Krakow Review of Economics and Management 2 (1012)
Zeszyty Naukowe Uniwersytet Ekonomiczny w Krakowie 2026

KREM, 2026, 2(1012): 43–62
ISSN 1898-6447
e-ISSN 2545-3238
https://doi.org/10.15678/krem.18776

Corporate Bankruptcy Forecasting before and after the COVID-19 Pandemic in Poland

Sergiusz Herman[1], **Bartłomiej Lach**[2]

[1] Poznań University of Economics and Business, Institute of Informatics and Quantitative Economics, Department of Econometrics, al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: sergiusz.herman@ue.poznan.pl, ORCID: https://orcid.org/0000-0002-2753-1982

[2] Analyx® sp. z o.o. sp.k., Krysiewicza 2, 61-887 Poznań, Poland, e-mail: lach.bartlomiej@gmail.com, ORCID: https://orcid.org/0000-0002-2023-0378

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 License (CC BY 4.0); https://creativecommons.org/licenses/by/4.0/

Suggested citation: Herman, S., & Lach, B. (2026). Corporate Bankruptcy Forecasting before and after the COVID-19 Pandemic in Poland. *Krakow Review of Economics and Management / Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, 2(1012), 43–62. https://doi.org/10.15678/krem.18776

ABSTRACT

Objective: The main objective of the study is to analyse the impact of the economic crisis caused by the COVID-19 pandemic on the forecasting of corporate bankruptcy in Poland. The analysis verifies the accuracy of models used to make forecasts and examines the determinants of corporate bankruptcy before and after the pandemic.

Research Design & Methods: The study used financial data from 121,000 companies for the 2015–2022 period. Five variable selection methods, eight classification methods and 1,000 different random learning and testing samples were used to perform the study.

Findings: The results indicate different determinants of corporate bankruptcy before and after the outbreak of the COVID-19 pandemic. Models constructed and tested before the pandemic had lower classification accuracy than models constructed and tested after the outbreak.

Implications/Recommendations: The results confirm the need to reconstruct and test models for forecasting corporate bankruptcy during periods of dynamic changes in capital markets, such as those triggered by the COVID-19 pandemic.

Contribution: The considerations presented in the article deepen knowledge of the impact of the economic crisis on the forecasting of corporate bankruptcy in Poland. To date, no research has been conducted with such a wide range of research methodology used in the context of the crisis caused by the COVID-19 pandemic.

Article type: original article.

Keywords: bankruptcy forecasting, crisis, COVID-19, machine learning.

JEL Classification: C38, C53, G33.

1. Introduction

The COVID-19 pandemic affected almost the entire world. According to the WHO, it contributed to more than 775 million illnesses and more than 7 million deaths worldwide (WHO, 2024). In response to its high dynamics, governments in most countries decided to impose numerous restrictions. Their goal was to limit the spread of the disease by limiting human contact. Numerous studies confirm that this goal was achieved, as indicated by a reduction in the number of infected people and deaths during this period (Cho, 2020; Lai *et al*., 2020; Mitze *et al*., 2020; Chernozhukov, Kasahara & Schrimpf, 2021). In addition to the positive impact of lockdowns on human health, the literature also contains many studies emphasising the negative impact of the introduced restrictions. Restrictions on the movement of people, resulting in reduced demand, disruptions in supply chains, and bans on business in many sectors of the economy contributed to the economic crisis. There are studies whose authors present the impact of lockdowns on the health of the entire economy (Coccia, 2020; Kanitkar, 2020; Ke & Hsiao, 2022). There are also a number of studies on the impact of the pandemic and related economic crisis on the financial health of companies. Thus, for example, Gopalakrishnan, Jacob, and Mohapatra (2022) indicate that COVID-19 and related lockdowns contributed to an increase in corporate indebtedness and that the propensity to incur debt depended on the severity of the restrictions imposed in the country. On the other hand, Shen *et al*. (2020) and Makni (2023) argue in their study that the COVID-19 pandemic significantly contributed to a decline in both investment and corporate earnings. Qin *et al*. (2020) indicate that the prevailing global situation caused an increase in the level of cash at the disposal of companies, while Ke (2022) indicates in his analysis that companies experienced an increase in the cost of equity capital during the period under review.

In order to mitigate the negative impact of the economic crisis on the financial health of companies, the governments of many countries decided to intervene by providing financial support to companies. Its absence would result in a drastic increase in the number of bankruptcies (Gourinchas *et al*., 2020). The literature emphasises that the economic crisis resulting from the pandemic is also associated with dynamic legislative changes, changes in financial reporting rules, or more intensive use of earnings management tools by companies. All this means that the effectiveness of models predicting corporate bankruptcy may change. Research on the impact of the economic crisis on the quality of forecasting models does not allow drawing clear conclusions about the direction of these changes.

Thus, one can find studies whose authors indicate that the economic crisis has no effect on the effectiveness of models predicting corporate bankruptcy. Nam and Jinn (2000), using a logit model and a sample of 142 companies, examined the impact of the Asian crisis on the process of forecasting corporate bankruptcy. The constructed model and the obtained results allowed them to conclude that the economic crisis did not have a direct impact on the probability of bankruptcy. The companies against which bankruptcy was declared had been characterised by poor financial condition long before the crisis. Thus, on this basis, it can be assumed that the currency crisis itself was also the result of the long-term poor financial condition of Korean companies.

At the same time, it is possible to identify in the literature only a few studies, where the results indicate that the accuracy of forecasting models is higher after the outbreak of the economic crisis. Thus, Almamy, Aston and Ngwa (2016), using a discriminant model and a sample of 1,090 companies from the UK, compared the accuracy of forecasts made before, during and after the global financial crisis (2007–2008). The results showed that the accuracy of forecasts made before the crisis was 15.1 to 15.9 percentage points (p.p.) worse than the results obtained during the crisis and 17.1 to 20 p.p. worse than the results obtained after the economic crisis. Similar results were also obtained by Liu *et al*. (2022). Like their predecessors, they also compared the quality of models predicting bankruptcy before and after the onset of the global financial crisis. The objects of the study were companies from seven countries in Western Europe. Seven machine-learning methods and three sampling methods were used. The quality of the models turned out to be higher for those constructed after the outbreak of the crisis than those built during the crisis.

The issue of the impact of the economic crisis on forecasting corporate bankruptcy was also addressed by Papík and Papíková (2023), who studied the COVID-19 pandemic period. They formulated conclusions contrary to the previously presented studies. The study was based on financial data of more than 90,000 companies operating in Slovakia from 2015–2019. Three classification methods were used: CatBoost, LightGBM and XGBoost. When constructing the models, it was decided to distinguish two pre-crisis periods, i.e., before the outbreak of the COVID-19 pandemic (2018–2019) and the crisis period (2020). The obtained results confirm that the quality of the forecasting models was lower for the economic crisis period. This was particularly evident for forecasts with a one-year horizon.

In the literature, one can find many studies whose authors attempt to forecast the bankruptcy of companies in Poland. Various classification methods are used for this purpose and the constructed models concern both the whole economy and its individual industries (Hadasik, 1998; Korol & Prusak, 2005; Hołda, 2006; Pociecha *et al*., 2014; Herman & Lach, 2024). There are also studies whose authors attempt to identify the main causes of bankruptcies in Poland. In light of the literature and their own research, Hołda and Strojny (2019) indicate that the causes of business bankruptcies in Poland have changed with socio-economic changes. At the beginning of the transformation, the predominant causes were mainly external, resulting from the macroeconomic situation of the country, e.g., market downturn, increasing domestic and foreign competition or a decline in demand. Over the years, however, the situation has changed. The authors point out that the causes of bankruptcy today should be found primarily within the organisation. They indicate as the main reason for bankruptcy the weakness of management, which results, for example, in overinvestment, poor company strategy, inadequate cooperation with the contractor.

The primary purpose of the study was to analyse the impact of the economic crisis caused by the COVID-19 pandemic on the forecasting of corporate bankruptcy in Poland. As indicated above, there are few studies related to the issue at hand and only one of them deals with the crisis caused by the SARS-CoV-2 virus. The conclusions of the described analyses are not unequivocal. In the conducted analysis, the accuracy of models which were used to make predictions constructed and tested both before and after the outbreak of the COVID-19 pandemic was verified. Unlike the studies presented above, the analysis considered and compared the quality of eight machine-learning methods, while the comparison itself was made on the basis of 1,000 random learning and testing samples. The use of such an approach makes the obtained results robust – independent of the methodology and test sample. The second objective of the study is to compare the determinants of company bankruptcy in the periods in question. In contrast to the study by Papík and Papíková (2023) five methods of variable selection were used for this purpose. The results of the study also allowed answering the question of which classification methods – ensemble or individual – proved more effective in predicting the bankruptcy of companies before and after the outbreak of the COVID-19 pandemic. This issue was not addressed in the studies presented earlier.

2. Methodology

Conducting an empirical study required gathering an adequate research sample of enterprises against which the relevant court had issued a decision to declare corporate bankruptcy (denoted as "bankrupts" in the study) and enterprises operating in the economy in good financial condition (hereinafter denoted as "healthy"). Two data sources were used to obtain information on enterprises that were declared bankrupt during the period under study: The Official Journal of the Court and Economic Monitor and the public portal National Debt Register. In this way, it was possible to draw 1,000 enterprises against which bankruptcy was declared – 500 against which bankruptcy was declared in the 2016–2019 period (before the outbreak of the pandemic) and 500 where the court decision took place in the 2020–2023 period (after the outbreak of the pandemic). After removing observations for which there were data gaps and outliers, the study was left with 946 bankrupt companies. The financial retrieved statements were always for the year preceding the court's decision to declare bankruptcy. In addition, for the purpose of empirical analysis, a database of 120,000 financial statements of companies that efficiently operated in the market during the period under study was collected. Table 1 shows the number of financial statements included in the study from consecutive years of analysis.

Table 1. Number of Bankruptcy and Healthy Companies per Year

Year of Financial Statements	Bankrupts	Healthy
2015	99	2,523
2016	115	6,824
2017	146	24,104
2018	125	26,549
2019	152	10,947
2020	112	16,288
2021	95	19,482
2022	102	13,283

Source: own study.

Table 2, in turn, presents information on the size of assets and sales revenue of the companies in the research sample.

Table 2. Number of Companies Included in the Survey According to the Size of Assets and Sales Revenue

Range (PLN Million)	Size of Assets	Sales Revenue
0–9.99	82,673	77,321
10–19.99	13,075	15,486
20–29.99	8,135	7,298

Table 2 cnt'd

Range (PLN Million)	Size of Assets	Sales Revenue
30–39.99	4,928	4,422
40–49.99	1,777	2,922
50–59.99	1,253	2,143
60–69.99	928	1,513
70–79.99	766	1,081
80–89.99	604	909
90–99.99	490	732
> 100	6,317	7,119

Source: own study.

Assets of 68% of the surveyed companies, while sales revenues of 64% of the entities did not exceed the size of PLN 10 million.

Based on the constructed dataset, the values of 21 financial ratios were calculated representing the basic areas of companies' activities, i.e., profitability, operating efficiency, liquidity and capital structure (Table 3).

Table 3. Financial Ratios Used in the Study

Category	Financial Ratio	Formula
Profitability	Return on assets (ROA) (%)	Net profit/assets
	Return on equity (ROE) (%)	Net profit/total equity
	Return on sales (ROS) (%)	Net profit/net sales
	Operating margin (%)	Operating profit/net sales
	Return on capital employed (%)	EBIT/(assets – current liabilities)
Efficiency	Inventory turnover (days)	Inventories/net sales × 365
	Trade receivable turnover (days)	Receivables/net sales × 365
	Total asset turnover	Net sales/assets
	Trade payables turnover	(Cost + inventory)/trade payables
	Working capital turnover	Net sales/(current assets – current liabilities)
Liquidity	Current ratio	Current assets/current liabilities
	Quick ratio	Current assets – inventories)/short-term liabilities
	Cash ratio	(Working assets – inventories – receivables)/short-term liabilities

Table 3 cnt'd

Category	Financial Ratio	Formula
Capital structure	Debt ratio (%)	Total liabilities/total assets
	Debt to equity ratio (%)	Total liabilities/shareholder's equity
	Cash to total assets (%)	Cash/assets
	Trade receivables to total assets (%)	Trade receivables/assets
	Inventories to total assets (%)	Inventories/assets
	Fixed assets to total assets (%)	Fixed assets/assets
	Current liabilities to total liabilities (%)	Current liabilities/total liabilities
	Equity to assets ratio (%)	Equity/assets

Source: own study.

The following five variable selection methods were used to construct business bankruptcy forecasting models:

– filter-based methods: using entropy (Peng, Long & Ding, 2005), Mann-Whitney statistics (Mann & Whitney, 1947) and the ReliefF algorithm (Kira & Rendell, 1992),

– a method based on model selection (wrappers) – a stepwise "progressive" method (Miller, 1984), which was based on a discriminant function and a criterion for improving classification accuracy,

– a method that is an integral part of the learning algorithm (embedded methods): based on the Gini index.

In the following analysis these methods are labelled as, respectively: "entropy," "Whitney," "relief," "stepwise" and "Gini." Corporate bankruptcy forecasting models were constructed based on eight classification methods. These were: four individual methods, i.e., the *k*-nearest neighbours method (Guo *et al*., 2003), the support vector machines (Cortes & Vapnik, 1995), neural networks (Rumelhart, Hinton & Williams, 1986), logistic regression (Berkson, 1944) and four ensemble methods, i.e. random forests (Breiman, 2001), the bagging algorithm (Breiman, 1996), and two gradient enhancement methods: Extreme Gradient Boosting (Chen & Guestrin, 2016) and Light Gradient-boosting Machine (Ke *et al*., 2019). These methods are denoted hereafter as "KNN," "SVM," "network," "logistic," "forest," "bagging," "XGBoost" and "LightGBM," respectively. Table 4 shows the main assumptions made using the above-mentioned methods. The study did not optimise the values of the parameters listed below as the purpose of the study is not to obtain the best possible forecasting model, but to compare the accuracy of the classification of models predicting the bankruptcy of companies before and after the outbreak of the COVID-19 pandemic.

In the study itself, each model for predicting corporate bankruptcy, both before and after the pandemic outbreak, was constructed and tested on the basis of a drawn, balanced sample, consisting of 450 bankrupt companies and 450 viable entities. The inclusion of an unbalanced sample often results in a significant deterioration in the obtained classification accuracy, especially for a smaller sample. In the case of an unbalanced sample, it could be necessary to solve this problem, for example, by means of a specific sampling method. However, its choice could, in turn, have a major impact on the results obtained in the study (Veganzones & Séverin, 2018).

Table 4. Assumptions Made for Each Classification Method

Method	Assumptions
KNN	*k* optimised using cross-validation, maximum number of nearest neighbours: 30
SVM	Kernel function: radial
Network	One hidden layer, number of neurons: the average of the number of neurons of the input and output layers
Logistic	Decision threshold: 0.5
Forest	Number of trees: 1,000, number of variables drawn: 5
Bagging	Number of iterations: 1,000
XGBoost	Number of iterations: 1,000, eta: 0.1
LightGBM	Number of iterations: 1,000, eta: 0.1

Source: own study.

The empirical study was conducted using a custom web application with a graphical user interface made in the R environment, to be used in the future for any multidimensional object classification problem.

3. Results and Analysis

At the outset of the study, it was verified whether the surveyed companies differed in terms of their financial health, described by the above-mentioned financial indicators, before and after the outbreak of the COVID-19 pandemic. The financial indicators did not follow a normal distribution. For this reason, the nonparametric Mann-Whitney test was used for comparative analysis. The corresponding results are shown in Tables 5 and 6.

Table 5. Comparative Analysis of Selected Financial Indicators for the Studied Groups before the Pandemic Outbreak

Financial Ratio	Indication	Median		*Z*-statistic	*p* value
		Bankrupts	Healthy		
Return on assets (ROA) (%)	X1	–23.64	4.63	–27.09	0.00
Return on equity (ROE) (%)	X2	23.19	16.68	1.73	0.08
Return on sales (ROS) (%)	X3	–18.00	2.58	–26.29	0.00
Operating margin (%)	X4	–14.56	3.59	–24.91	0.00
Return on capital employed (%)	X5	6.65	15.60	–3.19	0.00
Inventory turnover	X6	14.61	5.55	6.44	0.00
Trade receivable turnover	X7	39.38	38.45	1.58	0.11
Total asset turnover	X8	1.56	1.89	–2.71	0.01
Trade payables turnover	X9	0.00	0.03	–9.37	0.00
Working capital turnover	X10	–1.08	3.76	–16.06	0.00
Current ratio	X11	0.55	1.52	–23.70	0.00
Quick ratio	X12	0.37	1.20	–22.93	0.00
Cash ratio	X13	0.01	0.13	–15.36	0.00
Debt ratio (%)	X14	25.06	0.70	14.70	0.00
Debt to equity ratio (%)	X15	0.00	0.00	–10.69	0.00
Cash to total assets (%)	X16	1.44	5.44	–8.06	0.00
Trade receivables to total assets (%)	X17	18.89	20.12	–0.91	0.36
Inventories to total assets (%)	X18	5.90	2.80	4.57	0.00
Fixed assets to total assets (%)	X19	10.25	6.62	3.38	0.00
Current liabilities to total liabilities (%)	X20	98.46	100.00	–1.95	0.06
Equity to assets ratio (%)	X21	–14.29	41.67	–23.46	0.00

Notes: *Z*-statistic – the value of the test statistic for the *U* Mann-Whitney test.

Source: own compilation based on data from the EMIS professional database.

Table 6. Comparative Analysis of Selected Financial Indicators for Study Groups after the Pandemic Outbreak

Financial Ratio	Indication	Median		*Z*-statistic	*p* value
		Bankrupts	Healthy		
Return on assets (ROA) (%)	X1	–24.11	7.35	–26.90	0.00
Return on equity (ROE) (%)	X2	28.58	21.56	2.85	0.00
Return on sales (ROS) (%)	X3	–22.83	3.79	–25.46	0.00
Operating margin (%)	X4	–18.58	4.79	–23.88	0.00
Return on capital employed (%)	X5	12.64	20.05	–1.96	0.05

Table 6 cnt'd

Financial Ratio	Indication	Median		*Z*-statistic	*p* value
		Bankrupts	Healthy		
Inventory turnover	X6	7.73	5.66	1.52	0.13
Trade receivable turnover	X7	37.00	35.08	2.29	0.02
Total asset turnover	X8	1.14	2.01	–8.01	0.00
Trade payables turnover	X9	0.00	0.00	–10.05	0.00
Working capital turnover	X10	–0.35	4.00	–19.32	0.00
Current ratio	X11	0.46	1.68	–26.52	0.00
Quick ratio	X12	0.31	1.31	–24.94	0.00
Cash ratio	X13	0.00	0.16	–17.75	0.00
Debt ratio (%)	X14	32.26	2.78	16.13	0.00
Debt to equity ratio (%)	X15	–14.72	0.75	–18.13	0.00
Cash to total assets (%)	X16	0.51	6.42	–11.00	0.00
Trade receivables to total assets (%)	X17	17.06	18.78	–1.24	0.21
Inventories to total assets (%)	X18	2.66	2.98	–0.54	0.59
Fixed assets to total assets (%)	X19	6.70	4.90	2.86	0.00
Current liabilities to total liabilities (%)	X20	99.63	99.62	0.58	0.56
Equity to assets ratio (%)	X21	–42.70	45.44	–27.78	0.00

Notes: *Z*-statistic – the value of the test statistic for the *U* Mann-Whitney test.

Source: own compilation based on data from the EMIS professional database.

Based on the results presented in Tables 5–6, it can be concluded that in the period before the outbreak of the pandemic, only for four financial indicators (X2, X7, X17 and X20) the surveyed companies (at a significance level of 0.05) do not differ significantly in terms of financial health. Similarly, after the outbreak of the pandemic, such a situation occurs four times – for indicators X6, X17, X18 and X20. Healthy companies are characterised in most cases by a higher value of profitability indicators and liquidity of operations than bankrupts. In the case of bankrupt entities, one can see a decrease in the value of the aforementioned indicators in the studied periods, while profitability and liquidity for healthy companies increase after the outbreak of the pandemic. Thus, the differences between the medians of the used indicators for healthy and sick companies are in most cases higher after the outbreak of the pandemic than before 2020. This may indicate that the COVID-19 pandemic and the related economic crisis caused the difference between the financial condition in the studied groups of companies to increase significantly, and the financial indicators themselves gained discriminatory power.

In order to construct models for forecasting corporate bankruptcy, the eight classification methods mentioned earlier were used. In the case of individual classifiers: the *k*-nearest neighbours, the support vector machines, neural networks and logistic regression, the variables for the models were selected using filter-based selection methods (entropy, Whitney, relief) and the stepwise method. For ensemble classifiers, i.e., random forests, the bagging algorithm and gradient enhancement methods, the methods integral to the learning algorithm and based on the Gini index were used. In this way, results were obtained for 20 combinations: a classification method – a variable selection method. This procedure is intended to test the robustness of the results of the study to the research methodology used. The construction of a classification model for the period before and after the pandemic outbreak requires the division of the available research sample into a learning sample and a test sample. In all the carried out analyses, it was determined that 1,000 times 450 economically viable (healthy) enterprises and 450 bankrupt enterprises are drawn at random, and then the objects are randomly divided with a ratio of 75:25 into a learning sample and a test sample. For each draw, a learning sample was used to select variables using the methods described earlier. Based on these financial indicators, predictive models were constructed, the quality of which was checked on the basis of test samples. Figure 1 shows the validity of each financial indicator by area of business operation. The validity was assessed based on the average frequency of selection of financial indicators into the model using the four methods of variable selection presented earlier (entropy, Whitney, relief and stepwise). Based on the results, it can be concluded that the models constructed before and after the outbreak of the COVID-19 pandemic differ in the validity of financial indicators that describe different areas of the company's operations. For all the presented methods of variable selection, the importance of profitability indicators decreases after the outbreak of the pandemic. On the other hand, in the case of 3 out of the 4 methods, the importance of indicators included in the groups responsible for operational efficiency and financial structure increases during the studied period. In the case of indicators responsible for liquidity, no clear conclusion can be drawn as to the direction of change.

The next stage of the study compared the average (for 1,000 iterations) classification accuracy obtained for models:

– constructed and tested before the pandemic outbreak (before–before),

– constructed and tested after the pandemic outbreak (after–after),

– constructed before the pandemic outbreak and tested on observations from after the pandemic outbreak (before–after).

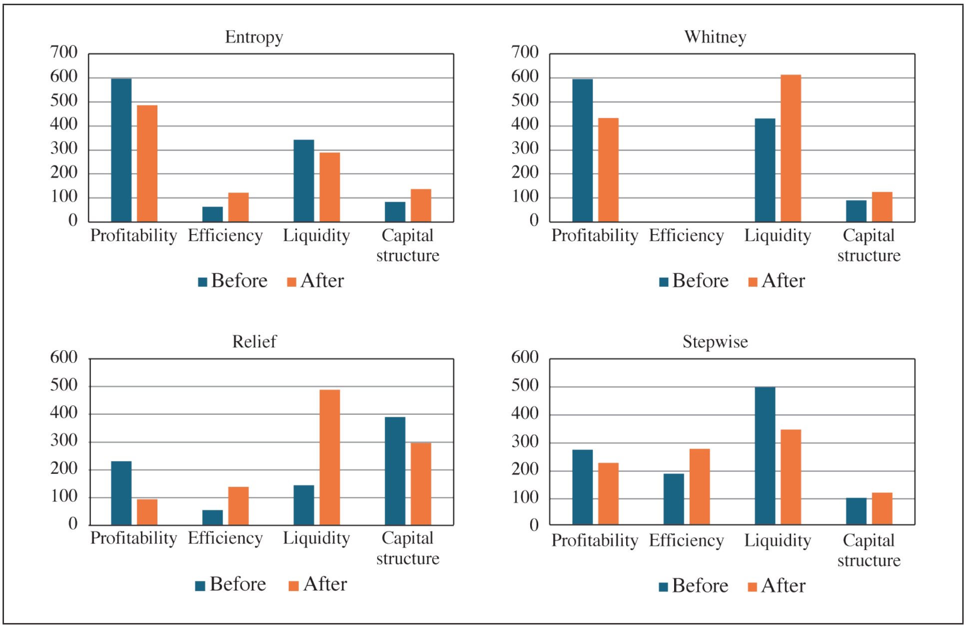

Fig. 1. Mean Frequency of Selection of Financial Indicators Representing Different Areas of Company Operations in Classification Models before and after the COVID-19 Pandemic Outbreak

Source: own compilation based on data from the EMIS professional database.

Tables 7–9 show the average value of global accuracy indices, accuracy for healthy companies, and accuracy for companies declared bankrupt.

Table 7. Average Global Classification Accuracy Obtained for Models Constructed and Tested before and after the Pandemic Outbreak

Classification Method	Variable Selection Method	Before–before (%)	After–after (%)	Before–after (%)	*H*-statistics	Statistically Significant Differences
Forest	Gini	82.7	85.8	84.9	732.8	all of them
Bagging	Gini	81.1	85.1	84.0	1,004.8	all of them
XGBoost	Gini	82.3	84.9	83.2	505.1	all of them
LightGBM	Gini	82.6	85.1	83.3	487.7	all of them
KNN	Entropy	77.6	81.0	79.8	638.1	all of them
SVM	Entropy	79.3	81.3	81.2	302.3	1–2; 1–3
Network	Entropy	79.6	83.5	82.0	852.8	all of them
Logistic	Entropy	79.0	82.1	82.0	585.6	1–2; 1–3
KNN	Whitney	78.0	81.3	80.1	616.7	all of them

Table 7 cnt'd

Classification Method	Variable Selection Method	Before–before (%)	After–after (%)	Before–after (%)	*H*-statistics	Statistically Significant Differences
SVM	Whitney	79.6	81.9	81.4	381.7	all of them
Network	Whitney	79.8	83.9	82.3	930.4	all of them
Logistic	Whitney	79.4	82.7	82.3	655.3	all of them
KNN	RelieF	72.3	72.6	74.8	143.2	all of them
SVM	RelieF	77.1	77.6	79.6	242.8	all of them
Network	RelieF	77.7	78.9	81.1	400.8	all of them
Logistic	RelieF	75.9	76.4	79.9	444.1	all of them
KNN	Stepwise	75.9	78.8	78.6	383.7	1–2; 1–3
SVM	Stepwise	77.5	76.7	80.1	218.0	all of them
Network	Stepwise	79.0	81.3	81.8	408.4	all of them
Logistic	Stepwise	77.7	79.2	80.9	283.5	all of them

Notes: *H*-statistics – the value of the test statistic for the Kruskal-Wallis test.

Source: own compilation based on data from the EMIS professional database.

Table 8. Average Classification Accuracy for Healthy Companies Obtained for Models Constructed and Tested before and after the Pandemic Outbreak

Classification Method	Variable Selection Method	Before–before (%)	After–after (%)	Before–after (%)	*H*-statistics	Statistically Significant Differences
Forest	Gini	82.3	85.0	84.5	279.1	all of them
Bagging	Gini	80.8	83.7	83.4	276.6	1–2; 1–3
XGBoost	Gini	82.2	84.9	83.9	268.4	all of them
LightGBM	Gini	82.4	85.0	84.2	259.5	all of them
KNN	Entropy	77.5	82.4	80.7	624.9	all of them
SVM	Entropy	85.7	86.2	87.8	137.6	all of them
Network	Entropy	79.0	84.1	82.0	448.5	all of them
Logistic	Entropy	83.2	87.5	85.9	336.4	all of them
KNN	Whitney	78.1	82.6	81.2	569.8	all of them
SVM	Whitney	86.2	88.0	88.2	167.3	1–2; 1–3
Network	Whitney	79.2	84.7	82.5	546.6	all of them
Logistic	Whitney	82.7	87.6	85.5	398.6	all of them
KNN	RelieF	73.6	73.1	76.4	147.7	1–3; 2–3
SVM	RelieF	81.4	76.4	84.0	525.8	all of them
Network	RelieF	76.6	79.7	80.0	160.5	1–2; 1–3
Logistic	RelieF	79.5	72.0	82.3	424.5	all of them

Table 8 cnt'd

Classification Method	Variable Selection Method	Before–before (%)	After–after (%)	Before–after (%)	*H*-statistics	Statistically Significant Differences
KNN	Stepwise	75.9	80.2	79.4	424.2	all of them
SVM	Stepwise	80.3	72.1	83.1	293.1	all of them
Network	Stepwise	78.3	81.1	81.7	183.3	all of them
Logistic	Stepwise	78.9	77.8	82.0	93.6	1–3; 2–3

Notes: *H*-statistics – the value of the test statistic for the Kruskal-Wallis test.

Source: own compilation based on data from the EMIS professional database.

Table 9. Average Classification Accuracy for Companies Declared Bankrupt Obtained for Models Constructed and Tested before and after the Pandemic Outbreak

Classification Method	Variable Selection Method	Before–before (%)	After–after (%)	Before–after (%)	*H*-statistics	Statistically Significant Differences
Forest	Gini	83.1	86.6	85.3	449.5	all of them
Bagging	Gini	81.3	86.5	84.5	738.4	all of them
XGBoost	Gini	82.4	84.8	82.4	265.1	1–2; 2–3
LightGBM	Gini	82.7	85.1	82.4	300.2	1–2; 2–3
KNN	Entropy	77.7	79.6	78.9	105.1	all of them
SVM	Entropy	72.9	76.3	74.6	228.6	all of them
Network	Entropy	80.2	83.0	81.9	145.9	all of them
Logistic	Entropy	74.8	76.7	78.1	160.0	all of them
KNN	Whitney	78.0	80.0	79.0	103.5	all of them
SVM	Whitney	73.0	75.9	74.6	179.6	all of them
Network	Whitney	80.5	83.2	82.1	140.2	all of them
Logistic	Whitney	76.1	77.8	79.1	196.1	all of them
KNN	RelieF	70.9	72.0	73.3	77.5	all of them
SVM	RelieF	72.8	78.8	75.2	467.9	all of them
Network	RelieF	78.8	78.2	82.3	163.7	1–3; 2–3
Logistic	RelieF	72.4	80.9	77.5	579.0	all of them
KNN	Stepwise	75.9	77.3	77.9	71.3	1–2; 1–3
SVM	Stepwise	74.7	81.3	77.0	415.9	all of them
Network	Stepwise	79.8	81.5	81.8	62.7	1–2; 1–3
Logistic	Stepwise	76.5	80.5	79.7	221.2	all of them

Notes: *H*-statistics – the value of the test statistic for the Kruskal-Wallis test.

Source: own compilation based on data from the EMIS professional database.

As can be seen in Table 7, for each (except SVM + stepwise) method used, the average global classification accuracy is higher for models constructed and tested after the pandemic outbreak than for models constructed and tested before the COVID-19 pandemic. These differences are always statistically significant. The maximum is as high as 4.1 p.p. Similarly, for models constructed before the pandemic outbreak and tested afterwards, the classification accuracy is always higher than when they were tested before the analysed event.

Analysing separately the accuracy of the classification of efficient enterprises (Table 8) and those declared bankrupt (Table 9), it should be noted that the situation is similar. This time, however, it is clear that the values in question are more differentiated. In the case of enterprises against which bankruptcy was declared, the average difference between the classification accuracy of models built and constructed after the outbreak of the pandemic and those for which construction and testing takes place before this event is 3.1 p.p. In the case of smoothly operating companies, the difference in question is lower and amounts to only 1.5 p.p. On this basis, it can be concluded that the different performance of forecasting models built before and after the outbreak of the pandemic is much more due to the difference in the accuracy of forecasting bankrupt companies.

The analysis of Tables 7–9 also allows us to draw interesting conclusions about the comparison of the classification quality of the used research methods. In the case of the level of average global relevance and the relevance of enterprises against which bankruptcy was declared, the ensemble methods: random forest, XGBoost, LightGBM, and the bagging algorithm consistently occupy the top four places. In the case of the classification of fit enterprises, support vector machines and logistic regression dominate.

4. Discussion

Two main conclusions can be drawn from the study. The first concerns the determinants of corporate bankruptcy in Poland. After the outbreak of the pandemic, profitability indicators lose importance, while financial indicators describing the area of operational efficiency and financial structure of companies gain. The results are similar to those obtained by Papík and Papíková (2023), who also indicated that the economic crisis causes a decline in the importance of profitability indicators and a much greater importance of leverage ratios. The second conclusion concerns the predictive ability of the models. For the most part, models constructed and tested before the pandemic outbreak have worse classification accuracy than those built on the same sample but tested after the pandemic outbreak and those constructed and tested after the pandemic outbreak. The results are analogous to those obtained by Almamy, Aston and Ngwa (2016) and Liu *et al*. (2022). They found that the accuracy of forecasting models is higher after an economic crisis outbreak as well.

The results of the study differ significantly from those obtained by Nam and Jinn (2000) and Papík and Papíková (2023), who indicated that the economic crisis has no effect or a positive effect on the quality of the forecasts made.

Bankruptcy forecasting models are crucial from the perspective of many stakeholders. Correct and early enough detection of companies' financial problems is crucial from the point of view of investors and creditors, who can thus reduce their losses (Srebro *et al*., 2021). As mentioned earlier, among the most frequently cited causes of corporate bankruptcy in Poland in recent years is poor management. An effective forecasting model can be a very helpful tool for enterprise managers. Managers who are alerted to impending problems could intervene well in advance and implement appropriate corrective measures in the enterprise. The results obtained in the study indicate that an effective forecasting model requires that the existence of an economic crisis be taken into account in its construction. As a result, the re-estimated model will be constructed on the basis of a different set of financial indicators and will achieve higher classification accuracy.

The first limitation of the study conducted was the period from which the financial statements of the surveyed companies originated. The horizon of the study was divided into two periods: before the outbreak of the pandemic and the related economic crisis, as well as after the outbreak. It was valuable to investigate what would happen to the models for forecasting corporate bankruptcy after the crisis resulting from the pandemic, how long the observed regularities would be present in the models for forecasting corporate bankruptcy. This is impossible due to the macroeconomic environment of Polish companies, resulting, among other things, from Russia's triggering of a full-scale war in Ukraine. Its effect was, among other consequences, a very high level of inflation and interest rates, which affected all companies severely.

The second limitation of the survey is the drawn research sample. The companies surveyed operated in various, often very different, industries of the economy. It would be worthwhile again, with a more extensive research sample, to conduct a similar survey separately, for individual industries of the economy. Taking into account industry specifics would make the observed relationships very interesting and valuable from the point of view of the survey recipients. The survey, as mentioned in the description of the research sample, focused on small businesses. It would be interesting to perform the analysis exclusively for large companies operating in the market. It seems that larger companies, by their established position in the market, are less exposed to macroeconomic changes. Thus, it would seem interesting to ask whether, in fact, for such entities, the observed correlations regarding bankruptcy forecasting would be repeated.

5. Conclusions

An empirical study was conducted based on a sample of 1,000 enterprises against which bankruptcy was declared in 2016–2023 and 120,000 efficient enterprises. The primary objective of the study was to analyse the impact of the economic crisis caused by the COVID-19 pandemic on the forecasting of corporate bankruptcy in Poland. To achieve it, five variable selection methods, eight classification methods and 1,000 random learning and testing samples were used.

The results of the study allow us to conclude that the determinants of corporate bankruptcy are different in the two studied periods. Taking into account the obtained results, it can be concluded that the accuracy of models for forecasting corporate bankruptcy differs depending on whether they were constructed and tested before or after the outbreak of the pandemic. The greater predictive ability of the models after the pandemic outbreak may be due to the fact that during this period, the difference between the financial condition of companies declared bankrupt and those operating efficiently in the market was greater than it was before the pandemic. Based on the results, it can also be concluded that ensemble methods have higher global accuracy and classification accuracy for bankrupt companies than individual methods. This is observed for both models constructed on data before and after the pandemic outbreak.

Authors' Contribution

The authors' individual contribution is as follows: Sergiusz Herman – conceptualisation, data curation, formal analysis, investigation, validation, visualisation, writing – review and editing; Bartłomiej Lach – conceptualisation, data curation, formal analysis, investigation, validation, visualisation, writing – review and editing.

Conflict of Interest

The authors declare no conflict of interest.

References

Almamy, J., Aston, J., & Ngwa, L. N. (2016). An Evaluation of Altman's Z-score Using Cash Flow Ratio to Predict Corporate Failure Amid the Recent Financial Crisis: Evidence from the UK. *Journal of Corporate Finance*, *36*, 278–285. https://doi.org/10.1016/j.jcorpfin.2015.12.009

Berkson, J. (1944). Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, *39*(227), 357–365. https://doi.org/10.1080/01621459.1944.10500699

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, *24*, 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. KDD ’16. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Chernozhukov, V., Kasahara, H., & Schrimpf, P. (2021). Causal Impact of Masks, Policies, Behavior on Early COVID-19 Pandemic in the U.S. *Journal of Econometrics*, *220*(1), 23–62. https://doi.org/10.1016/j.jeconom.2020.09.003

Cho, S.-W. (Stanley). (2020). Quantifying the Impact of Nonpharmaceutical Interventions during the COVID-19 Outbreak: The Case of Sweden. *Econometrics Journal*, *23*(3), 323–344. https://doi.org/10.1093/ECTJ/UTAA025

Coccia, M. (2020). The Relation between Length of Lockdown, Numbers of Infected People and Deaths of COVID-19, and Economic Growth of Countries: Lessons Learned to Cope with Future Pandemics Similar to COVID-19 and to Constrain the Deterioration of Economic System. *Science of the Total Environment*, *775*, 145801. https://doi.org/10.1016/j.scitotenv.2021.145801

Cortes, C., & Vapnik, V. (1995). Support-vector Networks. *Machine Learning*, *20*, 273–297. https://doi.org/10.1007/BF00994018

Gopalakrishnan, B., Jacob, J., & Mohapatra, S. (2022). COVID-19 Pandemic and Debt Financing by Firms: Unravelling the Channels. *Economic Modelling*, *114*, 105929. https://doi.org/10.1016/j.econmod.2022.105929

Gourinchas, P.-O., Kalemli-Özcan, Ș., Penciakova, V., & Sander, N. (2020). *SME Failures under Large Liquidity Shocks: An Application to the COVID-19 Crisis* (NBER Working Paper No. 27877). National Bureau of Economic Research. Retrieved from: https://www.nber.org/papers/w27877.pdf (accessed: 5.05.2024).

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-based Approach in Classification. In: R. Meersman, Z. Tari, D. C. Schmidt (Eds), *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science* (Vol. 2888). Springer. https://doi.org/10.1007/978-3-540-39964-3_62

Hadasik, D. (1998). *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*. Wydawnictwo Akademii Ekonomicznej w Poznaniu.

Herman, S., & Lach, B. (2024). Czy warto budować modele prognozowania upadłości przedsiębiorstw uwzględniające kryterium przynależności sektorowej? *Wiadomości Statystyczne. The Polish Statistician*, *69*(10), 41–58. https://doi.org/10.59139/ws.2024.10.3

Hołda, A. (2006). *Zasada kontynuacji działalności i prognozowanie upadłości w polskich realiach gospodarczych*. Wydawnictwo Akademii Ekonomicznej w Krakowie.

Hołda, A., & Strojny, K. (2019). Determinants of Entrepreneurs’ Bankruptcy in Poland – Systematics and Comparative Analysis of the Opinion of Management and Trustees. *Zeszyty Teoretyczne Rachunkowości*, *101*(157), 9–34. https://doi.org/10.5604/01.3001.0013.0753

Kanitkar, T. (2020). The COVID-19 Lockdown in India: Impacts on the Economy and the Power Sector. *Global Transitions*, *2*, 150–156. https://doi.org/10.1016/j.glt.2020.07.005

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2019). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: *31st Conference on Neural Information Processing Systems* (*NIPS 2017*) (pp. 3149–3157). NIPS.

Ke, X., & Hsiao, C. (2022). Economic Impact of the Most Drastic Lockdown during COVID-19 Pandemic – the Experience of Hubei, China. *Journal of Applied Econometrics*, *37*(1), 187–209. https://doi.org/10.1002/jae.2871

Ke, Y. (2022). The Impact of COVID-19 on Firms' Cost of Equity Capital: Early Evidence from U.S. Public Firms. *Finance Research Letters*, *46*, 102242. https://doi.org/10.1016/j.frl.2021.102242

Kira, K., & Rendell, L. A. (1992). A Practical Approach to Feature Selection. In: D. Sleeman, P. Edwards (Eds), *Machine Learning Proceedings* (pp. 249–256). Morgan Kaufmann. https://doi.org/10.1016/B978-1-55860-247-2.50037-1

Korol, T., & Prusak, B. (2005). *Upadłość przedsiębiorstw a wykorzystanie sztucznej inteligencji*. CeDeWu.

Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., Wesolowski, A., Santillana, M., Zhang, C., Du, X., Yu, H., & Tatem, A. J. (2020). Effect of Non-pharmaceutical Interventions to Contain COVID-19 in China. *Nature*, *585*(7825), 410–413. https://doi.org/10.1038/s41586-020-2293-x

Liu, Y., Zeng, Q., Li, B., Ma, L., & Ordieres-Meré, J. (2022). Anticipating Financial Distress of High-tech Startups in the European Union: A Machine Learning Approach for Imbalanced Samples. *Journal of Forecasting*, *41*(6), 1131–1155. https://doi.org/10.1002/for.2852

Makni, M. S. (2023). Analyzing the Impact of COVID-19 on the Performance of Listed Firms in Saudi Market. *Technological Forecasting and Social Change*, *187*, 122171. https://doi.org/10.1016/j.techfore.2022.122171

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

Miller, A. J. (1984). Selection of Subsets of Regression Variables. *Journal of the Royal Statistical Society*. *Series A: General*, *147*(3), 389–425. https://doi.org/10.2307/2981576

Mitze, T., Kosfeld, R., Rode, J., & Wälde, K. (2020). Face Masks Considerably Reduce COVID-19 Cases in Germany. *Proceedings of the National Academy of Sciences*, *117*(51), 32293–32301. https://doi.org/10.1073/pnas.2015954117

Nam, J.-H., & Jinn, T. (2000). Bankruptcy Prediction: Evidence from Korean Listed Companies during the IMF Crisis. *Journal of International Financial Management and Accounting*, *11*(3), 178–197. https://doi.org/10.1111/1467-646X.00061

Papík, M., & Papíková, L. (2023). Impacts of Crisis on SME Bankruptcy Prediction Models' Performance. *Expert Systems with Applications*, *214*, 119072. https://doi.org/10.1016/j.eswa.2022.119072

Peng, H., Long, F., & Ding, C. (2005). Feature Selection Based on Mutual Information Criteria of Max-dependency, Max-relevance, and Min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238. https://doi.org/10.1109/TPAMI.2005.159

Pociecha, J., Pawełek, B., Baryła, M., & Augustyn, S. (2014). *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*. Fundacja Uniwersytetu Ekonomicznego w Krakowie.

Qin, X., Huang, G., Shen, H., & Fu, M. (2020). COVID-19 Pandemic and Firm-level Cash Holding – Moderating Effect of Goodwill and Goodwill Impairment. *Emerging Markets Finance and Trade*, *56*(10), 2243–2258. https://doi.org/10.1080/1540496X.2020.1785864

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning Representations by Back-propagating Errors. *Nature*, *323*, 533–536. https://doi.org/10.1038/323533a0

Shen, H., Fu, M., Pan, H., Yu, Z., & Chen, Y. (2020). The Impact of the COVID-19 Pandemic on Firm Performance. *Emerging Markets Finance and Trade*, *56*(10), 2213–2230. https://doi.org/10.1080/1540496X.2020.1785863

Srebro, B., Mavrenski, B., Bogojević Arsić, V., Knežević, S., Milašinović, M., & Travica, J. (2021). Bankruptcy Risk Prediction in Ensuring the Sustainable Operation of Agriculture Companies. *Sustainability*, *13*(14), 7712. https://doi.org/10.3390/su13147712

Veganzones, D., & Séverin, E. (2018). An Investigation of Bankruptcy Prediction in Imbalanced Datasets. *Decision Support Systems*, *112*, 111–124. https://doi.org/10.1016/j.dss.2018.06.011

WHO. (2024). *Coronavirus Disease* (*COVID-2019*) *Situation Reports*. Retrieved from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (accessed: 7.06.2024).