support vector machines, neural networks, or ensemble models Pavlicko, Durica and Mazanec (2021) [21]. However, many prediction models are not suitable for estimating financial distress in each country around the world for different historical and economic aspects. Kovacova and Kliestikova (2017) [12] demonstrate that specific prediction models for selected integration groups are important in classifying

enterprises. Their results show that foreign models correctly identify only 50% of all Slovak companies. In addition, Liang, Tsai and Wu (2015) [22] explain that specific models are important for different countries or regions. In other words, these models consider various legal and economic aspects that affect financial stability and corporate performance.

In our case, Central European governments have transformed the state-run economy into a market-oriented economy. This transformation took place through the privatization of state-owned enterprises and the arrival of major investors from energy and telecommunications after the fall of the communist regime. In addition, many companies went bankrupt because their business activities were not competitive compared to other companies in Western Europe. During this period, the countries of Western Europe were part of an important integration group of the European Union with all the benefits such as free movement of goods, services, capital and labor. Other advantages included a common currency with the elimination of exchange rate losses and the Schengen area. These aspects have significantly contributed to the development of private enterprise and investment in Western Europe. The member states of the Visegrad Group have not achieved the economic performance of the countries even more than 25 years after the fall of the communist regime. Valaskova et al. (2018) [23] argue that each country provides different business conditions that have a significant impact on business performance. For this reason, their research presents a new prediction model for Slovak companies.

Czech Republic. Neumaierova and Neumaier (1995) [24] created the first prediction model for companies in the Czech Republic. Other models based on logit and decision trees analysis were modelled by Jakubík and Teplý (2011) [25] and Karas and Režňáková (2014) [26]. Jakubík and Teplý (2008) [5] contributed to progress in credit risk management. The prediction model was based on accounting data using binary logistic regression to evaluate financial stability in the corporate sector. The dataset includes 22 financial indicators divided into four groups, such as liquidity, solvency, profitability, and activity ratios. These data were obtained from Czech Capital Information Agency from 1993 to 2005. They demonstrated that interest coverage, gross profit margin, cash ratio, and return on equity have a positive impact on financial stability compared to leverage I, leverage II, and inventory ratio. The prediction model correctly classifies 80.41% of all companies, based on the Gini coefficient. Vochozka et al. [27] proposed a prediction model for the shipping sector using profitability, activity, liquidity, and debt ratios from Albertina database. The results demonstrate that the threshold value is 0.52 based on sensitivity analyses. In other words, the company will likely survive possible financial distress with a threshold value exceeding 0.52.

Slovak Republic. The prediction of financial distress is attractive for many researchers in the Slovak Republic. Chrastinová (1998) [28] and Gurčík (2012) [29] proposed a prediction model using MDA for the agricultural industry. CH-index and G-index models were inspired by Altman's Z-index model based on the set of variables. Hurtošová (2009) [30] created a logistic regression model including an assessment of the company's creditworthiness. Mihalovič (2016) [31] concentrated on the performance comparison of MDA and logit models for Slovak companies. The logit model achieved higher predictive performance in the training and testing sample. Jenčová et al. (2020) [32] forecasted failure distress in the electrical engineering industry based on accounting data from financial statements. The model includes five variables, namely, accounts payable turnover ratio, return on sales, quick ratio, financial leverage, and net working capital to assets. Moreover, return on sales, quick ratio, and net working capital to assets decrease the likelihood of bankruptcy. The model is good at evaluating classification quality based on ROC (more than 95%). Kovacova and Kliestik (2017) [33] developed models for bankruptcy prediction of Slovak companies using logit and probit methods. On the other hand, Štefko et al. (2020) [34] applied data envelopment analysis (DEA) to forecast financial distress in the heating industry. They emphasize that DEA is not an effective method as logistic regression is for weak performance in determining failed and non-failed companies.

Poland. Pisula (2012) [4] implemented an effective tool for estimating the probability of failure distress based on 225 Polish companies from the logistics sector using the logit model in Statistica Visual Basic. Similarly, Balina and Juszczyk (2014) [3] modelled financial distress in the transport sector. They compare the accuracy of selected foreign discriminant models using Altman II, Altman III, Springate, Legault, and Fredrikslust I. On the other hand, Lozinskaia et al. (2017) [35] identified key aspects of financial distress in the shipping sector using the logit model. They found that significant variables are Tobin Q, EBITDA, GDP, and logarithms of total assets. The results demonstrate that the model correctly predicts less than 70% of all in the testing sub-sample. In a wide context, research shows that financial and macroeconomic variables are significant in assuming financial distress. Brozyna et al. (2016) [2] applied linear discriminant analysis, logistic regression, classification trees, and k-nearest neighbors. They divide methods into statistical and non-statistical models based on the operational research methods and artificial intelligence methods. The statistical methods represent a discriminant analysis, linear regression, logistics regression, probit regression, classification tree, and k-nearest neighbors. On the other hand, non-statistical methods are mathematical programming, neural networks, fuzzy set theory, rough set theory, support vector machine, genetic algorithms, and experts systems. They collected data of companies in the logistics sector from Corporate Database EMIS (Emerging Markets Information Service). The dataset consists of 28 financial variables divided into liquidity, profitability, debt, operating effectiveness, and other variables on capital structure. They found out that ROC indicates more than 90% of all in each sub-samples expect from testing the sample of classification and regression tree in the one-year horizon. However, ROC shows that the network model is the best model than others in the two-year horizon. Berent et al. (2017) [36] summarized a wide spectrum of theoretical and empirical knowledge to identify key approaches in forecasting failure distress, such as multivariate discriminant analysis, logistic regression, k-nearest neighbors, classification trees, support vector machines, and neural networks. They used data on less than 15,000 Polish non-listed companies from Coface Poland Credit Management Services from 2006 to 2015 to model financial distress using the alternative tool − double stochastic Poisson process with the multi-period horizon in Matlab compared to traditional approaches. They applied various macroeconomic and market data from the Central Statistical Office of Poland. The financial ratios are divided into five groups, such as liquidity, profitability, rotation, leverage, and size. Moreover, the dataset includes macroeconomic and market variables such as nominal GDP, gross investments, and export growth. The model achieves an 81% of accuracy ratio two years ahead of default. Geise et al. (2021) [1] identified seven of 17 significant variables, namely, current assets, turnover, debt to assets ratio, operating profit to assets, gross profit to assets, operating profit plus amortization to short-term liabilities, current assets to assets ratio, and equity to assets ratio to estimate the likelihood of bankruptcy using a logit model for the construction sector in Poland. Moreover, the ROC results demonstrate that the model achieves high classification quality at the level of 99%. Finally, findings show current assets to total assets harm financial stability. On other hand, profitability decreases bankruptcy likelihood. Noga and Adamowicz (2021) [37] assumed failure distress using multiple discriminant analyses in the wood sector based on financial statements from Code District Courts and credit information bureaus. The model shows that current assets to current liabilities have the greatest impact on corporate bankruptcy compared to other variables.

Hungary. The first models to predict the bankruptcy in Hungarian companies were introduced in Hajdu and Virág (2001) [38] and Virág and Kristóf (2005) [39]. Virág and Kristóf (2005) [39] proposed a prediction model using an artificial neural network, but also traditional approaches such as discriminant analysis and logistic regression. The results indicate that failed and non-failed companies differ from each other based on quick liquidity ratio, return on sales, cash-flow to total debts, current assets to total assets, and accounts receivable to accounts payable using logistic regression. Moreover, discriminant analysis lies on quick liquidity ratio, cash-flow to total debts ratio, current assets to total

assets ratio, and cash-flow to total assets ratio. However, they claim that the logit model is better than discriminant analysis based on classification accuracy.

Other countries. Laitinen and Suvas (2013) [40] analyzed financial distress overall in 30 European countries. Moreover, they estimate financial distress using logistic regression and compare accuracy across countries. Grünberg and Lukason (2014) [41] modelled financial distress using logistic regression and neural networks in Estonia. The dataset includes data on 13 variables divided into solvency, capital structure, profitability, liquidity, efficiency, size, and others. One of the variables is the not financial indicator, namely, firm age at the time of bankruptcy. The results identify that equity to total assets, current assets to total assets, and natural logarithms of total sales are significant variables in logistic regression to distributing between failed and non-failed companies. Moreover, if the equity to total assets ratio increase, so bankruptcy probability decrease compared to other variables. Finally, the logit model is better to classify correctly non-failed companies. The neural network comprises cash to current liabilities, total liabilities to total assets, net profit to total assets, retained earnings to total assets, cash to total assets. They found out that accuracy for healthy companies is lower in the neural network than in logistic regression.

Finally, we summarize the significant predictors from logit models for determining financial distress based on Virág and Kristóf (2005) [39], Hurtošová (2009) [30], Jakubík and Teplý (2011) [25], Pisula (2012) [4], Delina and Packová (2013) [42], Balina and Juszczyk (2014) [3], Grünberg and Lukason (2014) [41], Harumova and Janisova (2014) [43], Vochodzka et al. (2015) [27], Brozyna et al. (2016) [2], Gulka (2016) [44], Kovacova and Kliestik (2017) [33], Lozinskaia (2017) [35], Durica et al. (2019) [45], Jenčová et al. (2020) [32], and Geise et al. (2021) [1]. Table 1 shows the 11 most common statistically significant indicators, based on bibliographic analysis, which were used in at least two models.

Table 1. The summary of significant variables from the logit model mainly for Central Europe.

Financial Variables (Expressed by Formula)	Total Number	Authors
current assets/ current liabilities	5	Pisula (2012) [4], Harumova and Janisova (2014) [43], Brozyna et al. (2016) [2], Kovacova and Kliestik (2017) [33], Durica et al. (2019) [45]
equity/total assets	5	Hurtošová (2009) [30], Grünberg and Lukason (2014) [41], Gulka (2016) [44], Kovacova and Kliestik (2017) [33], Geise et al. (2021) [1]
total debt/total assets	4	Pisula (2012) [4], Kovacova and Kliestik (2017) [33], Durica et al. (2019) [45], Geise et al. (2021) [1]
current assets/total assets	4	Virág and Kristóf (2005) [39], Balina and Juszczyk (2014) [3], Grünberg and Lukason (2014) [41], Geise et al. (2021) [1]
(current assets-inventory)/ current liabilities	2	Virág and Kristóf (2005) [39], Jenčová et al. (2020) [32]
cash and cash equivalents/ short-term liabilities	2	Brozyna et al. (2016) [2], Vochodzka et al. (2015) [27]
sales/total assets	2	Harumova and Janisova (2014) [43], Durica et al. (2019) [45]
(inventory/sales)*360	2	Hurtošová (2009) [30], Jakubík and Teplý (2011) [25]
total debt/equity	2	Jakubík and Teplý (2011) [25], Balina and Juszczyk (2014) [3]
cash flow/total debt	2	Virág and Kristóf (2005) [39], Delina and Packová (2013) [42]
EBITDA/sales	2	Harumova and Janisova (2014) [43], Jenčová et al. (2020) [32]

Similarly, Kováčová et al. (2019) [46] analyzed the methods and financial indicators used in the Visegrad Group. The results show that discriminant analysis and conditional probability are more often used than a neural network, decision trees, and data envelopment analysis. Moreover, the prediction models are based primarily on liquidity ratios (current ratio, quick ratio, working capital/total assets, cash ratio), debt ratios (liabilities/total assets, equity/total assets, cash flow/liabilities), profitability ratios (ROA, ROE, EBIT/total assets,

operating profit/total assets) and activity ratios (total revenues/total assets, total sales/total assets, cash-flow/total assets). Moreover, Prusak (2018) [47] provides an overview of techniques used in national models from Central Europe, Eastern Europe, and Baltic countries from 2016 to 2017.

3. Methodology

Data. This paper aims to present a prediction model to estimate business failure in the Slovak Republic, the Czech Republic, Poland, and Hungary as members of the Visegrad Group (V4). The Amadeus database collected 27 financial variables in Central Europe from 2016 to 2018. Table 2 shows the activity ratio, liquidity ratio, profitability ratio, and debt ratio. These ratios and their indexing are the same as in [48], for easier validation and comparison. The last-mentioned research contains a more detailed view of the most used financial variables in financial distress prediction.

Table 2. Economic indicators.

№	ID	Type	Financial Variable	Formula
1	X01	activity	asset turnover ratio	sales/total assets
2	X16	activity	current assets to sales ratio	current assets/sales
3	X18	activity	inventory to sales ratio	inventories/sales
4	X32	activity	net assets turnover ratio	net sales/total assets
5	X38	activity	total liabilities to sales ratio	total liabilities/sales
6	X06	leverage	debt to EBITDA ratio	total liabilities/EBITDA
7	X10	leverage	debt ratio	total liabilities/total assets
8	X11	leverage	current assets to total assets ratio	current assets/total assets
9	X14	leverage	solvency ratio	cash flow/total liabilities
10	X15	leverage	short-term debt ratio	current liabilities/total assets
11	X21	leverage	long-term debt ratio	non-current liabilities/total assets
12	X02	liquidity	current ratio	current assets/current liabilities
13	X12	liquidity	cash to total assets ratio	cash and cash equivalents/total assets
14	X22	liquidity	cash ratio	cash and cash equivalents/current liabilities
15	X23	liquidity	operating cash flow ratio	cash flow/current liabilities
16	X26	liquidity	quick ratio	(current assets—stock)/current liabilities
17	X36	liquidity	net working capital	current assets—current liabilities
18	X04	profitability	ROE	net income/shareholder's equity
19	X05	profitability	EBITDA margin	EBITDA/sales
20	X07	profitability	ROA	net income/total assets
21	X09	profitability	ROTA	EBIT/total assets
22	X13	profitability	Cash ROA	cash flow/total assets
23	X19	profitability	free cash flow to sales ratio	cash flow/sales
24	X20	profitability	net profit margin	net income/sales
25	X28	profitability	ROE (of EBIT)	EBIT/shareholder's equity
26	X31	profitability	cash flow to operating revenue ratio	cash flow/EBIT
27	X35	profitability	EBIT margin	EBIT/sales

Note: return on equity (ROE), return on assets (ROA), earnings after taxes (EAT), earnings before interest and taxes (EBIT), earnings before interest, taxes, depreciation, and amortization (EBITDA).

Table 3 demonstrates that approximately 16.5% of all companies in the Visegrad Group are unhealthy in the monitored period, but we can observe the differences in corporate bankruptcy among the countries. In Slovakia and the Czech Republic, about 21% are unhealthy companies. On the other hand, only 13% of all companies are in Poland and Hungary. Despite these differences, the model parameters are set for the whole group without any knowledge of country affiliation. However, we examine the performance of the model both for the Visegrad group as a whole and for each country in the group.

Table 3. Total sample.

| Country | The Year 2017 | | | | The Year 2018 | | | |
| | Number | | % | | Number | | % | |
	Non-Failed	Failed	Non-Failed	Failed	Non-Failed	Failed	Non-Failed	Failed
SK	122,946	32,178	79.26	20.74	122,846	32,278	79.19	20.81
CZ	76,634	20,845	78.62	21.38	76,633	20,846	78.61	21.39
PL	59,780	8 487	87.57	12.43	59,579	8 688	87.27	12.73
HU	298,713	47,999	86.16	13.84	299,189	47,523	86.29	13.71
Total	558,073	109,509	83.60	16.40	558,247	109,335	83.62	16.38%

Note: Slovak Republic (SK), Czech Republic (CZ), Poland (PL), Hungary (HU).

Data division for training, validation, and testing. The final setting of parameters of the proposed model originated from financial variables in 2016 and the corresponding financial distress statement in 2017. These data formed training and validation datasets. The proposed model was tested on financial variables in 2017 to predict business failure in 2018 for all companies—testing sample. We emphasize that the testing sample was not standardized at all, and the model was tested on a whole testing sample consisting of all companies.

Training and validation data standardization process. All records with substandard or missing values in columns of a respective variable were dropped out. As a substandard value was considered to be an entry where the value of a variable was not in a close interval around the variable mean. The close interval was determined as 50% of the standard deviation around the respective mean. The motivation for such an approach lies in the presumption that it is easier to distinguish among records with substandard values and it is more difficult to classify records that are close to the respective variable mean. We assume that model coefficients trained on such a sample would respectably differentiate between failing and non-failing companies with standard as well as substandard variable values.

Creating of k-folds with the balanced samples for training and validation. We extracted all standardized records of the failing companies and then supplemented them by a random selection of the non-failing companies of the same size to create a balanced sample fold for training and validation. To create *k*-folds for each combination of variables, the records of failing companies were the same, but the same-sized selection of non-failing companies differs among the folds. Each fold was randomly divided into training and validation subsamples in the ratio of 80 to 20.

The performance metrics for evaluation of the logit model. There are various metrics to validate the model prediction performance, but the informative value may be misleading in certain circumstances. This can be caused by an imbalance between positive and negative responses in the dataset or by a threshold setting of a model function. Hence, the two major metrics in our research are the Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) and Matthews Correlation Coefficient (MCC). AUC is the most common metric in the field, and it visualizes a trade-off between model sensitivity and specificity. The metric regards all possible thresholds (cut-off); therefore, it is resistant to any misleading due to the current setting. However, this metric is not resistant to an imbalanced dataset as shown in various research [49]. AUC that reaches more than 0.9 is usually considered an excellent result and lower than 0.6 as an insufficient result.

Other performance metrics are derived from the confusion matrix that can be constructed after the final decision about the definitive threshold setting of a model function. The confusion matrix (CM) is 2-by-2 sized and stores numbers of true positive (TP), true negative (TN), false positive (FP), and false-negative (FN) predictions. Subsequently, all derived metrics are affected by the threshold value. On the other hand, the MCC metric is fully resistant to an imbalance in the dataset [50] shown in their research. Therefore, the MCC (Equation (1)) is the second major metric applied in the validation process.

$$\text{MCC} = \frac{\text{TP} \cdot \text{TN} - \text{FP} \cdot \text{FN}}{\sqrt{(\text{TP} + \text{FP}) \cdot (\text{TP} + \text{FN}) \cdot (\text{TN} + \text{FP}) \cdot (\text{TN} + \text{FN})}}. \tag{1}$$

Another popular metric derived from the confusion matrix is the F_β Score, Equation (2). This metric is not completely resistant to an imbalanced dataset because it ignores true negative outcomes as shown in [51]. The advantage of this metric is the possibility to choose the coefficient β reflecting a ratio between the Sensitivity (also known as True Positive Ratio—TPR or as Recall, it is the ratio of TP to all positive entities, i.e., TP + FN) and Precision (ratio of TP predictions to all positive predictions, i.e., TP + FP).

$$F_\beta\text{Score} = \frac{(1 + \beta^2) \cdot \text{TP}}{(1 + \beta^2) \cdot \text{TP} + \beta^2 \cdot \text{FN} + \text{FP}}. \tag{2}$$

If it is not desired to distinguish between sensitivity and precision importance, the β coefficient is set to 1 and F_1 Score, i.e., the harmonic mean can be calculated according to Equation (3).

$$F_1\text{Score} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FN} + \text{FP}}. \tag{3}$$

While the F_1 Score is not considered the major metric in our validation process, the scores are shown in our result tables altogether with the Accuracy—ACC (ratio of truly identified entities, i.e., TP + FN to the total population, i.e., TP + TN + FP + FN), Sensitivity—TPR, and Specificity—TNR (also known as True Negative Rate, it is the ratio of TN to all negative entities, i.e., TN + FP).

Logit model. Many statistical methods can be used to create a predictive scoring model. As described in the literature review, many current models are based on artificial neural networks, support vector machines, ensemble models, and so on. These models achieve excellent results, but reproducibility is questionable and difficult. In other words, the less complicated models such as decision trees or logistic regression models are still very popular.

Logistic regression (logistic model or logit model) is used to determine the probability p of the categorical binary dependent variable (Y) such as a prosperity statement (healthy company: Y = 0, unhealthy company: Y = 1). It comes from linear regression that can be described as a linear combination w of n explanatory—independent variables, i.e., predictors X = $(x_1, x_2, \ldots, x_n)$ with the corresponding regression coefficients β_i and the intercept β_0, Equation (4).

$$w(\text{X}) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \quad \text{for } i = 1, 2, \ldots, n. \tag{4}$$

It is assumed that the sum of explanatory variables multiplied by the relevant coefficients is linearly related to the natural logarithm (referred to as the logit) describing the rate of business failure, i.e., financial distress of a company. The logit is linear in its parameters, may be continuous, and may range from negative to positive infinity. The logit model usually normalizes scoring function into the interval from 0 to 1 through a logistic transformation, known also as inverse logit transformation or exponential transformation, which describes the probability of business failure [25,52], Equation (5).

$$p = \text{P}(\text{Y} = 1 | \text{X}) = \frac{1}{1 + e^{-w(\text{X})}}. \tag{5}$$

Methodology background. In our previous research [21], an interesting behavior was found in the k-nearest neighbor method. The performance of this method was decreased for the phenomenon called the "curse of dimensionality" with the rising number of predictors. This obstacle is caused by high dimensional data, where the points drawn from a probability distribution tend to never be close together [53,54]. Similar behavior could be found in many types of research dealing with non-parametric regression and fixed-effect logit models [55–59]. In the discussion section of our previous research, we dealt with various logit and probit models from the region. The models with fewer sets of predictors achieved better results than the model with a higher number of predictors. This observation was an impulse for this research.

Research goals. The research implies two goals. First, to find out if the addition of more variables would lead to a relevant improvement in the validation sample and if it reflects in the testing sample. Second, to create an extremely simple logit model with the lowest number of predictors that is easy to reproduce and remember and compare it with similar and larger models used in this sphere to prove the relevance of our observations.

Methodology for proving the relevancy of the minimal predictor set. To find out whether the increasing number of variables in the predictor set of a logit model brings a decisive or negligible improvement or even deterioration. We chose the method of stepwise growth of a model predictor set. This method is based on choosing the best one-predictor model, then adding another predictor from a set of all variables to create the best two-predictor model, and so forth. The parameters of the logit models are estimated in each step and for each fold throughout the process.

This procedure is based on 10 folds. The average AUC value (of 10 validation sub-samples) represents the determination criterion for the best set of variables in each additional step. The procedure of stepwise addition is set up to finish when the set reaches the size of the eight predictors. Then the significance of prediction improvement measured via the ROC metric on the validation sub-samples is assessed, and when the improvement is less than 0.5% in the subsequent addition steps, then the added predictors are pruned from the set. The pruned set of predictors is the final set transferred to the following methodology step dealing with the tuning of the parameter estimation. This approach differs from the conventional research approach, where the variables are determined as statistically significant via a selected statistical test. The proposed methodology approach examines the improvement significance through the optics of the predictive performance achieved on the validation sub-sample, as is usual in machine learning.

Methodology for finding the parameters of the best minimal logit model. To find out the best minimal logit model according to both main metrics (AUC and MCC) we decided to create 50 folds for this procedure. It means that the training and validation dataset is divided into 50 folds of the training and validation subsamples. Hence, 50 generalized linear models (GLM), linked by the logit function, are created (each GLM with its estimated parameters) and validated on 50 corresponding validation samples. Then, the models are sorted according to the rank determined by the AUC rank and MCC rank achieved on the validation subsample where both metric ranks have the same weight. This way, both main metrics are represented in the final coefficient selection.

The first ranked model is selected, and its intercept and coefficient values are simplified. We emphasize that the goal was to create the best minimal logit model that is easy to reproduce and easy to remember. Therefore, simplification consists of multiplying the exponent in the logit model by such a coefficient that, as a result, the smallest member of the exponent is equal to one. As is known, such an adjustment affects only the steepness of the logit function, as shown in Figure 1, and does not affect the decision on a positive or negative outcome, i.e., the company financial statement.
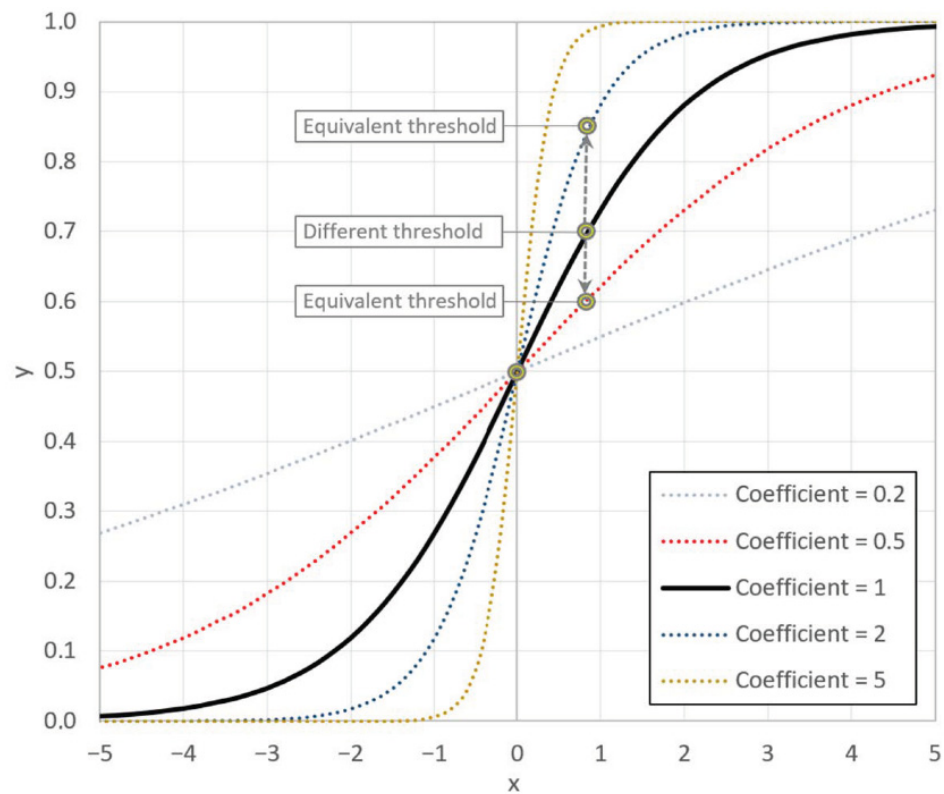
Figure 1. The coefficient value effect on the function slope.

As is clear from the figure, the proposed simplification gives the same result only at the threshold of 0.5. However, the corresponding threshold to different threshold value may be found as depicted in the figure. The method to calculate the required cut-off can be extracted from the ROC plot by moving the line with a specific slope from the upper-left corner to the left-right corner of the plot. When the line first intersects the ROC curve, the required cut-off is found. The slope of the line can be adjusted according to the cost preferences of the positive or negative misclassification.

To better understand the individual steps of the chosen methodology and dataset division, we provide the conceptual flowchart shown in Figure 2.
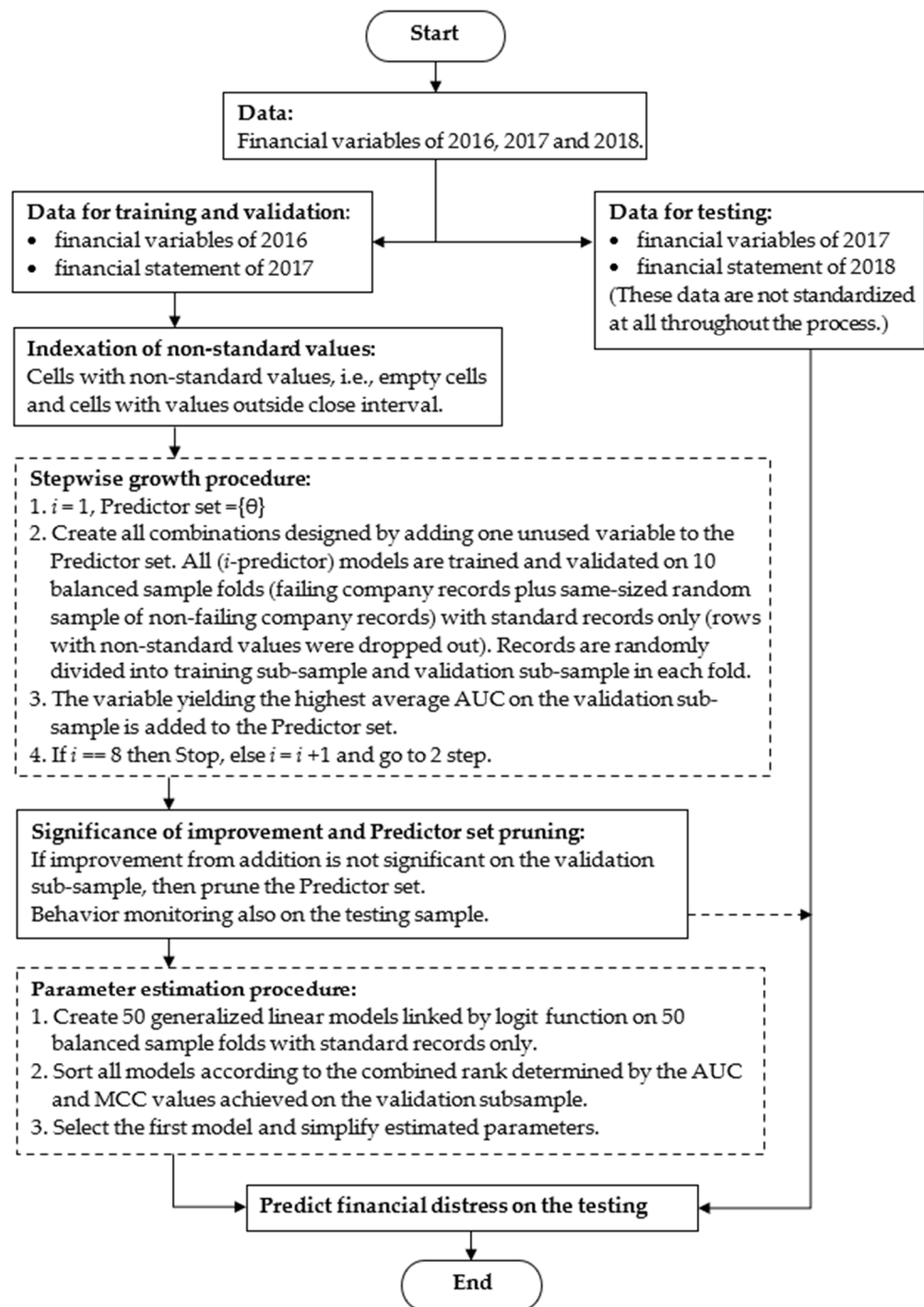
Figure 2. Methodology flowchart.

4. Results

This section is divided into two parts. The first part involves results supporting the aim of creating the model with minimal possible predictors achieving excellent prediction metrics. The second part involves the final estimation of the parameters of the minimalistic logit model altogether with the one-year-ahead prediction of financial distress for the whole Visegrad group as well as for each country separately.

Minimal predictor set. As described in the methodology section, we want to find out if the increasing number of variables in the predictor set of a logit model brings a decisive or negligible improvement or even deterioration. Figure 3 shows the development of the major performance metrics values in each step of the growth of the predictor set on the

validation sub-sample (dotted line). Moreover, we check the behavior on the testing sample (solid line) to monitor the progress of the performance metrics values.
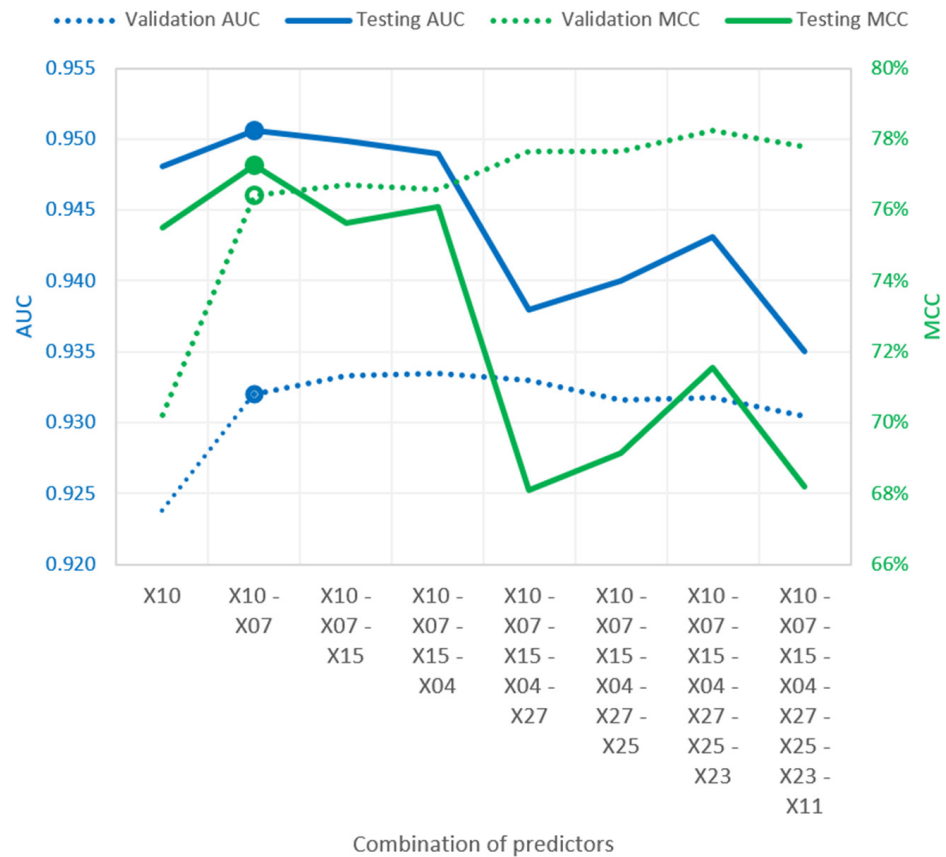

Figure 3. Average performance metrics in the different combinations of predictors on the validation and testing samples.

Table 4 shows the precise numbers of discarded records that did not pass the standardization process at each step. Moreover, the table shows the total number of records in each balanced fold, as well as the number of records in the training and validation sub-samples.

Table 4. Stepwise growth and record discarding in each step of the standardization process.

Model Composition	Non-Failed			Failed			Size of Balanced Sample in the Fold		
	Pass	Discard	Discard [%]	Pass	Discard	Discard [%]	Total	Training Sub-Sample	Validation Sub-Sample
X10	551,191	6882	1.23	107,577	1932	1.76	215,154	172,123	43,031
X10–X07	530,645	27,428	4.91	103,307	6202	5.66	206,614	165,291	41,323
X10 … X15	530,627	27,446	4.92	103,006	6503	5.94	206,012	164,810	41,202
X10 … X04	528,027	30,046	5.38	101,336	8173	7.46	202,672	162,138	40,534
X10 … X27	436,550	121,523	21.78	52,597	56,912	51.97	105,194	84,155	21,039
X10 … X25	352,144	205,929	36.90	48,531	60,978	55.68	97,062	77,650	19,412
X10 … X23	268,616	289,457	51.87	26,920	82,589	75.42	53,840	43,072	10,768
X10 … X11	253,060	305,013	54.65	23,388	86,121	78.64	46,776	37,421	9355

The best one-predictor model is based on the debt ratio (X10). All metrics show very satisfying results (AUC: 0.9236, MCC: 0.7021, F_1 Score: 0.8126, ACC: 0.8380, TPR: 0.7032, TNR: 0.9721). Note, all the values are averages obtained from the validation sub-samples. The runner-up is the short-term debt ratio (X15) one-predictor model but the margin between them is quite significant in all metrics. The second step, i.e., the two-predictor model, brings a noticeable improvement in all metrics but the margin between the best combination and the runner-up is very subtle. The winning combination involves X10

and X07 (ROA) predictors with the AUC value very close (it was equal to fourth decimal place) to the runner-up, the combination of X10 and X09 (ROTA). However, the winning combination achieves higher values in all metrics except for specificity.

The following additions to the predictor set do not bring a noticeable improvement. While the MCC metric is very subtly improved in the following steps of predictor addition on the validation sample, the AUC metric is stable with a slight decrease. Moreover, we also observe the performance metrics of the models on the testing sample. It turns out that the addition of subsequent predictors to the model would harm the model instead of an improvement.

The stagnation on the validation sample and even decline on the testing sample reveal that the combination of debt ratio (X10) and rentability of assets ratio (X07) are the right predictors for the minimalistic logit model.

Estimation of the parameters. All 50 generalized linear models linked by the logit function are sorted according to the combined ranking of AUC rank and MCC rank with equal weights achieved on its corresponding validation sub-sample. The results of the TOP 10 models can be seen in Table 5. Each GLM has its estimated coefficients as can be seen in the table.

Table 5. TOP 10 GLM models sorted by the combined ranking of AUC rank and MCC rank with equal weights and other parameters.

GLM No.	ACC	F_1 Score	MCC	AUC	β_0	β_1	β_2	AUC Rank	MCC Rank	Total Rank
4.	0.8947	0.8871	0.7966	0.9478	−1.7661	−0.5523	1.7696	2	1	1
24.	0.8825	0.8720	0.7754	0.9482	−1.6368	−0.5797	1.6073	1	6	2
48.	0.8828	0.8725	0.7757	0.9468	−1.6535	−0.5734	1.6269	10	5	3
22.	0.8777	0.8656	0.7681	0.9468	−1.5586	−0.5802	1.5052	7	10	4
37.	0.8869	0.8780	0.7823	0.9461	−1.7207	−0.5044	1.7184	17	4	5
40.	0.8900	0.8822	0.7871	0.9459	−1.7719	−0.5747	1.7792	20	3	6
11.	0.8743	0.8615	0.7617	0.9468	−1.5910	−0.5597	1.5420	11	12	7
7.	0.8707	0.8569	0.7554	0.9468	−1.5255	−0.5993	1.4640	8	16	8
17.	0.8738	0.8609	0.7609	0.9468	−1.5368	−0.6397	1.4771	12	14	9
45.	0.8659	0.8503	0.7483	0.9469	−1.4876	−0.5928	1.4108	5	24	10
...	...	...	...	...	...	...	...	...	...	...

The descriptive statistics including average, maximal, and minimal values of estimated coefficients and chosen tests of their statistical significance of all 50 GLMs are summarized in Table 6. Statistical significance expressed by *p*-value is recognizably lower than 0.001 in all instances. The table includes the standard error of coefficient estimate and *t*-statistic values.

Table 6. Descriptive statistical summary of the estimated coefficients and their statistical significance of all 50 GLMs.

Financial Variable	Type	ID	Coefficient	Type of Statistic	Estimate	Standard Error	*t*-Statistic	*p*-Value
Intercept	-	-	β_0	Average	−1.554	0.012	−127.509	0.000
				Maximum	−1.080	0.014	−105.480	0.000
				Minimum	−1.900	0.010	−139.950	0.000
ROA	profitability	X07	β_1	Average	−0.574	0.016	−34.910	<0.001
				Maximum	−0.469	0.018	−28.916	<0.001
				Minimum	−0.653	0.015	−41.890	<0.001
Debt ratio	leverage	X10	β_2	Average	1.500	0.012	120.313	0.000
				Maximum	1.943	0.015	132.870	0.000
				Minimum	0.904	0.009	99.748	0.000

The best minimalistic model resulting from our methodology in the form of a linear combination includes rentability on assets (X07) and the debt ratio (X10) with the following estimation of the coefficients, Equation (6).

$$w = -1.7661 - 0.5523 \cdot \text{X07} + 1.7696 \cdot \text{X10}. \tag{6}$$

The coefficient values of the proposed simplified version of the linear combination are given by Equation (7).

$$\text{PM} = -3.2 - \text{X07} + 3.2 \cdot \text{X10}. \tag{7}$$

The model score (probability of business failure) can be calculated by inserting the expression into the exponent coefficient of the reverse logistic transformation formula, Equation (8).

$$p = \text{P}(Y = 1|X) = \frac{1}{1 + e^{-\text{PM}}}. \tag{8}$$

Model prediction. The proposed model was tested on the testing sample to predict the financial distress of the Visegrad group companies. A standard threshold of 0.5 was used to distinguish between healthy and unhealthy companies (see Figure 4). On the left-hand side, there is a confusion chart altogether with relative (normalized) row and column values for the entire Visegrad group. Confusion charts for each country are on the right side of the figure.
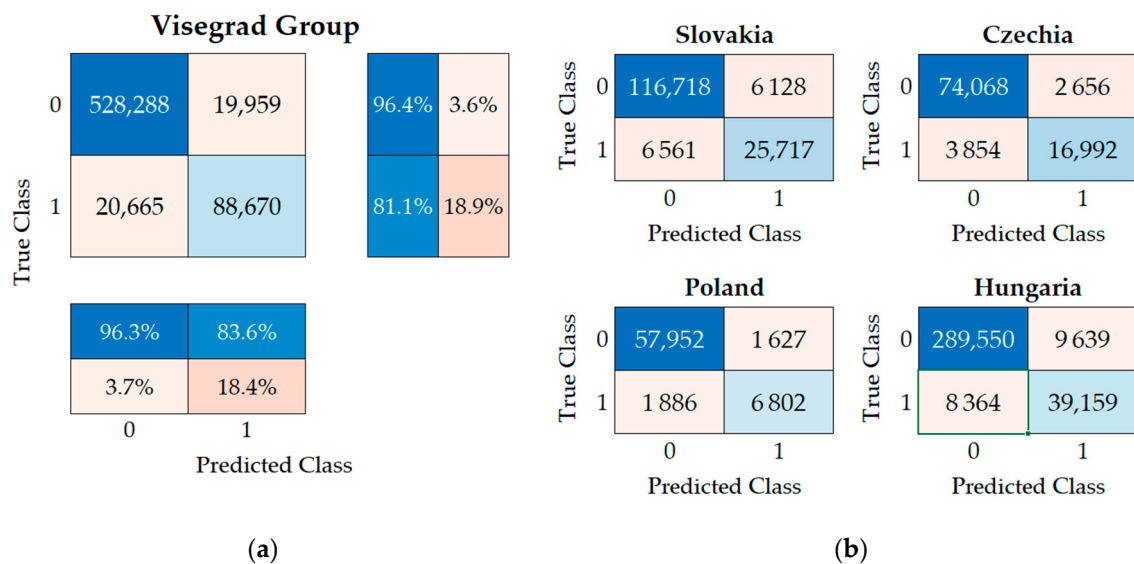

Figure 4. Confusion charts of model prediction on the testing sample. (**a**) Confusion chart of the Visegrad group with normalized row and column values, (**b**) confusion chart of each country separately.

Performance metrics derived from the confusion matrix are summarized in Table 7.

Table 7. Performance metrics of the proposed model on the testing sample.

Model	Accuracy	Sensitivity	Specificity	F_1 Score	MCC	AUC
Slovakia	0.9182	0.7967	0.9501	0.8021	0.7506	0.9434
Czechia	0.9341	0.8151	0.9665	0.8411	0.8003	0.9477
Poland	0.9485	0.7829	0.9727	0.7948	0.7655	0.9509
Hungary	0.9481	0.8240	0.9678	0.8131	0.7830	0.9542
Visegrad group	0.9391	0.8110	0.9642	0.8136	0.7773	0.9507

The shape of the ROC curves for the Visegrad Group as a whole and for each country in the group are depicted in Figure 5, where the whole area of the ROC curve is on the left-hand side and the detail of the upper left corner is on the right side of the figure.
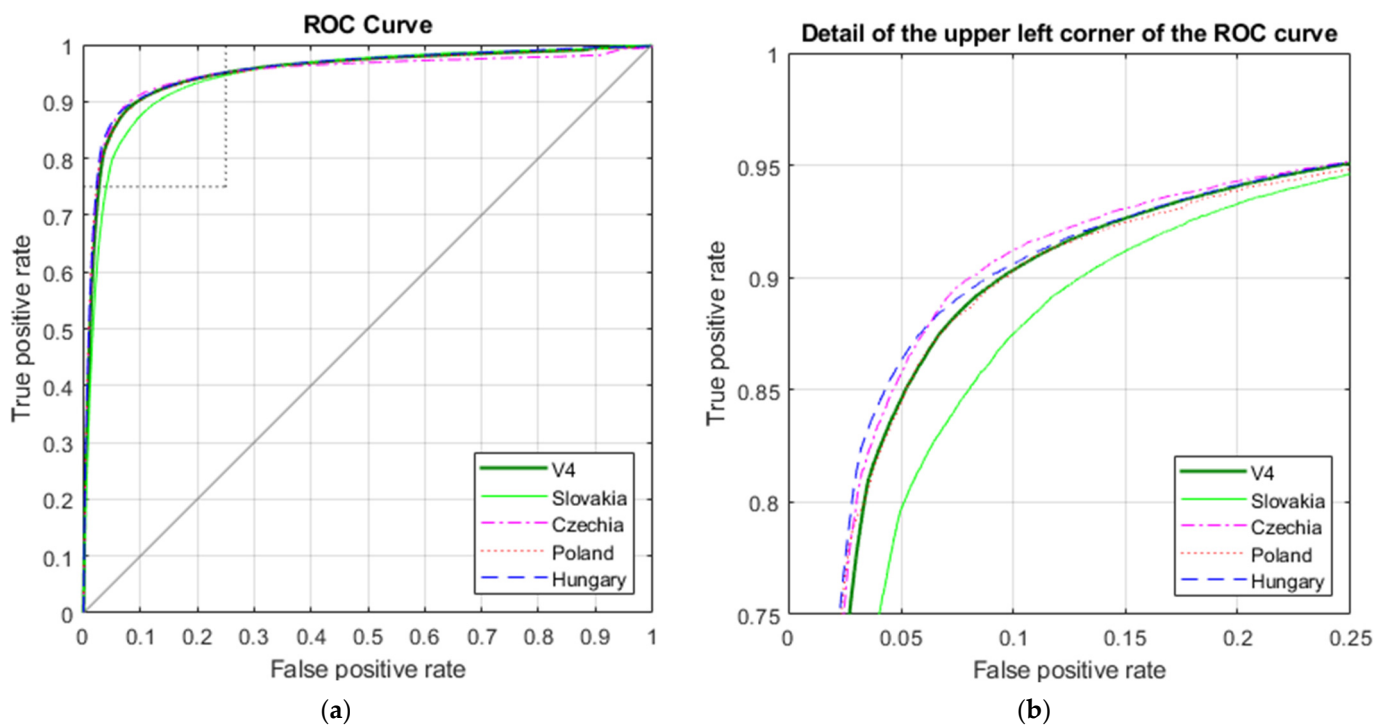
Figure 5. ROC curves of the Visegrad Group and its countries. (**a**) The shape of ROC curves on the whole interval, (**b**) the detail of the ROC curves.

All performance metrics achieve very good results on the testing sample. The model achieves very similar outputs for each country separately despite the optimization on the Visegrad group. We assume that the model may be used in any country without any restrictions. Moreover, the model proves a certain level of robustness.

5. Discussion

The primary aim of the research is to propose a simple and effective minimalistic logit model to predict business failure in various sectors in the Visegrad Group counties. The concentration on V4 countries lies in the specificity of this region, where the economy was transformed from a regulated to a free market in the recent past. However, the testing of the prediction power of the proposed model on the companies from outside this region would be beneficial, we have no appropriate dataset as mentioned in the limitation. The proposed model represents a useful tool for management that helps to make decisions to improve corporate performance before bankruptcy or can serve as an initial solution for other more sophisticated approaches. On behalf of the great size of the used dataset with more than half a million companies and achieved performance metrics values of very good quality such as an AUC of more than 0.95, MCC 78%, F_1 Score 81%, and an overall accuracy of 94%, we believe that the proposed model gives excellent results especially regarding minimalistic predictor set.

We verify and compare the prediction performance with other models on the same testing sample, in other words, the prediction of financial distress in 2018 based on financial variables in 2017. The comparison models were chosen based on three assumptions. First, the predictors of the chosen models had to match the same financial variables as our dataset. Therefore, we could not use the Altman model and its derivatives due to the absence of the market value index. Second, we preferred models used in the Visegrad Group countries or worldwide. Third, we found such models that differ in the number of predictors from small (3–5) to large (13–17) predictor sets. The last assumption has a secondary use to prove or disapprove our observations regarding the relationship between the number of predictors and the model performance (the potential occurrence of the curse of dimensionality). The

first group of selected models with a small predictor set is made of models from the research in [20,60,61] and the second group contains models from [48,62]. The compared models use the same formulas with the same parameter values as used in the relevant research and as shown in Table 8. In other words, the parameters were not re-estimated.

Table 8. The comparison of prediction models applied in V4 countries.

Authors	Year	Country	Formula for Exponent	Variables
Zmijewski *marked:* **Zmijewski**	1984	USA	$Zm = -4.336 - 4.513X07 + 5.679X10 - 0.004X02$	net income/total assets (X07), total liabilities/total assets (X10), current assets/current liabilities (X02)
Durica, Valaskova and Janoskova *marked:* **Durica**	2019	V4	$Z = 0.107 - 0.138CZ - 0.877HU - 0.599PL + 1.180SS - 0.863LS + 0.030X01 - 3.089 \cdot (X11 - X15) + 0.025X10 - 0.002X12 + 0.042X18 + 0.026X21 - 0.017X02 - 0.091X26 - 1.057X09 - 0.966X28 - 1.328X35 - 0.001X36$	sales/total assets (X01), current assets/total assets (X11), current liabilities/total assets (X15), total liabilities/total assets (X10), cash and cash equivalents/total assets (X12), inventories/sales (X18), non-current liabilities/total assets (X21), current assets/current liabilities (X02), (current assets—stock)/current liabilities (X26), EBIT/total assets (X09), EBIT/shareholder's equity (X28), EBIT/sales (X35), current assets—current liabilities (X36), dummy variables: Czech Rep. (CZ), Hungary (HU), Poland (PL), Small size company (SS), Large or very large company (LS)
Kliestik, Vrbka and Rowland *marked:* **Kliestik**	2018	V4	$y_{V4} = -1.470 + 0.024X02 - 0.589X04 - 1.158X07 + 1.870X10 - 0.452X11 + 0.613X12 + 1.030X15 - 0.012X22 + 0.731X09 + 0.173X28 - 0.475X35 + 0.244CZ + 0.522SK$	current assets/current liabilities (X02), net income/shareholder´s equity (X04), net income/total assets (X07), total liabilities/total assets (X10), current assets/total assets (X11), cash and cash equivalents/total assets (X12), current liabilities/total assets (X15), cash and cash equivalents/current liabilities (X22), EBIT/total assets (X09), EBIT/shareholder's equity (X28), EBIT/sales (X35), dummy variables: Czech Republic (CZ), Slovak Republic (SK)
Lukason and Laitinen *marked:* **Lukason**	2018	FR	$L = 1.599 - 0.475X22 - 2.154X14 - 4.307X10 - 0.073X01 - 5.880X09$	cash and cash equivalents/current liabilities (X22), cash flow/total liabilities (X14), total liabilities/total assets (X10), sales/total assets (X01), EBIT/total assets (X09)
Adamko, Kliestik and Kovacova *marked:* **Adamko**	2018	SK	$t = -1.1766 + 0.4838 \cdot (X11 - X15) - 0.1828X09 + 1.4733X10 - 1.3745X14$	current assets/total assets (X11), current liabilities/total assets (X15), EBIT/total assets (X09), total liabilities/total assets (X10), cash flow/total liabilities (X14)
Proposed **Logit Model**	2021	V4	$PM = -3.2 - X07 + 3.2X10$	net income/total assets (X07), total liabilities/total assets (X10)

The predictor working capital to total assets ratio X08 in Durica, Valaskova and Janoskova model and Adamko, Kliestik and Kovacova model was calculated as subtraction of X11 (current assets/total assets) and X15 (current liabilities/total assets) due to its absence in our dataset.

The Zmijewski model (1984) is the oldest of all compared models with a focus on US companies. However, besides the proposed model, it is the strongest model among the compared ones according to the shape of the ROC curve depicted in Figure 6. Moreover, again besides the proposed model, it is the model with the smallest predictor set of three items. Lukason and Adamko models with four and five predictors are in the following group with very similar performance metrics. Kliestik and Durica models with 13 respectively 17 predictors are in the last group based on the performance metrics. Table 9 shows all major metrics of the compared models. Our proposed model contains the lowest number of predictors, and it performs better on the testing sample than all compared models in all performance metrics. These results support our research approach.
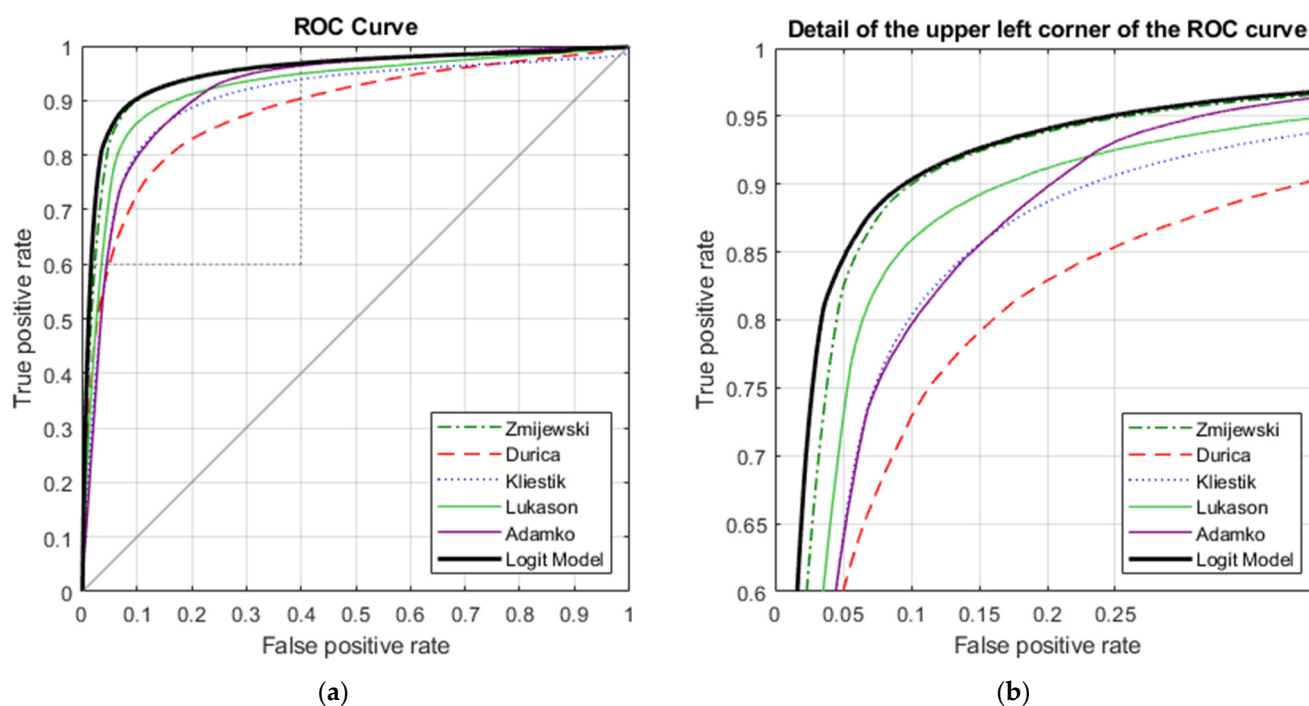
Figure 6. ROC curves of compared models. (**a**) The shape of ROC curves on the whole interval, (**b**) the detail of the ROC curves.

Table 9. Performance metrics of compared models on the testing sample.

Model	Number of Predictors	Accuracy	Sensitivity	Specificity	F_1 Score	MCC	AUC
Zmijewski	3	0.8456	0.9309	0.8289	0.6638	0.6162	0.9447
Durica	17	0.7855	0.8431	0.7742	0.5628	0.4870	0.8822
Kliestik	13	0.7032	0.9290	0.6589	0.5062	0.4386	0.9030
Lukason	5	0.8326	0.9052	0.8184	0.6392	0.5843	0.9238
Adamko	4	0.8481	0.8592	0.8459	0.6494	0.5880	0.9196
Logit Model	2	0.9391	0.8110	0.9642	0.8136	0.7773	0.9507

Our model provides a simple classification tool for companies to identify potential problems of financial stability reliably and quickly based on this comparison. The model is based on two predictors, namely, return on assets and debt ratio. The results demonstrate that the debt ratio is one of the key factors for possible financial failure as in previous research by Zmijewski (1984) [20], Durica et al. (2019) [45], Kliestik et al. (2018) [62], Lukason and Laitinen (2018) [61] and Adamko et al. (2018) [60]. Many businesses increase their debt levels to increase business performance. In addition to the debt ratio, the return on assets is an important indicator contributing to a comprehensive assessment of financial distress. This indicator is statistically significant, similar to the scientific studies by Zmijewski (1984) [20] and Kliestik et al. (2018) [62], in contrast to Durica et al. (2019) [45], Lukason and Laitinen (2018) [61] and Adamko et al. (2018) [60]. Although all the compared models have a higher number of significant indicators, our proposed model exceeds their performance metrics. We find that activity and liquidity indicators, which explain the situation in the short term, do not affect financial stability so considerably. In other words, companies should adopt long-term plans to develop future business activities. This model can serve as a tool for assessing the return on assets, assuming an increase in indebtedness in maintaining the same default probability as a decisive credibility factor from the perspective of stakeholders. Thus, a company or other entity finds out what profitability level must be brought by increased indebtedness without affecting the bankruptcy likelihood. The results reveal that the company should effectively manage the corporate performance with

an emphasis on the extent of a company's leverage. The model represents a useful tool to identify indicators of potential financial distress. The management can make decisions to improve corporate performance before potential bankruptcy.

Table 6 demonstrates that net income to total assets has a similar effect on the bankruptcy probability in the Visegrad Group as Zmijewski model and the Kliestik model. However, our results show that this indicator contributes less to the bankruptcy probability than other compared models. In addition, the debt ratio is the only significant indicator found in all models. As can be seen, the debt ratio has a higher coefficient than the net income to total assets in all models using both indicators.

Moreover, Lukason et al. (2016) [63] explored how the failure distress of young manufacturing companies differs in 11 European countries. They found out that failed young manufacturing micro-firms go through different failure processes and several corporate failures differ due to the firm's age. Moreover, they focus on a statistically significant association between failure process and a firm's country and a firm´s engagement in exporting. They identified 11 financial variables, especially eight financial ratios and three variables describing changes in financial report variables based on literature review from scientific researchers. Laitinen and Suvas (2013) [40] model for Bosnia and Herzegovina achieves the greatest absolute coefficients for all variables. The performance of logit models is very different. The ROC demonstrates that Polish and Finland's models are the best compared to the model for companies in Malta, Greece, Bulgaria, and the United Kingdom. The universal model achieves an AUC at a level of 77.70%. Finally, they demonstrated that there is a significant difference among the European countries. These studies show the differences among the individual countries and show that testing the proposed model on data from different countries would be beneficial in revealing the level of model universality, yet we cannot acquire such data, which is one of the major limitations.

The prediction model provides comprehensive information on bankruptcy probability to companies, stakeholders, and policymakers in all-region. This logit model serves as a crucial screening tool for identifying potential problems in corporate governance. Financial managers can make decisions on optimal loan portfolio management without threats to future business activities. Moreover, we recommend using predictive models to make decisions about international and domestic investment as a part of a comprehensive approach to risk management before entering a new market and building business relationships with new business partners. On other hand, financial institutions need effective prediction models to assess credit risk to make a pragmatic decision on potential business loans. These decisions aim at building confidence in banking and credit institutions and banking stability. Finally, this tool is useful for making political decisions to eliminate potential costs for regional disparities or economic cycles in the country.

Limitations. One of the main limitations is caused by difficult access to financial indicators over a longer period and thus the inability to concern with the trend. Secondly, the dataset limitation to the Visegrad group countries does not allow the proposed model to be tested on data outside this area. Third, we cannot use market value indices because most of the companies is not listed on the stock exchange. For absence of this index, we do not compare our model to Altman model and its derivatives.

Future research. The results showed that the exaggerated predictor-set growth does not need to bring any significant improvement. Hence, future research should focus on hybridization, or an ensemble of small logit models, or decision trees consisting of different minimalistic logit models in respective branches. We emphasize that neural networks or support vector machines may improve results. However, the main disadvantages of these methods are almost impossible reproducibility, high computational time consumption, a loss of universality, and the more demanding variable set requirements. However, we think that another area of research may also focus on the optimal management of the debt ratio to have a positive impact on business performance. This issue is promising from the point of view of many authors. Frank and Goyal (2015) [64] explain that the inverse relationship between profitability and leverage is serious for the trade-off theory. Alnori (2021) [65]

demonstrates that the relationship between profitability and leverage ratios is U-shaped in US companies. On other hand, Stryckova (2017) [66] demonstrates that the leverage ratio harms return on equity as a performance indicator in various sectors in the Czech Republic except for the mining and quarrying industry. Hoang et al. (2020) [67] deal with the impact of the debt ratio on corporate performance. The results show that there is a negative statistically significant relationship between return on assets and debt ratio based on the correlation matrix. In addition, regression models demonstrate that the debt ratio reduces business performance based on the fixed effects regression, random-effects model, and ordinally least squares regression model.

6. Conclusions

We created a universal tool for a one-year-ahead prediction of financial distress of companies in Central Europe focusing on simple model reproducibility based on a minimal number of predictors and high predictive performance. We identified the significant variables using logistic regression as a classifying method altogether with monitoring the significance of improvement in the stepwise growth of the predictor set. The final estimation of model parameters was conducted via cross-validation and verified on a large testing sample. Moreover, the results were compared with other universal or regional prediction logit models on the same testing sample to demonstrate the effectiveness and robustness of the proposed model. We found that the financial stability of a company is mainly affected by the return on assets and debt ratio.

We emphasize that the model should be simply interpreted and reproduced. Moreover, we hope that the proposed model will provide prompt information about the financial stability of the company without looking for a huge number of financial indicators from the financial statements. We think that a simple model with excellent results is desirable in today's dynamic world, especially for traditional users including owners, managers, investors, banks, creditors, suppliers, employees, and others. These scientific findings may represent a basic pillar for searching for a better prediction model in further research in credit risk management. Finally, this paper contributes to the development of current theoretical and empirical knowledge on financial prevention using traditional and alternative models based on a wide range of relevant financial and non-financial indicators.

References

1. Geise, A.; Kuczmarska, M.; Pawlowski, J. Corporate Failure Prediction of Construction Companies in Poland: Evidence from Logit Model. *Eur. Res. Stud. J.* **2021**, *24*, 99–116. [CrossRef]
2. Brozyna, J.; Mentel, G.; Pisula, T. Statistical Methods of the Banrkuptcy Prediction in the Logistic Sector in Poland and Slovakia. *Transform. Bus. Econ.* **2016**, *15*, 93–114.
3. Balina, R.; Juszczyk, S. Forecasting Bankruptcy Risk of International Commercial Road Transport Companies. *Int. J. Manag. Enterp. Dev.* **2014**, *13*, 1–20. [CrossRef]
4. Pisula, T. The Usage of Scoring Models to Evaluate the Risk of Bankruptcy on the Example of Companies from the Transport Sector. *Sci. J. Rzesz. Univ. Technol. Ser. Manag. Mark.* **2012**, *19*, 133–151. [CrossRef]
5. Jakubík, P.; Teply, P. *The Prediction of Corporate Bankruptcy and Czech Economy's Financial Stability through Logit Analysis*; IES Working Paper No. 19/2008; Institute of Economic Studies (IES), Charles University in Prague: Prague, Czech Republic, 2008.
6. Indriyanti, M. The Accuracy of Financial Distress Prediction Models: Empirical Study on the World's 25 Biggest Tech Companies in 2015–2016 Forbes's Version. *KnE Soc. Sci.* **2019**, *3*, 442–450. [CrossRef]
7. Sun, J.; Li, H.; Huang, Q.-H.; He, K.-Y. Predicting Financial Distress and Corporate Failure: A Review from the State-of-the-Art Definitions, Modeling, Sampling, and Featuring Approaches. *Knowl.-Based Syst.* **2014**, *57*, 41–56. [CrossRef]
8. Taffler, R.J. The Assessment of Company Solvency and Performance Using a Statistical Model. *Account. Bus. Res.* **1983**, *13*, 295–308. [CrossRef]
9. Fulmer, J.G.; Moon, J.E.; Gavin, T.A.; Erwin, M.J. A Bankruptcy Classification Model for Small Firms. *J. Com. B Len.* **1984**, *66*, 25–37.
10. Altman, E.I.; Iwanicz-Drozdowska, M.; Laitinen, E.K.; Suvas, A. Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *J. Int. Financ. Manag. Account.* **2017**, *28*, 131–171. [CrossRef]
11. Springate, G.L.V. *Predicting the Possibility of Failure in a Canadian Firm: A Discriminant Analysis*; Simon Fraser University: Burnaby, BC, Canada, 1978.
12. Kovacova, M.; Kliestikova, J. Modelling Bankruptcy Prediction Models in Slovak Companies. *SHS Web Conf.* **2017**, *39*, 01013. [CrossRef]
13. Smith, R.F.; Winakor, A.H. Changes in the Financial Structure of Unsuccessful Industrial Corporations. *Bull. Univ. Ill. Urbana-Champaign Campus Bur. Bus. Res.* **1935**, *51*, 44.
14. FitzPatrick, P.J. A Comparison of the Ratios of Successful Industrial Enterprises with Those of Failed Companies. *Certif. Public Account.* **1932**, *6*, 727–731.
15. Ramser, J.R.; Foster, L.O. *A Demonstration of Ratio Analysis*; Bulletin 40; Bureau of Business Research, University of Illinois: Urbana, IL, USA, 1931.
16. Merwin, C.L. *Financing Small Corporations in Five Manufacturing Industries, 1926–1936*; NBER Books; National Bureau of Economic Research, Inc.: Cambridge, MA, USA, 1942.
17. Altman, E.I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *J. Financ.* **1968**, *23*, 589–609. [CrossRef]
18. Beaver, W.H. Financial Ratios As Predictors of Failure. *J. Account. Res.* **1966**, *4*, 71–111. [CrossRef]
19. Ohlson, J.A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *J. Account. Res.* **1980**, *18*, 109–131. [CrossRef]
20. Zmijewski, M.E. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *J. Account. Res.* **1984**, *22*, 59–82. [CrossRef]
21. Pavlicko, M.; Durica, M.; Mazanec, J. Ensemble Model of the Financial Distress Prediction in Visegrad Group Countries. *Mathematics* **2021**, *9*, 1886. [CrossRef]
22. Liang, D.; Tsai, C.-F.; Wu, H.-T. The Effect of Feature Selection on Financial Distress Prediction. *Knowl.-Based Syst.* **2015**, *73*, 289–297. [CrossRef]
23. Valaskova, K.; Kliestik, T.; Kovacova, M.; Radisic, M.; Mirica, C.-O. Bankruptcy Prediction in Specific Economic Conditions of Slovakia: Multiple Discriminant Analysis. In *Vision 2020: Sustainable Economic Development and Application of Innovation Management, Proceedings of the 32nd International Business Information Management Association Conference, Seville, Spain, 15–16 November 2018*; Soliman, K.S., Ed.; International Business Information Management Assoc-Ibima: Norristown, PA, USA, 2018; pp. 6786–6798.
24. Neumaier, I.; Neumaierová, I. Try to calculate your index IN95. *Terno* **1995**, *5*, 7–10.
25. Jakubík, P.; Teplý, P. The JT Index as an Indicator of Financial Stability of Corporate Sector. *Prague Econ. Pap.* **2011**, *20*, 157–176. [CrossRef]
26. Karas, M.; Režňáková, M. A Parametric or Nonparametric Approach for Creating a New Bankruptcy Prediction Model: The Evidence from the Czech Republic. *Int. J. Math. Models Methods Appl. Sci.* **2014**, *8*, 214–223.
27. Vochozka, M.; Straková, J.; Váchal, J. Model to Predict Survival of Transportation and Shipping Companies. *Naše More* **2015**, *62*, 109–113. [CrossRef]
28. Chrastinová, Z. *Methods of Assessing Economic Creditworthiness and Predicting the Financial Situation of Agricultural Enterprises*; VUEPP: Bratislava, Slovakia, 1998; ISBN 978-80-8058-022-3.
29. Gurčík, L. G-index—The financial situation prognosis method of agricultural enterprises. *Agric. Econ. Zemědělská Ekon.* **2012**, *48*, 373–378. [CrossRef]

30. Hurtošová, J. Construction of a Rating Model, a Tool for Assessing the Creditworthiness of a Company [Konštrukcia Ratingového Modelu, Nástroja Hodnotenia Úverovej Spôsobilosti Podniku]. Ph.D. Thesis, Economic University in Bratislava, Bratislava, Slovakia, 2009.

31. Mihalovič, M. Performance Comparison of Multiple Discriminant Analysis and Logit Models in Bankruptcy Prediction. *Econ. Sociol.* **2016**, *9*, 101–118. [CrossRef] [PubMed]

32. Jenčová, S.; Štefko, R.; Vašaničová, P. Scoring Model of the Financial Health of the Electrical Engineering Industry's Non-Financial Corporations. *Energies* **2020**, *13*, 4364. [CrossRef]

33. Kovacova, M.; Kliestik, T. Logit and Probit Application for the Prediction of Bankruptcy in Slovak Companies. *Equilib. Q. J. Econ. Econ. Policy* **2017**, *12*, 775–791. [CrossRef]

34. Štefko, R.; Horváthová, J.; Mokrišová, M. Bankruptcy Prediction with the Use of Data Envelopment Analysis: An Empirical Study of Slovak Businesses. *J. Risk Financial Manag.* **2020**, *13*, 212. [CrossRef]

35. Lozinskaia, A.; Merikas, A.; Merika, A.; Penikas, H. Determinants of the Probability of Default: The Case of the Internationally Listed Shipping Corporations. *Marit. Policy Manag.* **2017**, *44*, 837–858. [CrossRef]

36. Berent, T.; Bławat, B.; Dietl, M.; Krzyk, P.; Rejman, R. Firm's Default—New Methodological Approach and Preliminary Evidence from Poland. *Equilibrium* **2017**, *12*, 753–773. [CrossRef]

37. Noga, T.; Adamowicz, K. Forecasting Bankruptcy in the Wood Industry. *Eur. J. Wood Wood Prod.* **2021**, *79*, 735–743. [CrossRef]

38. Hajdu, O.; Virág, M. A Hungarian Model For Predicting Financial Bankruptcy. *Társad. És Gazd. Közép-És Kelet-Európában Soc. Econ. Cent. East. Eur.* **2001**, *23*, 28–46.

39. Virág, M.; Kristóf, T. Neural Networks in Bankruptcy Prediction—A Comparative Study on the Basis of the First Hungarian Bankruptcy Model. *Acta Oeconomica* **2005**, *55*, 403–426. [CrossRef]

40. Laitinen, E.; Suvas, A. International Applicability of Corporate Failure Risk Models Based on Financial Statement Information: Comparisons across European Countries. *J. Financ. Econ.* **2013**, *1*, 1–26. [CrossRef]

41. Grünberg, M.; Lukason, O. Predicting Bankruptcy of Manufacturing Firms. *Int. J. Trade Econ. Financ.* **2014**, *5*, 93–97. [CrossRef]

42. Delina, R.; Packová, M. Validation of Predictive Bankruptcy Models in the Conditions of the Slovak Republic [Validácia Predikčných Bankrotových Modelov v Podmienkach SR]. *Ekon. Manag.* **2013**, *16*, 101–112.

43. Harumova, A.; Janisova, M. Rating Slovak Enterprises by Scoring Functions. *Ekon. Cas.* **2014**, *62*, 522–539.

44. Gulka, M. Bankruptcy prediction model of commercial companies operating in the conditions of the Slovak Republic [Model predikcie úpadku obchodných spoločností podnikajúcich v podmienkach SR]. *Forum Stat. Slovacum* **2016**, *12*, 16–22.

45. Durica, M.; Valaskova, K.; Janošková, K. Logit Business Failure Prediction in V4 Countries. *Eng. Manag. Prod. Serv.* **2019**, *11*, 54–64. [CrossRef]

46. Kovacova, M.; Kliestik, T.; Valaskova, K.; Durana, P.; Juhaszova, Z. Systematic Review of Variables Applied in Bankruptcy Prediction Models of Visegrad Group Countries. *Oeconomia Copernic.* **2019**, *10*, 743–772. [CrossRef]

47. Prusak, B. Review of Research into Enterprise Bankruptcy Prediction in Selected Central and Eastern European Countries. *Int. J. Financ. Stud.* **2018**, *6*, 60. [CrossRef]

48. Durica, M.; Frnda, J.; Svabova, L. Decision Tree Based Model of Business Failure Prediction for Polish Companies. *Oeconomia Copernic.* **2019**, *10*, 453–469. [CrossRef]

49. Adamko, P.; Kliestik, T. Proposal for a Bankruptcy Prediction Model with Modified Definition of Bankruptcy for Slovak Companies. In Proceedings of the 2nd Multidisciplinary Conference, Madrid, Spain, 2–4 November 2016; pp. 1–7.

50. Chicco, D.; Jurman, G. The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genom.* **2020**, *21*, 6. [CrossRef]

51. Sasaki, Y. The Truth of the F-Measure. *Sch. Comput. Sci.* **2007**, *5*. Available online: https://www.cs.odu.edu/~{}mukka/cs795 sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf (accessed on 18 January 2022).

52. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 2013; ISBN 978-1-118-54835-6.

53. Chow, J.C.K. Analysis of Financial Credit Risk Using Machine Learning. Ph.D. Thesis, Aston University, Birmingham, UK, 2017.

54. Pestov, V. Is the K-NN Classifier in High Dimensions Affected by the Curse of Dimensionality? *Comput. Math. Appl.* **2013**, *65*, 1427–1437. [CrossRef]

55. D'Haultfœuille, X.; Iaria, A. A Convenient Method for the Estimation of the Multinomial Logit Model with Fixed Effects. *Econ. Lett.* **2016**, *141*, 77–79. [CrossRef]

56. Frölich, M. Non-parametric Regression for Binary Dependent Variables. *Econ. J.* **2006**, *9*, 511–540. [CrossRef]

57. Sopitpongstorn, N.; Silvapulle, P.; Gao, J. *Local Logit Regression for Recovery Rate*; Social Science Research Network: Rochester, NY, USA, 2017.

58. Valencia, C.; Cabrales, S.; Garcia, L.; Ramirez, J.; Calderona, D. Generalized Additive Model with Embedded Variable Selection for Bankruptcy Prediction: Prediction versus Interpretation. *Cogent Econ. Financ.* **2019**, *7*, 1597956. [CrossRef]

59. Zhang, Y.; Wang, S.; Ji, G. A Rule-Based Model for Bankruptcy Prediction Based on an Improved Genetic Ant Colony Algorithm. *Math. Probl. Eng.* **2013**, *2013*, e753251. [CrossRef]

60. Adamko, P.; Klieštik, T.; Kováčová, M. An GLM Model for Prediction of Crisis in Slovak Companies. In *Economics and Management: How to Cope With Disrupted Times, Proceedings of the 2nd International Scientific Conference—EMAN 2018, Ljublana, Slovenia, 22 March 2018*; Association of Economists and Managers of the Balkans: Belgrade, Serbia, 2018; pp. 223–228.

61. Lukason, O.; Laitinen, E.K. Failure of Exporting and Non-Exporting Firms: Do the Financial Predictors Vary? *Rev. Int. Bus. Strategy* **2018**, *28*, 317–330. [CrossRef]

62. Kliestik, T.; Vrbka, J.; Rowland, Z. Bankruptcy Prediction in Visegrad Group Countries Using Multiple Discriminant Analysis. *Equilibrium* **2018**, *13*, 569–593. [CrossRef]

63. Lukason, O.; Laitinen, E.K.; Suvas, A. Failure Processes of Young Manufacturing Micro Firms in Europe. *Manag. Decis.* **2016**, *54*, 1966–1985. [CrossRef]

64. Frank, M.Z.; Goyal, V.K. The Profits–Leverage Puzzle Revisited. *Rev. Financ.* **2015**, *19*, 1415–1453. [CrossRef]

65. Alnori, F. Exploring Nonlinear Linkage between Profitability and Leverage: US Multinational versus Domestic Corporations. *J. Int. Financ. Manag. Account.* **2021**, *32*, 311–335. [CrossRef]

66. Stryckova, L. The Relationship Between Company Returns and Leverage Depending on the Business Sector: Empirical Evidence from the Czech Republic. *J. Compet.* **2017**, *9*, 98–110. [CrossRef]

67. Hoang, T.T.; Hoang, L.T.; Phi, T.K.; Nguyen, M.T.; Phan, M.Q. The Influence of the Debt Ratio and Enterprise Performance of Joint Stock Companies of Vietnam National Coal and Mineral Industries Holding Corp. *J. Asian Financ. Econ. Bus.* **2020**, *7*, 803–810. [CrossRef]