Article

Domain Knowledge Features versus LASSO Features in Predicting Risk of Corporate Bankruptcy—DEA Approach

Martina Mokrišová * and **Jarmila Horváthová**

Faculty of Management and Business, University of Prešov, Konštantínova 16, 080 01 Prešov, Slovakia; jarmila.horvathova@unipo.sk
* Correspondence: martina.mokrisova@unipo.sk

Abstract: Predicting the risk of corporate bankruptcy is one of the most important challenges for researchers dealing with the issue of financial health evaluation. The risk of corporate bankruptcy is most often assessed with the use of early warning models. The results of these models are significantly influenced by the financial features entering them. The aim of this paper was to select the most suitable financial features for bankruptcy prediction. The research sample consisted of enterprises conducting a business within the Slovak construction industry. The features were selected using the domain knowledge (DK) approach and Least Absolute Shrinkage and Selection Operator (LASSO). The performance of VRS DEA (Variable Returns to Scale Data Envelopment Analysis) models was assessed with the use of accuracy, ROC (Receiver Operating Characteristics) curve, AUC (Area Under the Curve) and Somers' D. The results show that the DK+DEA model achieved slightly better AUC and Somers' D compared to the LASSO+DEA model. On the other hand, the LASSO+DEA model shows a smaller deviation in the number of identified businesses on the financial distress frontier. The added value of this research is the finding that the application of DK features achieves significant results in predicting businesses' bankruptcy. The added value for practice is the selection of predictors of bankruptcy for the analyzed sample of enterprises.

Keywords: data envelopment analysis; domain knowledge; feature; LASSO; risk of corporate bankruptcy

1. Introduction

Research shows that no company can be sure of its future even in times of peace and prosperity. The problem of companies' risk of bankruptcy is highly relevant today and is being addressed by many researchers. The acceleration in interest in its solution was caused by the events of the last few years (COVID-19, war in Ukraine), especially in Europe. It is necessary to catch earlier signals of bankruptcy, to which business managers should pay increased attention in order to prevent bankruptcy. For this purpose, various methods of selecting bankruptcy prediction features, as well as various bankruptcy prediction models, are suitable. It is proven that domain knowledge plays a significant role in the given process and, when combined with a suitable prediction method, can provide significant results. This is confirmed by the studies of several authors. It is possible to mention the studies of Veganzones and Severin (2021), who selected features based on their popularity in the prior literature, the study of Min and Lee (2008), who used expert opinion, or the study of Zhou et al. (2015), who applied domain knowledge approach. Often used features in bankruptcy prediction are Altman's (1968) features. They were used in the study of Hu (2009) and that of De Andrés et al. (2011). Barboza et al. (2017) combined the features of Altman (1968) with the features of Carton and Hofer (2006), which have a greater impact on financial performance models in the short term. Similarly, Du Jardin (2015) applied financial ratios traditionally used in the literature since Altman (1968). These ratios were chosen based on the main financial dimensions which govern bankruptcy. Tseng and Hu (2010) used features inspired by the research of Lin (1999) and Lin and Piesse (2004).

Several studies (Kirkos 2015; Zvarikova et al. 2017; Kovacova et al. 2019) were published in which the authors examined the occurrence of individual features in bankruptcy prediction models. We followed up the results of the study of Kovacova et al. (2019), who made a review of the most often used bankruptcy prediction features in Visegrad-group countries.

Based on the above mentioned, the research question was as follows: Which way of selecting financial features for DEA model ensures higher performance of the model: the domain knowledge approach or one of data mining techniques—LASSO regression?

This paper follows previous research aimed at finding the most appropriate method of selecting features for DEA models. In previous studies, we can rarely see the comparison of domain knowledge and data mining techniques when selecting features. The mentioned approaches are mostly considered individually. This study is focused on filling this gap in the research. The LASSO+DEA approach is applied, and its results are compared with the selection of features based on expert opinion and their use in DEA (DK+DEA approach). The performances of the LASSO+DEA approach and the DK+DEA approach are compared.

In line with the above mentioned, the aim of this paper was to select the most suitable financial features for bankruptcy prediction based on the comparison of the performance of DEA prediction models.

The remainder of the paper is structured as follows: The Literature Review Section presents different approaches to defining bankruptcy risk and lists studies dealing with methods and features applied in bankruptcy prediction. The Materials and Methods Section describes the research sample and methods used for feature selection and bankruptcy prediction. The Results Section offers the results of feature selection with the use of the domain knowledge and LASSO methods and uses them to create VRS DEA model. The Discussion Section compares the results of the DK+DEA and LASSO+DEA models and discusses them from the point of view of their performance and applied features. The Conclusion Section presents the contributions, added value, limitations and future direction of this research.

2. Literature Review

Determining corporate bankruptcy risk is one of the main challenges of economic and financial research as well as one the most important issues for investors and decision-makers (Korol 2019). Predicting, measuring and assessing the risk of bankruptcy of a company is of particular interest to investors before investing their capital, as the optimization of risk is a prerequisite for the maximum capital profit of the investment, which will ensure payment of dividends. However, value maximization can only occur if capital providers selectively choose a profitable and sustainable business from which they can obtain the maximum share of business income (Agustia et al. 2020). The risk of bankruptcy is an important topic in many scientific articles, which is primarily reflected in the implications for the stakeholders' decisions (Lukason and Camacho-Miñano 2019). Bankruptcy risk (insolvency) can be understood as "the company's inability to meet maturing obligations resulting either from current operations, whose achievement conditions the continuation of activity, or from compulsory levies" (Bordeianu et al. 2011, p. 250). According to Achim et al. (2012), the risk of business bankruptcy is closely related to economic and financial risk. While financial risk is determined by the level of indebtedness, economic risk is dependent on the ratio of fixed and variable costs. It can be said that, in general, knowledge of these risks makes it possible to quantify the risk of bankruptcy of the company. Bankruptcy risk is the risk of a company no longer being able to meet its debt obligations. This risk is also referred to as the risk of failure or insolvency (Campbell 2011).

Bankruptcy risk represents a constant threat to businesses, which determines how long they will survive (Khan et al. 2020). If a business goes bankrupt, in fact, the probability of bankruptcy in connected businesses increases (Battiston et al. 2007), which can have a negative effect on the entire economy. Therefore, predicting the risk of bankruptcy is the

subject of many research studies dealing with the search for the most suitable bankruptcy prediction model as well as the features describing bankruptcy the best.

Research on bankruptcy prediction dates back to Fitzpatrick (1932), who was the first to examine the financial conditions of bankrupt and non-bankrupt firms by comparing the values of their financial ratios. He found that there are significant differences between bankrupt and non-bankrupt companies, especially between liquidity, debt and turnover indicators (Fejér-Király 2015). In the early days of the development of bankruptcy prediction models, discriminant analysis (DA) was very popular. Beaver (1966) applied univariate discriminant analysis to investigate the predictive ability of 30 financial ratios. The best discriminating factor was identified as the working capital/debt ratio. The second one was the net income/total assets ratio (Gameel and El-Geziry 2016). Despite the criticism, this method was a starting point for the development of other models. The most famous bankruptcy-risk-scoring model, known as Z-score, was published by Altman in 1968 (Voda et al. 2021). This model was developed with the use of multiple discriminant analysis. Since the introduction of Altman's model, many other authors (Deakin 1972; Altman et al. 1977; Norton and Smith 1979; Taffler 1983) developed their models based on multiple discriminant analysis. In the 1980s, logistic regression analysis was developed, followed by probit analysis. The first logistic regression model intended to predict the financial situation of businesses was developed by Ohlson (1980). In the next period, many authors (Kim and Gu 2006; Mihalovic 2016; Barreda et al. 2017; Khan 2018; Affes and Hentati-Kaffel 2019) compared the accuracy of the multiple discriminant analysis model and the logistic regression model. These two models were the most used parametric models in bankruptcy prediction (Fejér-Király 2015). Probit analysis has not been as widely used as logistic regression. The first probit model was developed by Zmijewski (1984), followed by Zavgren (1985). Since the 1990s, the development of computer science has enabled the use of more computationally demanding methods in bankruptcy prediction. These methods are mainly non-parametric. Within them, Mousavi et al. (2023) identifies two main groups: machine learning and artificial intelligence, and operation research. Most used methods within the machine learning and artificial intelligence group include artificial neural networks, such as those used by Messier and Hansen (1988), Odom and Sharda (1990), Atiya (2001) and Abid and Zouari (2002), decision trees (Frydman et al. 1985; Chen et al. 2011; Stankova and Hampel 2018), the Bayesian models (Sarkar and Sriram 2001; Aghaie and Saeedi 2009; Cao et al. 2022), genetic algorithms (Kingdom and Feldman 1995; Alfaro-Cid et al. 2007; Bateni and Asghari 2020), modeling based on rough sets (Ahn et al. 2000; Wang and Wu 2017) and support vector machines (Huang et al. 2004; Olson et al. 2012).

The main method within operation research is Data Envelopment Analysis. This method by Simak (1997) was firstly used when predicting corporate failure. In his master thesis, he compared the results of DEA with the results of Altman's Z-score. In recent years, numerous models based on Data Envelopment Analysis have been developed to predict bankruptcy and their results were compared with the results achieved based on other techniques. Cielen et al. (2004) found that DEA outperformed a discriminant analysis model and a rule induction (C5.0) model in terms of their classification accuracy. Ouenniche and Tone (2017) proposed the out-of-sample evaluation of decision-making units by applying DEA. Out-of-sample framework was based on an instance of case-based reasoning methodology. They found that "DEA as a classifier is a real contender to Discriminant Analysis, which is one of the most commonly used classifiers by practitioners" (Ouenniche and Tone 2017, p. 249). Premachandra et al. (2009) compared the results of an additive DEA model with the results of a Logit model. They found that DEA outperformed the Logit model in evaluating bankruptcy out of sample. Condello et al. (2017, p. 2186) found that DEA has "a greater capacity for bankruptcy prediction, while Logit Regression and Discriminant Analysis perform better in non-bankruptcy and overall prediction in the short term". Janova et al. (2012) achieved similar results. They found that the additive DEA model seems to perform well in correctly identifying bankrupt agricultural businesses. On

the other hand, it is less powerful when identifying non-bankrupt agricultural businesses. The performance of DEA models is assessed mainly with the use of sensitivity, specificity, or overall accuracy. In this regard, Premachandra et al. (2011) pointed out that the cut-off point of 0.5 traditionally used to classify bankrupt and non-bankrupt businesses may not be appropriate for the DEA model. According to these authors "depending on the precision with which predictions for bankrupt and non-bankrupt businesses need to be done, the decision maker has to determine an appropriate cut-off point", Premachandra et al. (2011, p. 623). Stefko et al. (2020) determined the optimal cut-off of the additive DEA model at a point in which the sum of sensitivity and specificity is the highest. Stankova and Hampel (2023) selected an optimal threshold by applying the Youden index and distance from the corner. They found that "selecting a suitable threshold improves specificity visibly with only a small reduction in the total accuracy" (Stankova and Hampel 2023, p. 129).

In the development of the above-mentioned models, the variables included in the model are as important as the method applied (Nurcan and Köksal 2021). In order to select appropriate variables from high-dimensional datasets, various dimensionality reduction methods can be applied. Depending on whether the original features are transformed into new features or not, feature extraction methods and feature selection methods are differentiated (Wang et al. 2016; Li et al. 2020). Feature extraction methods transform existing features into a lower-dimensional space (new set of features) while preserving the original relative distance between the features (Subasi and Gursoy 2010; Li et al. 2020). Well-known feature extraction methods often used in current research include Principal Component Analysis (Adisa et al. 2019; Karas and Reznakova 2020), Multidimensional Scaling (Tang et al. 2020) and Isometric Mapping (Gao et al. 2020). Since the new set of features is different from the original ones, it may be difficult to interpret them (Wang et al. 2016). When using feature selection methods, the original features are sorted according to specific criteria and features with the highest ranking are selected to form a subset (Li et al. 2020). Among the feature selection methods, we can differentiate between filter, wrapper, embedded and combined methods (Liu et al. 2018). Filter methods examine each feature independently while ignoring the individual performance of the feature in the relation to the group. Within filter methods, researchers frequently use *t*-test (Chandra et al. 2009; Xiao et al. 2012), correlation analysis (Zhou et al. 2012) and stepwise methods (Lin et al. 2010). Wrapper methods use machine learning algorithms to evaluate the performance of selected feature subsets. Within them, decision trees (Ratanamahatana and Gunopulos 2003), Naive Bayes (Chen et al. 2009), artificial neural networks (Ledesma et al. 2008) and genetic algorithms (Amini and Hu 2021) are often used. The results of wrapper methods are often superior to the results of filter methods; however, the computational cost of wrapper methods is high. Embedded methods integrate feature selection and learning procedures. Important embedded techniques are regularization approaches which have recently become more and more interesting, for example, LASSO (Fonti and Belitser 2017; Cao et al. 2022; Paraschiv et al. 2021), and Elastic net (Jones et al. 2016; Amini and Hu 2021). Combined methods include different types of feature selection measures, such as filter and wrapper.

Various methodologies have been applied to select features for DEA models. Cielen et al. (2004) used variables according to their efficiency to predict bankruptcy in prior research. Similarly, Psillaki et al. (2010) focused on financial ratios which appeared to be most successful in previous studies. Premachandra et al. (2009) approached this issue in the same way. When creating DEA models, they used ratios which were applied in past bankruptcy literature, and some of them were the same as the ratios used by Altman (1968) and Cielen et al. (2004). The ratios selected by Premachandra et al. (2009) were later applied in the study of Condello et al. (2017) and other studies. Min and Lee (2008) combined expert opinion and factor analysis when selecting features for DEA models. The resulting set of indicators contained the most relevant financial classification dimensions, while taking into account the mathematical relationships among ratios as well. Sueyoshi and Goto (2009) applied Principal Component Analysis to reduce the number of financial

factors in order to reduce the computational burden of the DEA-DA model. Stefko et al. (2021) used Principal Component Analysis and Multidimensional Scaling when selecting inputs and outputs for DEA models. Huang et al. (2015) selected variables for DEA models based on gray relational analysis. They proved this method to be an effective technique for obtaining variables for DEA models. Gray relational analysis was later used in this way by Nurcan and Köksal (2021) as well. Lee and Cai (2020) were dealing with the curse of dimensionality in DEA. They proposed the LASSO variable selection technique and combined it in a sign-constrained convex nonparametric least squares (SCNLS) to support estimating the production function using DEA for small datasets. They also proved that this approach provides useful guidelines for DEA with small datasets. Chen et al. (2021) were inspired by their approach and proposed a simplified two-step LASSO+DEA approach to handle the dimensionality of data entering the DEA models via LASSO. They used standard cross-validation LASSO to select an optimal number of regressors. These regressors were used in the DEA model. As an important advantage of this approach against the study of Lee and Cai (2020), Chen et al. (2021) state that tuning parameter λ was not chosen manually, but it was determined based on optimizing the classical cross-validation criterion to optimally select the relevant variables.

3. Materials and Methods

In this paper, two approaches to feature selection were compared. As the first one, domain knowledge was applied. As the second technique, feature selection based on LASSO was used. Based on these two approaches, two sets of variables were chosen. With the use of these data, VRS DEA models were formed, the results of which were assessed and compared with the use of accuracy, ROC curve, AUC and Somers' D.

3.1. Description of the Research Sample

The input database for the prediction of financial distress of companies operating under SK NACE 41-Construction of buildings consisted of data from the financial statements of 2660 companies. The database of the financial statements of these companies was provided by CRIF–Slovak Credit Bureau, s.r.o. (CRIF 2023).

In order to prepare the research sample for analysis, businesses with zero sales and incomplete records were removed. Since the DEA method is sensitive to outliers, it was necessary to identify and remove them from the analyzed sample. For this purpose, kernel density estimates (Scott 1992) were created for all analyzed indicators using the Epanechnikov kernel function, which was applied in studies by Produit et al. (2010), Gyamerali et al. (2019), Moraes et al. (2021). After excluding the outliers, we continued to work with a sample of 1349 businesses. In order to use the DEA method, the assumption of bankruptcy was established. The analyzed businesses were divided into prosperous and non-prosperous ones based on the criteria reflecting valid Slovak legislation and practice. Non-prosperous businesses included businesses which fulfilled the following criteria: negative EAT (earnings after taxes), equity to liabilities ratio lower than 0.08, and current ratio lower than 1 (Valaskova et al. 2017). The analyzed sample contained 1282 prosperous and 67 non-prosperous businesses.

The construction industry was chosen because it is one of the few industries that can have a stabilizing effect on the economy; this segment is an indicator of economic development and affects the development of other industries and segments of the economy (MTSR 2019; PS Stavby 2021). Therefore, it is necessary to pay attention to the prediction of financial difficulties of companies operating in this industry and to identify possible risks these companies have to face.

3.2. Selection of Financial Features

Financial ratios are vital for predicting the bankruptcy of firms. They were used in this way by many researchers (Snircova 1997; Platt and Platt 2006; Li and Sun 2008; Chen

and Du 2009; Memić 2015; Horvathova and Mokrisova 2019). In this study, the following financial ratios were considered (see Table 1).

Table 1. Financial features' formulas.

Indicator	Formula	Indicator	Formula
Net working capital to current assets (NWCCA)	$\frac{net\ working\ capital}{current\ assets}$	Total revenues to total assets (TRTA)	$\frac{total\ revenues}{total\ asets}$
Net working capital to total assets (NWCTA)	$\frac{net\ working\ capital}{total\ assets}$	Total asset turnover ratio (TATR)	$\frac{total\ sales}{total\ assets}$
Netto cash flow to total assets (NCFTA)	$\frac{netto\ cash\ flow}{total\ assets}$	Liabilities to equity ratio (LER)	$\frac{liabilities}{equity}$
Netto cash flow to liabilities (NCFL)	$\frac{netto\ cash\ flow}{liabilities}$	Liabilities to total assets (LTA)	$\frac{liabilities}{total\ assets}$
Netto cash flow to short-term liabilities (NCFSL)	$\frac{netto\ cash\ flow}{short-term\ liabilities}$	Equity to total assets (ETA)	$\frac{equity}{total\ assets}$
Cash ratio (CaR)	$\frac{financial\ assets}{short-term\ liabilities}$	Financial leverage (FL)	$\frac{total\ assets}{equity}$
Quick ratio (QR)	$\frac{short-term\ receivables+financial\ assets}{short-term\ liabilities}$	Return on assets with EBIT (ROA$_{EBIT}$)	$\frac{EBIT}{assets}$
Current ratio (CuR)	$\frac{short-term\ assets}{short-term\ liabilities}$	Return on equity (ROE)	$\frac{EAT}{equity}$
Equity to fixed assets ratio (EFAR)	$\frac{equity}{fixed\ assets}$	Return on sales (ROS)	$\frac{EAT}{sales}$
Equity and long-term liabilities to fixed assets ratio (ELLFAR)	$\frac{equity+long-term\ liabilities}{fixed\ assets}$	Return on costs (ROC)	$\frac{EAT}{costs}$
Short-term liabilities to total assets (SLTA)	$\frac{short-term\ liabilities}{total\ assets}$	Return on assets with EAT (ROA$_{EAT}$)	$\frac{EAT}{assets}$
Receivables turnover ratio (RTR)	$\frac{sales}{short-term\ receivables}$	Return on sales with EBITDA (ROS$_{EBITDA}$)	$\frac{EBITDA}{sales}\times100$
Short-term liabilities turnover ratio (SLTR)	$\frac{sales}{short-term\ liabilities}$	Cost ratio (CR)	$\frac{costs}{revenues}$

Sources: Snircova (1997); Platt and Platt (2006); Li and Sun (2008); Chen and Du (2009); Memić (2015); Horvathova and Mokrisova (2019).

The selection of input parameters for predicting bankruptcy was carried out using the domain knowledge approach. Financial features were selected based on the research of Kovacova et al. (2019), as follows: the three most frequently used features were selected from each group of indicators mentioned in this research (see Table 2). To avoid the occurrence of highly correlated features within the selected set, the correlation matrix was applied. From pairs of highly correlated indicators with a correlation coefficient higher than 0.9 (Delina and Packova 2013), the indicator with a higher frequency of usage was selected.

Table 2. Most frequently used bankruptcy prediction features in V4 countries.

Activity ratios	Liquidity ratios
Total revenues to total assets	Current ratio
Total asset turnover ratio	Quick ratio
Cash flow to total assets	Working capital to total assets
Profitability ratios	**Debt ratios**
Return on assets with EAT	Liabilities to total assets
Return on equity	Equity to total assets
Return on assets with EBIT	Cash flow to liabilities

Source: Kovacova et al. (2019).

The second set of financial features was selected with the use of LASSO penalized logistic regression. A logistic regression model is determined by the probability of success of the dependent variable, while this category is coded as 1 and another category is coded as 0. For k independent variables, the probability that the dependent variable is equal to 1 is expressed as follows (1) (Wu et al. 2009; Rabaca et al. 2023):

$$P(y_i = 1) = \frac{e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}}}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}}} \tag{1}$$

where y_i is the response for observation i, x_{ji} is the j-th predictor for the observation i, β_j is the regression coefficient for the j-th predictor, and β_0 is the intercept.

The logit is expressed by the logarithm of the odds as follows (2) (Rabaca et al. 2023):

$$Logit[P(y_i = 1)] = log\frac{P(y_i = 1)}{P(y_i = 0)} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ji} \tag{2}$$

LASSO is a particular case of penalized least squares regression with a penalty function L1 (Muthukrishnan and Rohini 2016). LASSO penalized logarithmic likelihood function that needs to be maximized can be written as follows (3) (Rabaca et al. 2023; In Hastie et al. 2009):

$$l_\lambda^L(\beta) = \sum_{i=1}^{n} \left[y_i \beta_0 + \sum_{j=1}^{k} y_i \beta_j x_{ji} - log\left(1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}} \right) \right] \tag{3}$$

where λ is the penalty parameter, and n is the number of observations.

Penalty parameter $\lambda \geq 0$ controls the amount of regularization applied to the estimate (Zhao and Yu 2006). The optimal value of λ (λ_{min}) is usually determined with the use of a 10-fold cross-validation method (Liu et al. 2021). The advantage of LASSO is that it improves the prediction accuracy and interpretability of the model by combining the good properties of ridge regression and subset selection. If there is a high correlation in the group of predictors, LASSO selects only one of them and shrinks the rest to zero (Muthukrishnan and Rohini 2016).

3.3. Method Used for Bankruptcy Prediction

To identify businesses that are threatened with bankruptcy, the DEA method was applied. The DEA model was built in the DEAFrontier software (Zhu 2023). Since this software cannot work with negative values, in accordance with the approach of the software creator, a positive constant has been added to the values of indicators (Seiford and Zhu 2002). According to Silva Portela et al. (2004), if the research sample contains negative values, it is necessary to use a model with the application of variable returns to scale. In accordance with the mentioned approach, the VRS DEA model was applied in this paper. This model assumes variable return to scale, and it was developed by Banker et al. (1984). The dual input-oriented VRS DEA model can be written as follows (4):

Minimize $\qquad\qquad\qquad \theta_q - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+\right),$
Subject to
$\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q x_{iq}, \qquad\qquad\qquad\qquad\qquad\qquad i = 1,2,\ldots,m,$
$\sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq} \qquad\qquad\qquad\qquad\qquad\qquad k = 1,2,\ldots,r, \qquad (4)$
$\sum_{j=1}^{n} \lambda_j = 1,$
$s_i^- \geq 0, \ s_k^+ \geq 0.$

where θ_q is the value of objective function, ε is the non-Archimedean infinitesimal value, x_{ij} and y_{kj} are the inputs and outputs of the DMU_j, x_{iq} and y_{kq} are the inputs and outputs of

the DMU_q, m and r are the number of inputs and outputs, respectively, n is the number of DMUs, λ_j is convex coefficient, and s_i^- and s_k^+ are the input and output slack variables.

With the use of the VRS DEA model, businesses which possibly have financial difficulties were identified. In line with the approach of Premachandra et al. (2009), two sets of features were divided into inputs and outputs, as follows: The smaller (inferior) values in the financial ratios, which could possibly cause financial failure, were considered to be input variables. In contrast, the larger (superior) values in those ratios, which could cause financial failure, were classified as output variables. In this approach, businesses that possibly have financial difficulties form the financial distress frontier. Score of these businesses are equal to 1. Financially healthy businesses are then expected to lie inside the financial distress possibility set, which is shaped by the financial distress frontier.

Premachandra et al. (2011) pointed to the need to find a suitable cut-off value of the DEA model at which the classification accuracy of the given model is optimal. In this paper, the Youden index was used to determine the optimal cut-off, which is calculated as follows: *Youden index* = *Sensitivity* + *Specificity* − 1 (Hajian-Tilaki 2018). The optimal cut-off point is determined at the point where the maximum of *Sensitivity* + *Specificity* is achieved (Youden 1950; In Yin and Tian 2014).

4. Results

From the results presented in Table 3, it is clear that the analyzed sample of companies achieved the required liquidity values, which means that most of the companies are able to pay their liabilities. Since liquidity is one of the representatives of the financial risk of companies and its low values can put companies in a state of financial distress, these results can be evaluated positively. Equally good results are indicated by the median of the indicator net working capital to current assets, which represents 21%. This value is not optimal, but it can be considered acceptable from the point of view of financial risk. The results of the profitability indicators can also be evaluated positively, as the median of them is positive and ranges from 9% to 2%. The costs ratio (0.98) also corresponds to it. The results of this indicator gives companies room for profit creation.

The total asset turnover ratio reaches a value of (1.46 or 1.43), which can be considered an adequate turnover rate considering the subject of business activity.

Less good results are indicated by indebtedness values. The share of liabilities in total assets is up to 68%, 56% of which are short-term liabilities. Liabilities are 1.69 times higher than the company's equity and, thus, the indicator liabilities to equity ratio does not reach the required optimal value. It is precisely the indebtedness of businesses that can be considered a weak point of the analyzed sample, which represents a risk of financial distress for them.

Table 3. Descriptive statistics of financial features.

	Mean	Median	Minimum	Maximum	Std. Dev.
NWCCA	0.07	0.21	−5.84	0.97	0.87
NWCTA	0.18	0.15	−1.08	0.97	0.40
NCFTA	0.10	0.07	−0.58	0.80	0.14
NCFL	0.33	0.11	−5.05	7.59	0.83
NCFSL	0.41	0.14	−5.14	7.66	0.95
CaR	1.50	0.32	−0.76	32.91	3.69
QR	2.45	1.11	0.02	32.91	4.24
CuR	2.75	1.26	0.15	32.91	4.42
EFAR	3.25	1.01	−45.69	116.55	9.53
ELLFAR	3.82	1.34	−44.97	125.01	10.09
SLTA	0.56	0.57	0.00	1.50	0.31
RTR	12.35	4.71	0.03	195.11	24.50
SLTR	5.34	2.99	0.00	47.03	6.97
TRTA	1.83	1.46	0.00	9.72	1.50
TATR	1.81	1.43	0.00	9.72	1.51

Table 3. *Cont.*

	Mean	**Median**	**Minimum**	**Maximum**	**Std. Dev.**
LER	4.78	1.69	−47.42	135.61	13.46
LTA	0.64	0.68	0.03	1.50	0.31
ETA	1.92	0.47	−0.33	37.63	4.53
FL	5.78	2.69	−46.42	136.61	13.46
ROA_{EBIT}	0.08	0.04	−0.58	0.79	0.15
ROE	0.11	0.09	−1.97	1.87	0.42
ROS	0.03	0.02	−2.58	1.74	0.21
ROC	0.06	0.02	−0.66	2.01	0.21
ROA_{EAT}	0.05	0.03	−0.58	0.65	0.13
ROS_{EBITDA}	0.10	0.06	−2.02	9.28	0.33
CR	0.96	0.98	0.28	2.97	0.20

Source: authors.

Based on the research of Kovacova et al. (2019) and the procedure described in Section 3.2. *Selection of financial features*, selected DK features were as follows: Total revenues to total assets, Current ratio, Net working capital to total assets, Return on assets with EAT, Return on equity, Netto cash flow to liabilities and Liabilities to total assets. These features were selected with the use of the correlation matrix.

The most relevant predictors according to LASSO penalized logistic regression were identified by optimizing the value of λ_{min} using 10-fold cross validation. At the optimal lambda value 0.0071, 7 financial ratios out of 26 exhibit non-zero coefficients (see Table 4). These indicators are as follows: Liabilities to total assets, Return on costs, Return on equity, Short-term liabilities to total assets, Net working capital to total assets, Netto cash flow to total assets, and Total asset turnover ratio. The coefficients of the rest of indicators were shrunk to zero. A similar approach was used in the study of Chen et al. (2021), who used LASSO, while tuning parameter λ was selected based on optimizing cross-validation criterion. In this way, the authors selected the relevant variables optimally before deploying DEA on these variables. A simplified LASSO+DEA approach was also used by Lee and Cai (2020). However, these authors chose to manually tune parameter λ.

Table 4. Coefficients of the indicators for λ_{min}.

LASSO Logistic Regression. Modeled probability that bankruptcy = yes Model Lambda = 0.0071; %Dev = 0.5617	
Indicator	**Estimates**
Intercept	−8.67
Liabilities to total assets	4.32
Return on costs	−4.18
Return on equity	−1.73
Short-term liabilities to total assets	1.81
Net working capital to total assets	−1.45
Netto cash flow to total assets	−1.31
Total asset turnover ratio	−0.20
Financial leverage	0.00
Equity and long-term liabilities to fixed assets ratio	0.00
Liabilities to equity ratio	0.00
Current ratio	. . .

Source: authors.

Features selected with the use of the DK approach and LASSO penalized logistic regression were used as inputs and outputs for the VRS DEA models. Two VRS DEA models were formulated—the model with the application of DK features (DK+DEA) and the model with the application of LASSO features (LASSO+DEA). Their results are compared in Table 5. In the case of the DK+DEA model, there were 41 businesses which lie on the financial distress frontier. LASSO+DEA model identified 13 less businesses lying on the

financial distress frontier. In the case of DK+DEA, the most numerous group of enterprises is located in the efficiency interval $\langle 0.9; 0.8 \rangle$; on the contrary, in the case of LASSO+DEA, the largest number of identified enterprises is in the interval $\langle 0.5; 0.4 \rangle$.

Table 5. Efficiency results of DEA models.

Efficiency	DK	LASSO
1	44	31
$\langle 1; 0.9 \rangle$	225	24
$\langle 0.9; 0.8 \rangle$	616	32
$\langle 0.8; 0.7 \rangle$	297	37
$\langle 0.7; 0.6 \rangle$	124	71
$\langle 0.6; 0.5 \rangle$	41	207
$\langle 0.5; 0.4 \rangle$	2	559
$\langle 0.4; 0.3 \rangle$	0	360
$\langle 0.3; 0.2 \rangle$	0	28
$\langle 0.2; 0.1 \rangle$	0	0

Source: authors.

For better comparability of the results, the optimal cut-offs for both models were determined with the use of the Youden index.

The optimal cut-off of LASSO+DEA model was determined at the level of 0.59. The classification accuracy for bankrupt businesses at this cut-off was 79.10% (see Table 6). The classification accuracy for non-bankrupt businesses achieved a higher value, 86.66%. The overall classification accuracy of LASSO+DEA model at a cut-off of 0.59 was 86.29%.

Table 6. Results of LASSO+DEA model at optimal cut-off 0.59.

	Predicted: Yes	Predicted: No	% Correct
Observed: yes	53	14	79.10
Observed: no	171	1111	86.66

Source: authors.

In the case of the DK+DEA model, the optimal cut-off was determined at the level of 0.89. At this cut-off, the DK+DEA model achieved high classification accuracy for bankrupt businesses, 97.01%, and lower classification accuracy for non-bankrupt businesses, 78.72 (see Table 7). The overall classification accuracy of DK+DEA model was 79.63%. A slightly higher overall classification accuracy of the DEA model with the application of DK features (85.1%) was achieved by Cielen et al. (2004). DEA models using DK features developed by Premachandra et al. (2009) achieved an overall classification accuracy of 74–86%. Similar to our results, these models achieved higher classification accuracy for bankrupt businesses.

Table 7. Results of DK+DEA model at optimal cut-off 0.89.

	Predicted: Yes	Predicted: No	% Correct
Observed: yes	65	2	97.01
Observed: no	273	1010	78.72

Source: authors.

Based on the results presented in Tables 6 and 7, we can conclude that the DK+DEA model performs better when identifying bankrupt businesses. It means that features selected via the DK approach are more suitable for bankruptcy prediction. The selection of DK features, with the application of which the DEA model with a higher classification accuracy was created, represents the fulfillment of the aim set in this paper.

The confirmation of this result can also be seen on the ROC curve (see Figure 1). The results show that both DEA models achieved excellent classification accuracy; however, the classification accuracy of the DK-DEA model was slightly higher.
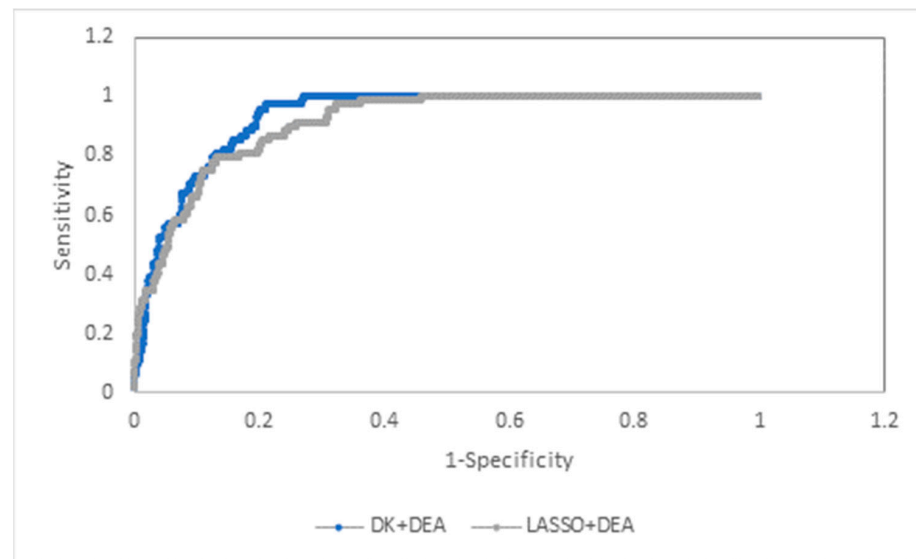
Figure 1. Comparison of ROC curves of DK+DEA and LASSO+DEA models.

5. Discussion

The summary of the research results shows interesting findings. By applying different features, the models achieved different classification accuracies. Table 8 and Figure 2 show the comparison of the bankruptcy prediction results achieved using the DEA model when applying features selected via DK and LASSO.

Table 8. Summary of DK+DEA and LASSO+DEA results.

Item	DK	LASSO
Number of businesses on financial distress frontier	44	31
Number of businesses in financial distress possibility set	1305	1.318
Match in the number of businesses on financial distress frontier	28	28
Difference in the number of businesses on financial distress frontier	16	3
Difference in the number of businesses on financial distress frontier compared to the assumption of bankruptcy	23	36

Source: authors.
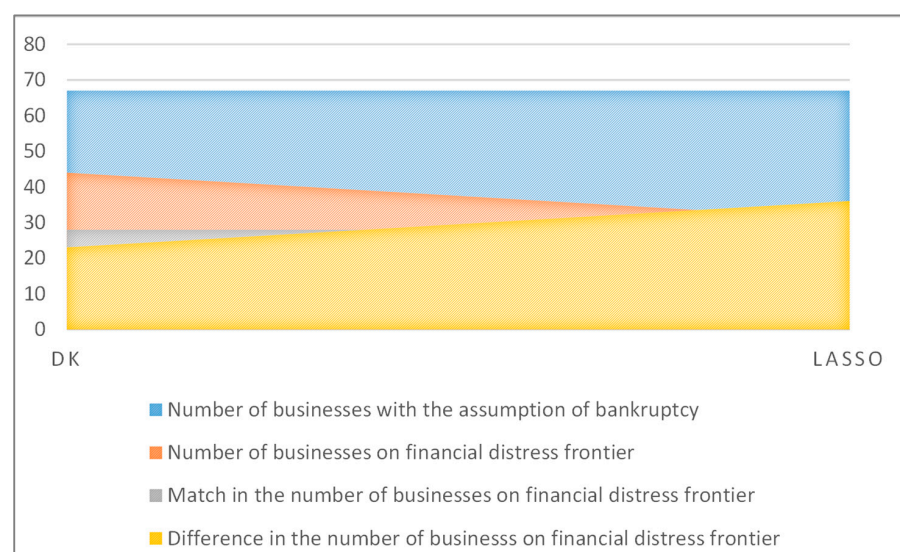
Figure 2. Comparison of classification accuracy of DK+DEA and LASSO+DEA models. Source: authors.

The analysis shows that when applying DK features, the VRS DEA model confirmed the assumption of bankruptcy in 44 businesses, which is 13 more businesses than when applying LASSO features. It is 23 fewer businesses than the assumption of bankruptcy. However, when applying LASSO features, it is 36 fewer businesses compared to the assumption of bankruptcy.

On the other hand, in the case of LASSO, only 3 of 31 businesses were incorrectly identified. These results indicate that the DK+DEA model has a better classification accuracy in relation to the assumption of bankruptcy. However, LASSO+DEA shows a smaller deviation in the number of identified businesses on the financial distress frontier.

Based on the above results, the application of feature selection using the LASSO method appears to be more appropriate. However, it is necessary to continue the analysis and apply other procedures and methods.

To analyze the results in more detail, it is necessary to specify the features used in both cases. They are presented in Table 9.

Table 9. DEA inputs and outputs in two sets of financial features.

LASSO	DK
Liabilities to total assets	*Liabilities to total assets*
Short—term liabilities to total assets	Current ratio
Return on equity	*Return on equity*
Return on costs	Return on assets with EAT
Netto cash flow to total assets	Netto cash flow to liabilities
Total asset turnover ratio	Total revenues to total assets
Net working capital to total assets	*Net working capital to total assets*

Source: authors.

Agreement in the selection of indicators occurred in the case of three indicators, highlighted in italics in Table 9. However, the selection based on the experience of experts seems to be more relevant, as it also includes Current ratio and Insolvency ratio (Netto cash flow to liabilities). Many authors consider these indicators to be important predictors of bankruptcy. This can be confirmed by the definitions of financial health of several authors. Szilagyi (2004) defined a financially healthy business, and, within his definition, he pointed out that such a business is not expected to become insolvent and does not show any sign of a threat to its existence, and it is even able to adequately cover the risks related to indebtedness. The importance of ability to pay was also pointed out by Koh et al. (2015), who defined financial distress as a situation when a business cannot pay the amount owed on the due date. Platt et al. (1995) argue that financial distress occurs when the total value of a company's assets is lower than the total value of creditors' claims. In the long term, this situation can lead to forced liquidation or bankruptcy. For this reason, financial distress is often referred to as a harbinger of bankruptcy and is related to the availability of liquid funds and credit (Hendel 1996). Gestel et al. (2006) characterize financial distress and financial failure as the result of chronic losses that cause a disproportionate increase in liabilities accompanied by a loss of assets' value. It is possible to mention other authors who talk about the ability to repay obligations as an important predictor of bankruptcy (Campbell 2011; Achim et al. 2012).

This means that a financially healthy company is able to pay its obligations and has fulfilled the purpose of its existence—to be profitable. Therefore, the indicators Current liquidity, Netto cash flow to liabilities and Return on assets in DK have their justification. This selection seems to be much more relevant than the LASSO selection.

On the other hand, it should be pointed out that the Current ratio was used as one of the criteria when establishing the assumption of bankruptcy. This indicator was selected as one of the DK features as well. This fact could affect the results of the DK+DEA model.

Table 10 shows the overall performance of the constructed DEA models. We can see that the DK+DEA model achieved slightly better AUC and Somers' D compared to the LASSO+DEA model. Based on it, we can conclude that the selection of DK features is

more appropriate than the selection of LASSO features when predicting the bankruptcy of businesses. If we compare the achieved results with the literature, the results of previous studies are slightly different. Zhou et al. (2015) found that there is no significant difference between the classification performance of models with feature selection guided by data mining techniques and that of those guided by domain knowledge. The findings of Lin et al. (2014), who revealed that a model with LASSO-based feature selection achieved a slightly higher performance in terms of accuracy as well as AUC compared to DK, are also slightly different. However, the comparability of these studies depends on several factors, e.g., research sample, used model, etc.

Table 10. Performance of DEA models.

	DK+DEA	LASSO+DEA
AUC	0.9258	0.9074
Somers' D	0.8516	0.8148

Source: authors.

6. Conclusions

In this research, the features for DEA models were selected with the use of the domain knowledge approach and the LASSO approach. According to DK, the following bankruptcy prediction indicators were chosen: Total revenues to total assets, Current ratio, Net working capital to total assets, Return on assets with EAT, Return on equity, Netto cash flow to liabilities, and Liabilities to total assets. LASSO identified the following predictors of bankruptcy: Liabilities to total assets, Return on costs, Return on equity, Short-term liabilities to total assets, Net working capital to total assets, Netto cash flow to total assets, and Total asset turnover ratio. Subsequently, the performance of the DK+DEA and LASSO+DEA models was compared. Performance was different for both selections at different cut-offs. For the selection of features according to LASSO, the optimal cut-off was 0.59, which means that from this value, businesses were identified as bankrupt. In the case of selecting features based on DK, the optimal cut-off value was at the level of 0.89. Based on this fact, it can be concluded that in the case of DK feature selection, more indicators were identified as predictors of businesses' bankruptcy. Important predictors of bankruptcy found with the DK application include Current ratio, Insolvency ratio (Netto cash flow to liabilities) and Return on assets, which are missing in features selected via LASSO. These features are significant predictors of bankruptcy that are applied in many bankruptcy prediction studies (Reznakova and Karas 2014; Lin et al. 2014; Pavlicko and Mazanec 2022).

The contribution of the paper is the application of DK and LASSO features and VRS DEA model in the evaluation of the financial failure of businesses. The results revealed that the DK+DEA model achieved higher classification and prediction accuracy compared to the LASSO+DEA model. On the other hand, there is a smaller deviation in the number of identified businesses on the financial distress frontier in the LASSO+DEA model.

The added value of this research lies in pointing out the importance of the indicators Current ratio and Return on assets, which were the criteria used to establish the assumption of bankruptcy. Since these indicators entered the DK+DEA model as well, this model achieved higher classification accuracy compared to LASSO+DEA. Therefore, it is necessary to pay more attention to the selection of criteria for determining the assumption of bankruptcy and subsequently to the selection of features based on DK. The managerial implications of this research enable companies and managers from the construction industry to focus on those features that are decisive for the area of evaluating the financial health of companies.

A limitation of the given research was missing and insufficient data. Another limitation was the occurrence of a relatively large number of outliers. Future research will be focused on confirming the significance of selected indicators for predicting the financial failure of companies, and especially on the Current ratio and its use in identifying prosperous and non-prosperous businesses.

References

Abid, Fathi, and Anis Zouari. 2002. Predicting corporate financial distress: A neural networks approach. *Finance India* 16: 601–12. Available online: https://ssrn.com/abstract=1300290 (accessed on 1 June 2023).

Achim, Monica Violeta, Codruta Mare, and Sorin Nicolae Borlea. 2012. A statistical model of financial risk bankruptcy applied for Romanian manufacturing industry. *Procedia Economics and Finance* 3: 132–37. [CrossRef]

Adisa, Juliana Adeola, Samuel Olusegun Ojo, Pius Adewale Owolawi, and Agnieta Beatrijs Pretorius. 2019. Financial Distress Prediction: Principle Component Analysis and Artificial Neural Networks. Paper presented at 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC), Vanderbijlpark, South Africa, November 21–22.

Affes, Zeineb, and Rania Hentati-Kaffel. 2019. Predicting US Banks Bankruptcy: Logit Versus Canonical Discriminant Analysis. *Computational Economics* 54: 199–244. [CrossRef]

Aghaie, Arezoo, and Ali Saeedi. 2009. Using Bayesian Networks for Bankruptcy Prediction: Empirical Evidence from Iranian Companies. Paper presented at 2009 International Conference on Information Management and Engineering, Kuala Lumpur, Malaysia, April 3–5.

Agustia, Dian, Nur Pratama A. Muhammad, and Yani Permatasari. 2020. Earnings management, business strategy, and bankruptcy risk: Evidence from Indonesia. *Heliyon* 6: E03317. [CrossRef]

Ahn, Byeong S., Sung Sik Cho, and Chang Yun Kim. 2000. The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications* 18: 65–74. [CrossRef]

Alfaro-Cid, Eva, Ken Sharman, and Anna I. Esparcia-Alcazar. 2007. A genetic programming approach for bankruptcy prediction using a highly unbalanced database. In *Workshops on Applications of Evolutionary Computation*. Edited by Mario Giacobini. Berlin: Springer, pp. 169–78. [CrossRef]

Altman, Edward I. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23: 589–609. [CrossRef]

Altman, Edward I., Robert G. Haldeman, and Paul Narayanan. 1977. ZETA ANALYSIS, a new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance* 1: 29–54. [CrossRef]

Amini, Fatemeh, and Guiping Hu. 2021. A two-layer feature selection method using Genetic Algorithm and Elastic Net. *Expert Systems with Applications* 166: 114072. [CrossRef]

Atiya, Amir F. 2001. Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks* 12: 929–35. [CrossRef]

Banker, Rajiv D., Abraham Charnes, and William Wager Cooper. 1984. Some models for estimating technical scale inefficiencies in Data Envelopment Analysis. *Management Science* 30: 1078–92. [CrossRef]

Barboza, Flavio, Herbert Kimura, and Edward Altman. 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications* 83: 405–17. [CrossRef]

Barreda, Albert A., Yoshimasa Kageyama, Dipendra Singh, and Sandra Zubieta. 2017. Hospitality Bankruptcy in United States of America: A Multiple Discriminant Analysis-Logit Model Comparison. *Journal of Quality Assurance in Hospitality and Tourism* 18: 86–106. [CrossRef]

Bateni, Leila, and Farshid Asghari. 2020. Bankruptcy Prediction Using Logit and Genetic Algorithm Models: A Comparative Analysis. *Computational Economics* 55: 335–48. [CrossRef]

Battiston, Stefano, Domenico Delli Gatti, Mauro Gallegati, Bruce Greenwald, and Joseph E. Stiglitz. 2007. Credit chains and bankruptcy propagation in production networks. *Journal of Economic Dynamics & Control* 31: 2061–84. [CrossRef]

Beaver, William H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research* 4: 71–111. [CrossRef]

Bordeianu, Gabriela-Daniela, Florin Radu, Marius Dumitru Paraschivescu, and Willi Păvăloaia. 2011. Analysis models of the bankruptcy risk. *Economy Transdisciplinary Cognition* 14: 248–59. Available online: https://www.ugb.ro/etc/etc2011no1/FIN-1-full.pdf (accessed on 20 July 2023).

Campbell, R. Harvey. 2011. Bankruptcy Risk. Available online: https://financial-dictionary.thefreedictionary.com/Bankruptcy+risk (accessed on 1 June 2023).

Cao, Yi, Xiaoquan Liu, Jia Zhai, and Shan Hua. 2022. A two-stage Bayesian network model for corporate bankruptcy prediction. *International Journal of Finance & Economics* 27: 455–72. [CrossRef]

Carton, Robert B., and Charles W. Hofer. 2006. *Measuring Organizational Performance*. Cheltenham: Edward Elgar Publishing.

Chandra, D. Karthik, Vadlamani Ravi, and Indranil Bose. 2009. Failure prediction of dotcom companies using hybrid intelligent techniques. *Expert Systems with Applications* 36: 4830–7. [CrossRef]

Chen, Hui-Ling, Bo Yang, Gang Wang, Jie Liu, Xin Xu, Su-Jing Wang, and Da-You Liu. 2011. A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method. *Knowledge-Based Systems* 24: 1348–59. [CrossRef]

Chen, Jingnian, Houkuan Huang, Shengfeng Tian, and Youli Qu. 2009. Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications* 36: 5432–35. [CrossRef]

Chen, Wei-Sen, and Yin-Kuan Du. 2009. Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications 36 Pt* 2: 4075–86. [CrossRef]

Chen, Ya, Mike G. Tsionas, and Valentin Zelenyuk. 2021. LASSO+DEA for small and big wide data. *Omega* 102: 120419. [CrossRef]

Cielen, Anja, Ludo Peeters, and Koen Vanhoof. 2004. Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research* 154: 526–32. [CrossRef]

Condello, Silva, Antonio Del Pozzo, Salvatore Loprevite, and Bruno Ricca. 2017. Potential and limitations of D.E.A. s a bankruptcy prediction tool in the light of a study on Italian listed companies. *Applied Mathematical Sciences* 11: 2185–207. [CrossRef]

CRIF. 2023. *Financial Statements of Analyzed Businesses*. Bratislava: Slovak Credit Bureau, s.r.o.

De Andrés, Javier, Pedro Lorca, Francisco Javier de Cos Juez, and Fernando Sánchez-Lasheras. 2011. Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications* 38: 1866–75. [CrossRef]

Deakin, Edward B. 1972. A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research* 10: 167–79. [CrossRef]

Delina, Radoslav, and Miroslava Packova. 2013. Validácia predikčných bankrotových modelov v podmienkach SR [Validation of bankruptcy prediction models in Slovak conditions]. *Ekonomie a Management [Economics and Management]* 16: 101–12. Available online: https://otik.uk.zcu.cz/bitstream/11025/17515/1/2013_3%20Validacia%20predikcnych%20bankrotivych%20modelov%20v%20podmienkach%20SR.pdf (accessed on 13 June 2023).

Du Jardin, Philippe. 2015. Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research* 242: 286–303. [CrossRef]

Fejér-Király, Gergely. 2015. Bankruptcy Prediction: A Survey on Evolution, Critiques, and Solutions. *Acta Universitatis Sapientiae, Economics and Business* 3: 93–108.

Fitzpatrick, Paul J. 1932. *A Comparison of the Ratios of Successful Industrial Enterprises with Those of Failed Companies*. Washington, DC: The Accountants' Publishing Company.

Fonti, Valeria, and Eduard Belitser. 2017. *Feature Selection Using LASSO*. VU Amsterdam Research Paper in Business Analytics. Amsterdam: VRIJE Universiteit Amsterdam, pp. 1–25. Available online: https://www.semanticscholar.org/paper/Paper-in-Business-Analytics-Feature-Selection-using-Fonti-Belitser/24acd159910658223209433cf4cbe3414264de39 (accessed on 5 July 2023).

Frydman, Halina, Edward I. Altman, and Duen-Li Kao. 1985. Introducing recursive portioning for financial classification: The case of financial distress. *Journal of Finance* 40: 269–91. [CrossRef]

Gameel, Mohamed, and Khairy El-Geziry. 2016. Predicting Financial Distress: Multi Scenarios Modeling Using Neural Network. *International Journal of Economics and Finance* 8: 159–66. [CrossRef]

Gao, Shuzhi, Sixuan Zhang, Yimin Zhang, and Yue Gao. 2020. Operational reliability evaluation and prediction of rolling bearing based on isometric mapping and NoCuSa-LSSVM. *Reliability Engineering and System Safety* 201: 106968. [CrossRef]

Gestel, Tony Van, Bart Baesens, Johan A. K. Suykens, and Dirk Van den Poel. 2006. Bayesian Kernel Based Classification for Financial Distress Detection. *European Journal of Operational Research* 172: 979–1003. [CrossRef]

Gyamerali, Samuel Asante, Philip Ngare, and Dennis Ikpe. 2019. Crop Yield Probability Density Forecasting via Quantile Random Forest and Epanechnikov Kernel Function. Available online: http://ir.mksu.ac.ke/handle/123456780/4393 (accessed on 7 August 2023).

Hajian-Tilaki, Karimolla. 2018. The choice of methods in determining the optimal cut-off value for quantitative diagnostic test evaluation. *Statistical Methods in Medical Research* 27: 2374–83. [CrossRef] [PubMed]

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed. Berlin: Springer.

Hendel, Igal. 1996. Competition under financial Distress. *The Journal of Industrial Economics* 5: 309–24. [CrossRef]

Horvathova, Jarmila, and Martina Mokrisova. 2019. *Efektívnosť Podniku Verzus jeho Bankrot [Business Efficiency Versus Business Bankruptcy]*. Prešov: Bookman.

Hu, Yi-Chung. 2009. Bankruptcy prediction using ELECTRE-based single-layer perceptron. *Neurocomputing* 72: 3150–7. [CrossRef]

Huang, Chao, Chong Dai, and Miao Guo. 2015. A hybrid approach using two-level DEA for financial failure prediction and integrated SE-DEA and GCA for indicators selection. *Applied Mathematics and Computation* 251: 431–41. [CrossRef]

Huang, Zan, Hsinchun Chen, Chia-Jung Hsu, Wun-Hwa Chen, and Saushan Wu. 2004. Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems* 37: 543–58. [CrossRef]

Janova, Jitka, Jan Vavrina, and David Hampel. 2012. DEA as a tool for bankruptcy assessment: The agribusiness case study. Paper presented at 30th International Conference Mathematical Methods in Economics 2012, Karviná, Czech Republic, September 11–13; Available online: http://mme2012.opf.slu.cz/proceedings/pdf/065_Janova.pdf (accessed on 1 July 2023).

Jones, Stewart, David Johnstone, and Roy Wilson. 2016. Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks. *Journal of Business Finance & Accounting* 44: 3–34. [CrossRef]

Karas, Michal, and Mária Reznakova. 2020. Cash flow indicators in the prediction of financial distress. *Engineering Economics* 31: 525–35. [CrossRef]

Khan, Khurram A., Robert Dankiewicz, Yana Kliuchnikava, and Judit Oláh. 2020. How do entrepreneurs feel bankruptcy? *International Journal of Entrepreneurial Knowledge* 1: 89–101. [CrossRef]

Khan, Usama E. 2018. Bankruptcy Prediction for Financial Sector of Pakistan: Evaluation of Logit and Discriminant Analysis Approaches. *Pakistan Journal of Engineering Technology and Science* 6: 210–20. [CrossRef]

Kim, Hyunjoon, and Zheng Gu. 2006. Predicting Restaurant Bankruptcy: A Logit Model in Comparison with a Discriminant Model. *Journal of Hospitality & Tourism Research* 30: 474–93. [CrossRef]

Kingdom, J., and K. Feldman. 1995. *Genetic Algorithms for Bankruptcy Prediction*. Search Space Research Report No. 01-95. London: Search Space Ltd.

Kirkos, Efstathios. 2015. Assessing methodologies for intelligent bankruptcy prediction. *Artificial Intelligence Review* 43: 83–123. [CrossRef]

Koh, SzeKee R., Robert B. Durand, Lele Dai, and Milicent Chang. 2015. Financial distress: Lifecycle and corporate Restructuring. *Journal of Corporate Finance* 33: 19–33. [CrossRef]

Korol, Tomasz. 2019. Dynamic bankruptcy prediction models for European enterprises. *Journal of Risk and Financial Management* 12: 185. [CrossRef]

Kovacova, Maria, Tomas Kliestik, Katarina Valaskova, Pavol Durana, and Zuzana Juhaszova. 2019. Systematic review of variables applied in bankruptcy prediction models of Visegrad group countries. *Oeconomia Copernicana* 10: 743–72. [CrossRef]

Ledesma, Sergio, Gustavo Cerda, Gabriel Aviña, Donato Hernández, and Miguel Torres. 2008. Feature Selection Using Artificial Neural Networks. In *Advances in Artificial Intelligence*. Edited by Alexander Gelbukh and Eduardo F. Morales. Paper presented at Micai 2008, October 27–31. Berlin/Heidelberg: Springer, vol. 5317.

Lee, Chia-Yen, and Jia-Ying Cai. 2020. LASSO variable selection in data envelopment analysis with small datasets. *Omega* 91: 102019. [CrossRef]

Li, Hui, and Jie Sun. 2008. Ranking-order case-based reasoning for financial distress prediction. *Knowledge-Based Systems* 21: 868–78. [CrossRef]

Li, Mengmeng, Haofeng Wang, Lifang Yang, You Liang, Zhigang Shang, and Hong Wan. 2020. Fast hybrid dimensionality reduction method for classification based on feature selection and grouped feature extraction. *Expert Systems with Applications* 150: 113277. [CrossRef]

Lin, Fengyi, Deron Liang, and Wing-Sang Chu. 2010. The role of non-financial features related to corporate governance in business crisis prediction. *Journal of Marine Science and Technology* 18: 504–13. [CrossRef]

Lin, Fengyi, Deron Liang, Ching-Chiang Yeh, and Jiu-Chieh Huang. 2014. Novel feature selection methods to financial distress prediction. *Expert Systems with Applications* 41: 2472–81. [CrossRef]

Lin, Lin, and Jenifer Piesse. 2004. Identification of corporate distress in UK industrials: A conditional probability analysis approach. *Applied Financial Economics* 14: 73–82. [CrossRef]

Lin, Lin. 1999. Does Takeover Help Distressed Acquirers to Escape from Bankruptcy? Some Lessons from the UK Industrial Sector. Ph.D. thesis, University of London, London, UK.

Liu, Weihua, Shan Zeng, Guiju Wu, Hao Li, and Feifei Chen. 2021. Rice Seed Purity Identification Technology Using Hyperspectral Image with LASSO Logistic Regression Model. *Sensors* 21: 4384. [CrossRef] [PubMed]

Liu, Xiao-Ying, Yong Liang, Sai Wang, Zi-Yi Yang, and Han-Shuo Ye. 2018. A Hybrid Genetic Algorithm With Wrapper-Embedded Approaches for Feature Selection. *IEEE Access* 6: 22863–74. [CrossRef]

Lukason, Oliver, and María-del-Mar Camacho-Miñano. 2019. Bankruptcy Risk, Its Financial Determinants and Reporting Delays: Do Managers Have Anything to Hide? *Risks* 7: 77. [CrossRef]

Memić, Deni. 2015. Assessing credit default using logistic regression and multiple discriminant analysis: Empirical evidence from Bosnia and Herzegovina. *Interdisciplinary Description of Complex Systems* 13: 128–53. [CrossRef]

Messier, William F., Jr., and James V. Hansen. 1988. Inducing rules for expert system development: An example using default and bankruptcy data. *Management Science* 34: 1412–5. [CrossRef]

Mihalovic, Matus. 2016. Performance comparison of multiple discriminant analysis and logit models in bankruptcy prediction. *Economics & Sociology* 9: 101–18. [CrossRef]

Min, Joe H., and Young-Chan Lee. 2008. A practical approach to credit scoring. *Expert Systems with Applications* 35: 1762–70. [CrossRef]

Moraes, Caroline P. A., Denis G. Fantinato, and Aline Neves. 2021. Epanechnikov Kernel for pdf Estimation Applied to Equlization and Blind Source Separation. *Signal Processing* 189: 108251. [CrossRef]

Mousavi, Muhammad M., Jamal Ouenniche, and Kaoru Tone. 2023. A dynamic performance evaluation of distress prediction models. *Journal of Forecasting* 42: 756–84. [CrossRef]

MTSR. 2019. *Ročenka slovenského Stavebníctva 2019 [Yearbook of the Slovak Construction 2019]*. Bratislava: Ministry of Transport of the Slovak Republic. Available online: https://www.mindop.sk/ministerstvo-1/vystavba-5/stavebnictvo/dokumenty-a-materialy/rocenky-stavebnictva (accessed on 22 March 2023).

Muthukrishnan, Ramakrishnan, and R. Rohini. 2016. LASSO: A feature selection technique in predictive modeling for machine learning. Paper presented at 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, India, October 24.

Norton, Curtis L., and Ralph E. Smith. 1979. A comparison of general price level and historical cost financial statements in the prediction of bankruptcy. *The Accounting Review* 54: 72–87. Available online: https://www.jstor.org/stable/246235 (accessed on 15 July 2023).

Nurcan, Ebru, and Can Deniz Köksal. 2021. Determination of Financial Failure Indicators by Gray Relational Analysis and Application of Data Envelopment Analysis and Logistic Regression Analysis in BIST 100 Index. *Iranian Journal of Management Studies* 14: 163–87. [CrossRef]

Odom, Marcus D., and Ramesh Sharda. 1990. A neural network model for bankruptcy prediction. Paper presented at 1990 IJCNN International Joint Conference on Neural Networks, San Diego, CA, USA, June 17–21.

Ohlson, James A. 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research* 18: 109–31. [CrossRef]

Olson, David L., Dursun Delen, and Yanyan Meng. 2012. Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems* 52: 464–73. [CrossRef]

Ouenniche, Jamal, and Kaoru Tone. 2017. An out-of-sample evaluation framework for DEA with application in bankruptcy prediction. *Annals of Operations Research* 254: 235–50. [CrossRef]

Paraschiv, Florentina, Markus Schmid, and Ranik Raaen Wahlstrøm. 2021. Bankruptcy Prediction of Privately Held SMEs Using Feature Selection Methods. *SSRN Electronic Journal*, 1–64. [CrossRef]

Pavlicko, Michal, and Jaroslav Mazanec. 2022. Minimalistic Logit Model as an Effective Tool for Predicting the Risk of Financial Distress in the Visegrad Group. *Mathematics* 10: 1302. [CrossRef]

Platt, Harlan D., and Marjorie B. Platt. 2006. Understanding differences between financial distress and bankruptcy. *Review of Applied Economics* 2: 141–57. Available online: https://ageconsearch.umn.edu/record/50146/?ln=en (accessed on 20 March 2023).

Platt, Harlan D., Marjorie B. Platt, and Guangli Chen. 1995. Sustainable Growth Rate of Firms In Financial Distress. *Journal of Economics and Finance* 19: 147–51. [CrossRef]

Premachandra, I. M., Gurmeet S. Bhabra, and Toshiyuki Sueyoshi. 2009. DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research* 193: 412–24. [CrossRef]

Premachandra, I. M., Yao Chen, and Jon Watson. 2011. DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment. *Omega* 3: 620–6. [CrossRef]

Produit, Timothée, Nicolas Lachance-Bernard, Emanuele Strano, Sergio Porta, and Stéphane Joost. 2010. A network based Kernel density estimator applied to Barcelona economic activities. Paper presented at 2010 International conference on Computational Science and its Applications (ICCSA)-Volume Part I, New York, NY, USA, March 23–26.

PS Stavby. 2021. Rok 2021–aký bol pre nás (a pre stavebníctvo)? [The Year 2021-How It Was for Us (and for the Construction Industry?)]. Available online: https://www.psgroup.sk/psstavby/hodnotenie-roka-2021-ps-stavby (accessed on 20 February 2023).

Psillaki, Maria, Ioannis E. Tsolas, and Dimitris Margaritis. 2010. Evaluation of credit risk based on firm performance. *European Journal of Operational Research* 201: 873–81. [CrossRef]

Rabaca, Viera, José M. Pereira, and Mário Basto. 2023. Logit Ridge and Lasso in predicting business failure. *Global Journal of Accounting and Economy Research* 4: 33–46. [CrossRef]

Ratanamahatana, Chotirat A., and Dimitrios Gunopulos. 2003. Feature selection for the naive bayesian classifier using decision trees. *Applied Artificial Intelligence* 17: 475–87. [CrossRef]

Reznakova, Mária, and Michal Karas. 2014. Bankruptcy Prediction Models: Can the Prediction Power of the Models be Improved by Using Dynamic Indicators? *Procedia Economics and Finance* 12: 565–74. [CrossRef]

Sarkar, Sumit, and Ram S. Sriram. 2001. Bayesian models for early warning of bank failures. *Management Science* 47: 1457–75. [CrossRef]

Scott, David W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: Wiley.

Seiford, Lawrence M., and Joe Zhu. 2002. Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research* 142: 16–20. [CrossRef]

Silva Portela, Maria Conceição A., Emmanuel Thanassoulis, and Gary Simpson. 2004. Negative data in DEA: A directional distance approach applied to bank branches. *Journal of the Operational Research Society* 55: 1111–21. [CrossRef]

Simak, Paul C. 1997. DEA Based Analysis of Corporate Failure. Master's thesis, University of Toronto, Toronto, ON, Canada.

Snircova, Jana. 1997. Ways to prognosticate financial situation of Slovak businesses. *BIATEC* 5: 15–22. Available online: https://nbs.sk/_img/documents/_publik_nbs_fsr/biatec/rok1997/biatec_4_1997.pdf (accessed on 2 April 2023).

Stankova, Michaela, and David Hampel. 2018. Bankruptcy Prediction of Engineering Companies in the EU Using Classification Methods. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* 66: 1347–56. [CrossRef]

Stankova, Michaela, and David Hampel. 2023. Optimal threshold of data envelopment analysis in bankruptcy prediction. *SORT-Statistics and Operations Research Transactions* 47: 129–50. [CrossRef]

Stefko, Robert, Jarmila Horvathova, and Martina Mokrisova. 2020. Bankruptcy prediction with the use of data envelopment analysis: An empirical study of Slovak businesses. *Journal of Risk and Financial Management* 13: 212. [CrossRef]

Stefko, Robert, Jarmila Horvathova, and Martina Mokrisova. 2021. The Application of Graphic Methods and the DEA in Predicting the Risk of Bankruptcy. *Journal of Risk and Financial Management* 14: 220. [CrossRef]

Subasi, Abdulhamit, and M. Ismail Gursoy. 2010. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications* 37: 8659–66. [CrossRef]

Sueyoshi, Toshiyuki, and Mika Goto. 2009. DEA-DA for bankruptcy-based performance assessment: Misclassification analysis of Japanese construction industry. *European Journal of Operational Research* 199: 576–94. [CrossRef]

Szilagyi, Mikuláš. 2004. Analýza finančného zdravia z pohľadu znalca [Analysis of financial health from an expert's point of view]. *Soudní inženýrství [Forensic Engineering]* 15: 239–40. Available online: http://www.sinz.cz/archiv/docs/si-2004-04-239-240.pdf (accessed on 4 April 2023).

Taffler, Richard J. 1983. The assessment of company solvency and performance using a statistical model. *Accounting and Business Research* 13: 295–308. [CrossRef]

Tang, Xueying, Zhi Wang, Qiwei He, Jingchen Liu, and Zhiliang Ying. 2020. Latent Feature Extraction for Process Data via Multidimensional Scaling. *Psychometrika* 85: 378–97. [CrossRef]

Tseng, Fang-Mei, and Yi-Chung Hu. 2010. Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks. *Expert Systems with Applications* 37: 1846–53. [CrossRef]

Valaskova, Katarina, Lucia Svabova, and Marek Durica. 2017. Verifikácia predikčných modelov v podmienkach slovenského poľnohospodárskeho sektora [Verification of prediction models in the conditions of the Slovak agricultural sector]. *Ekonomika Management Inovace [Economics Management Innovations]* 9: 30–38. Available online: http://emijournal.cz/wp-content/uploads/2017/12/03_verifikacia-predik%C4%8Dn%C3%BDch-modelov.pdf (accessed on 15 April 2023).

Veganzones, David, and Eric Severin. 2021. Corporate failure prediction models in the twenty-first century: A review. *European Business Review* 33: 204–26. [CrossRef]

Voda, Alina D., Gabriela Dobrotă, Diana Mihaela Țîrcă, Dănuț Dumitru Dumitrașcu, and Dan Dobrotă. 2021. Corporate bankruptcy and insolvency prediction model. *Technological and Economic Development of Economy* 27: 1039–56. [CrossRef]

Wang, Lipo, Yaoli Wang, and Qing Chang. 2016. Feature selection methods for big data bioinformatics: A survey from the search perspective. *Methods* 111: 21–31. [CrossRef] [PubMed]

Wang, Lu, and Chong Wu. 2017. Business failure prediction based on two-stage selective ensemble with manifold learning algorithm and kernel-based fuzzy self-organizing map. *Knowledge-Based Systems* 121: 99–110. [CrossRef]

Wu, Tong T., Yi F. Chen, Trevor Hastie, Eric Sobel, and Kenneth Lange. 2009. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* 25: 714–21. [CrossRef] [PubMed]

Xiao, Zhi, Xianglei Yang, Ying Pang, and Xin Dang. 2012. The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster–Shafer evidence theory. *Knowledge-Based Systems* 26: 196–206. [CrossRef]

Yin, Jingjing, and Lili Tian. 2014. Joint Inference about Sensitivity and Specificity at Optimal Cut-Off Point Associated with Youden Index. *Computational Statistics and Data Analysis* 77: 1–13. [CrossRef]

Youden, William J. 1950. Index for Rating Diagnostic Tests. *Cancer* 3: 32–5. [CrossRef]

Zavgren, Christine V. 1985. Assessing the Vulnerability to Failure of American Industrial Firms: A logistic Analysis. *Journal of Business Finance & Accounting* 12: 19–45. [CrossRef]

Zhao, Peng, and Bin Yu. 2006. On model selection consistency of Lasso. *Journal of Machine Learning Research* 7: 2541–63. Available online: https://www.jmlr.org/papers/volume7/zhao06a/zhao06a.pdf (accessed on 3 June 2023).

Zhou, Ligang, Dong Lu, and Hamido Fujita. 2015. The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches. *Knowledge-Based Systems* 85: 52–61. [CrossRef]

Zhou, Ligang, Kin K. Lai, and Jerome Yen. 2012. Empirical models based on features ranking techniques for corporate financial distress prediction. *Computers & Mathematics with Applications* 64: 2484–96. [CrossRef]

Zhu, Joe. 2023. *DEA Frontier Software*. Worcester: Foisie Business School.

Zmijewski, Mark E. 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research* 22: 59–82. [CrossRef]

Zvarikova, Katarina, Erika Spuchlakova, and Gabriela Sopkova. 2017. International comparison of the relevant variables in the chosen bankruptcy models used in the risk management. *Oeconomia Copernicana* 8: 145–57. [CrossRef]