Risk assessment of Polish joint stock companies using machine learning methods

Wroclaw University of Economics and Business

Aleksandra Szymura

aleksandra.szymura@ue.wroc.pl

INTRODUCTION

Risk assessment plays an important role in companies functioning properly. It allows companies to make their decision-making processes more effective, investors to minimize the negative effects of their investments, and banks to check potential borrowers before lending to them. An assessment of the company can be made for several identified risks e.g. bankruptcy, insolvency, or loss of liquidity. The first mentions of business evaluation methods appeared in the literature in the first half of the 20th century, while statistical methods began to be used for this purpose in the 1960s. One of the pioneers in this field was Altman (Altman 1968), who presented a model for predicting corporate bankruptcy. He used discriminant analysis to achieve this, which was popular in the literature for the following years. The 1980s marked the growth in the use of logistic regression (Zmijewski 1984). Over time more advanced machine learning methods have been applied. The 1990s belonged to researchers who began to use neural networks (Lacher et al. 1995), and the 21st century marked the beginning of the popularity of gradient boosting methods (Pham and Ho 2021).

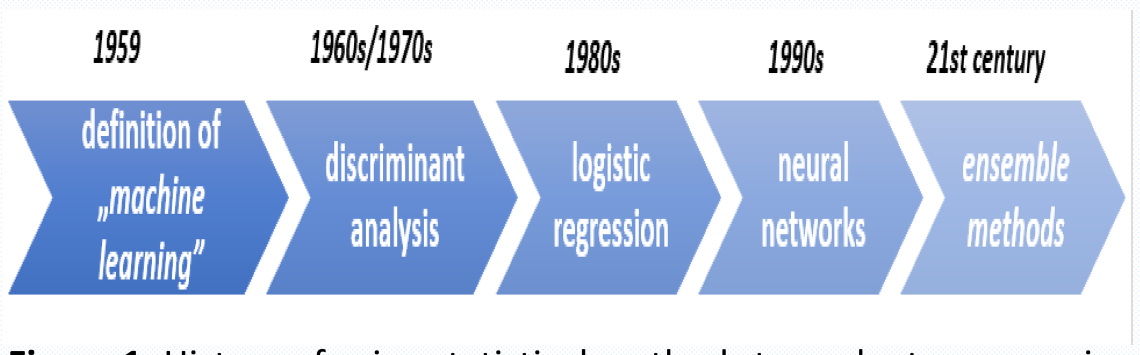

Figure 1. History of using statistical methods to evaluate companies.

METHODS

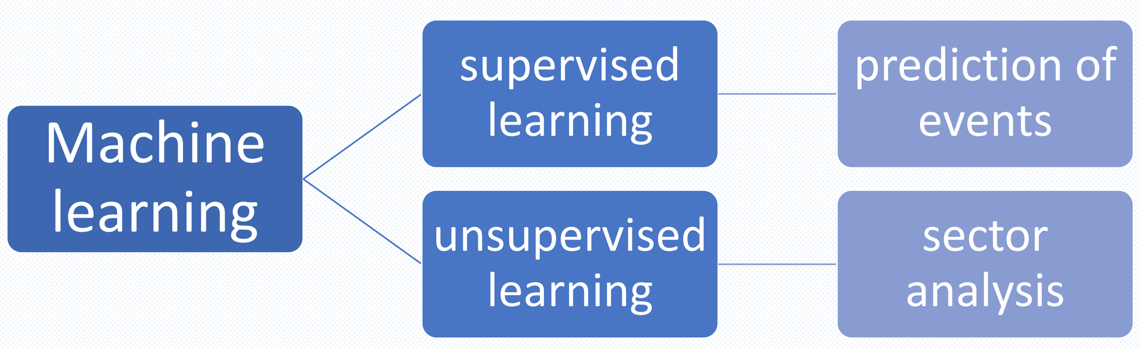

Figure 2. Methods used in research.

EXPLAINABILITY

- **global** (for liquidity risk model)

Figure 3. Example plot of SHAP values.

- **local** (for liquidity risk model)

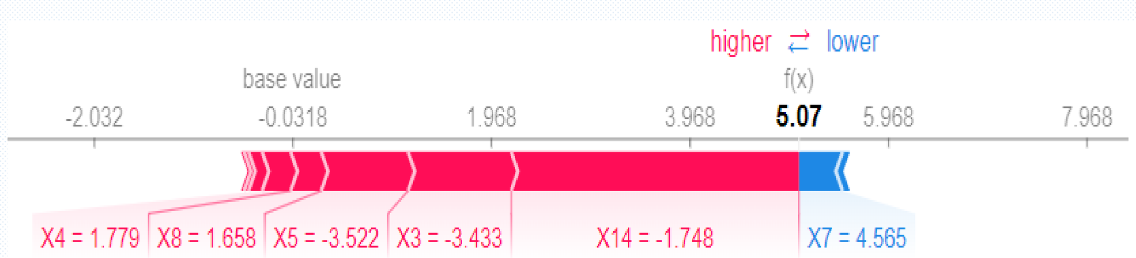

Figure 4. Example plot of SHAP values for one company.

OBJECTIVE

Presentation of the methodology to assess the risk of Polish joint stock companies.

RISK ASSESSMENT – DATASET

Independent variables: financial indicators of companies' activities from 2016-2018

Dependent variables: financial information on companies from 2017-2019

Table 1. Number of businesses included in the analysis by year.

Year	Number of Companies
2017	305
2018	311
2019	312

RISK ASSESSMENT – RESULTS

Sample stratification: by year, sales, dependent variable

Train/test split: 70/30

Class balancing method: SMOTE

Methods: classification models

- **risk of penalties or compensation payments**

Table 2. Mean value of metrics for a 10-fold process run.

Model	AUC	Cohen's Kappa
Logistic regression	0.6522	0.1903
Decision tree	0.5913	0.1767
XGBoost	0.7159	0.2754
LightGBM	0.7178	0.2716
CatBoost	0.7321	0.3027

- **liquidity risk**

Table 3. Mean value of metrics for a 10-fold process run.

Model	AUC	Cohen's Kappa
Logistic regression	0.8505	0.4017
Decision tree	0.6955	0.3447
XGBoost	0.8574	0.5011
LightGBM	0.8664	0.4969
CatBoost	0.8648	0.5015

- **risk of loss from operations**

Table 4. Mean value of metrics for a 10-fold process run.

Model	AUC	Cohen's Kappa
Logistic regression	0.8436	0.5382
Decision tree	0.6771	0.3466
XGBoost	0.8101	0.4618
LightGBM	0.8164	0.4852
CatBoost	0.8299	0.5008

RISK ASSESSMENT WITHIN SECTOR – DATASET

Variables: financial indicators of banks' activities from the 2018 and 2020 years.

Table 5. Number of banks included in the analysis by year.

Year	Number of banks
2018	16
2020	14

RISK ASSESSMENT WITHIN SECTOR – RESULTS

Methods: Isolation forest, DBSCAN, PCA

- **2018**

Table 6. Banks that are labeled as anomalies.

Isolation forest	DBSCAN
IDEA BANK	IDEA BANK
BANK POLSKIEJ SPOLDZIELCZOSCI	BANK POLSKIEJ SPOLDZIELCZOSCI
	EURO BANK
	NEST BANK

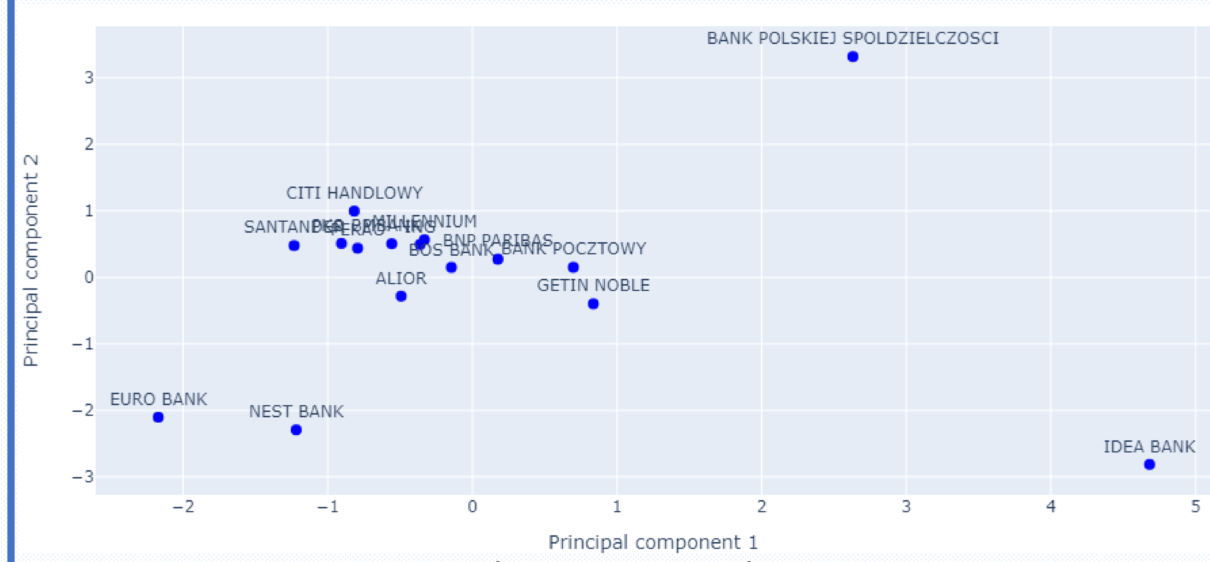

Figure 5. Banking sector analysis using PCA.

- **2020**

Table 7. Banks that are labeled as anomalies.

Isolation forest	DBSCAN
GETIN NOBLE	GETIN NOBLE
NEST BANK	NEST BANK
	BANK POLSKIEJ SPOLDZIELCZOSCI
	CITI HANDLOWY

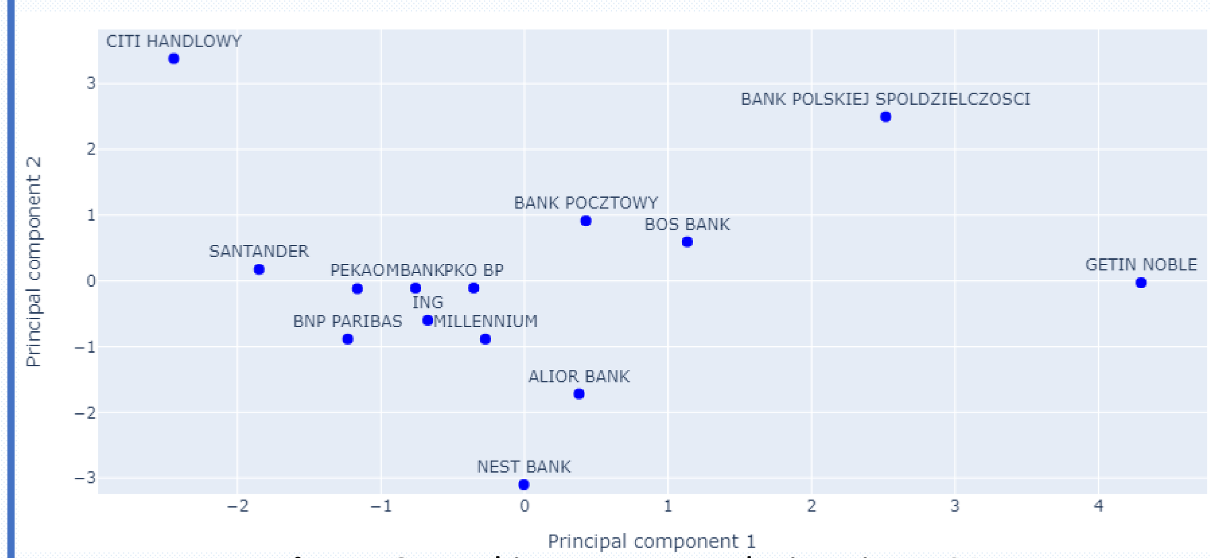

Figure 6. Banking sector analysis using PCA.

SUMMARY

- Gradient boosting methods are more stable and effective than *white-box* models.
- Anomaly detection methods can be used to identify companies that are better or worse than others in a given period in the sector.

Table 8. Effectiveness and stability of the created risk models.

Model	Effectiveness	Stability
Penalties	CatBoost Decision tree	XGBoost Decision tree
Liquidity	LightGBM/CatBoost Decision tree	Decision tree Logistic regression/XGBoost
Loss	Logistic regression Decision tree	XGBoost Decision tree

REFERENCES

1. Altman, Edward I. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance 23: 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x.
2. Lacher, R. C., Pamela K. Coats, Shanker C. Sharma, and L. Franklin Fant. 1995. A neural network for classifying the financial health of a firm. European Journal of Operational Research 85: 53–65. https://doi.org/10.1016/0377-2217(93)E0274-2.
3. Pham, Xuan T.T., and Tin H. Ho. 2021. Using boosting algorithms to predict bank failure: An untold story. International Review of Economics & Finance 76: 40–54. https://doi.org/10.1016/j.iref.2021.05.005.
4. Zmijewski, Mark E. 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. Journal of Accounting Research 22: 59–82. https://doi.org/10.2307/2490859.