Predicting Patterns of Firms' Vulnerability to Economic Crises Using Open Data, Synthetic Minority Oversampling Technique and Machine Learning

Mohsan Ali(✉), Euripidis Loukis, and Yannis Charalabidis

University of the Aegean, Samos, Greece
{mohsan,eloukis,yannisx}@aegean.gr

Abstract. One of the most serious problems of the market-based economies are the instabilities of the economic activity (business cycles), which are sometimes large and lead to recessionary economic crises of different intensities, geographical scopes and durations, and have quite negative consequences for firms. Governments, in order to reduce these negative consequences of economic crises, which can lead to high levels of unemployment and poverty, as well social unrest and political extremism, undertake various interventions, such as large-scale economic stimulus programs. However, in order to maximize their cost-effectiveness, as well as the economic and social value they generate, it is necessary that they are properly targeted and directed to/focused on the most vulnerable firms to the economic crisis. This paper describes a methodology that can be quite useful for this: it enables the prediction of multi-dimensional patterns of individual firms' vulnerability to economic crisis with respect to the main aspects of their financial situation. For this purpose, Machine Learning algorithms are used, in combination with the Synthetic Minority Oversampling Technique (SMOTE) for increasing their performance, which are trained using open government data from Statistical Authorities. Furthermore, a first application/validation of the proposed methodology is presented, using open data from the Greek Statistical Authority about 363 firms for the severe Greek economic crisis period 2009–2014, which gave satisfactory results.

Keywords: Economic Crises · Open data · Machine Learning

1 Introduction

One of the most serious problems of the market-based economies are the instabilities of economic activity (business cycles), which sometimes can be large and result in recessionary economic crises of different intensities, geo-graphical scopes and durations [1–7]. These economic crises have quite negative consequences for society and the economy [1, 2, 8–11]. Due to the decrease of overall economic activity and incomes during economic crises leads most firms face a deterioration in many important aspects

K. Hinkelmann et al. (Eds.): BIR 2023, LNBIP 493, pp. 188–196, 2023.
https://doi.org/10.1007/978-3-031-43126-5_14

of their economic situation: most firms experience decrease of their sales, and therefore their revenue and their liquidity, increase of their debts, and reductions investments and personnel employment, while some firms cannot survive and go bankrupt. However, these negative impacts of economic crisis differ significantly among firms [1–3, 12]: some firms exhibit higher capabilities to cope with the economic crises and therefore higher vulnerability to them, while some other firms cannot sufficiently cope with the crises and are more vulnerable.

So, it is one of the most important challenges of governments face to reduce as much as possible these severe negative consequences for the society and the economy of the economic crises that repeatedly appear, which can result in high levels of unemployment, poverty and social exclusion, as well social unrest, and political extremism. For this purpose, governments undertake huge interventions, such as large-scale economic stimulus programs, which include the provision to firms of tax rebates, financial assistance, subsidies, financial support for investments, low-interest (or even no-interest) loans, etc. [13–16]. These important government interventions, and especially the large-scale economic stimulus programs, are more cost-effective and generate more economic and social value if they are properly targeted and directed to/focused on the most vulnerable firms to the economic crisis.

This paper describes a methodology that can be quite useful for achieving this focus: it enables the prediction of the multi-dimensional 'patterns' of individual firms' vulnerability to economic crisis with respect to the main aspects of their financial situation (such as sales revenue, liquidity, debts, investment, employment, etc.). For this purpose, we employ Artificial Intelligence (AI) algorithms from the area of Machine Learning (ML) [17, 18], which are used in order to construct a set of prediction models of the vulnerability to economic crises of an individual firm with respect the main aspects of their financial situation (i.e. the degree of deterioration of each of them during an economic crisis); as independent variables are used the characteristics of each individual firm. For the training of these prediction models are used open government data (OGD) [19, 20] from Statistical Authorities. Furthermore, a first application/validation of the proposed methodology is presented, which gives satisfactory results.

Our paper consists of four sections. In the following Sect. 2 the proposed methodology is described, and then in the Sect. 3 the abovementioned application of it is presented; lastly, the conclusions are summarized in the final Sect. 4.

2 Proposed Methodology

The proposed methodology aims to predict the multi-dimensional 'pattern of vulnerability' to an economic crisis (VEC) of an individual firm, which is defined as a vector having as components the degrees of deterioration of the main aspects of firm's financial situation, such as sales revenue, liquidity, debts, investment, employment, etc. (they can be measured in a 5-levels Likert scale: not at all, small, moderate, large, very large), during an economic crisis (Fig. 1):

$$\mathrm{VEC} = [\mathrm{VEC}_1, \mathrm{VEC}_2, \ldots, \mathrm{VEC}_\mathrm{N}]$$

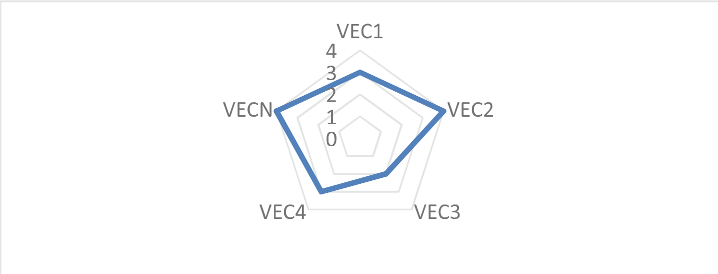

Fig. 1. Multidimensional Pattern of Firm's Vulnerability to Economic Crisis

So, for each of these components/dimensions of firm's vulnerability to economic crisis VEC_1, VEC_2, ..., VEC_N we construct a prediction model of it (having it as dependent variable). In order to determine the appropriate independent variables of these prediction models we have been based on theoretical foundations from management sciences. In particular, several frameworks have been developed concerning the main elements of a firm that determine its performance, with the 'Leavitt's Diamond' framework being the most widely recognized one, which includes five main elements: strategy, processes, people, technology, and structure [21, 22]. We can expect that these five main elements of the 'Leavitt's Diamond' framework will be the main determinants of the performance of a firm both in normal economic periods and in economic crisis ones.

So, the prediction models of firm's economic vulnerability concerning the main aspects of its financial situation VEC_i will include five corresponding groups of independent variables concerning:

a) strategy (e.g., degree of adoption of the main competitive advantage strategies, such as cost leadership, differentiation, focus, innovation, etc.)
b) processes (e.g., main characteristics of firm's processes, such as complexity, flexibility, etc.)
c) people (e.g., shares of firm's human resources having different levels of education or specific skills, certifications, etc.)
d) technology (e.g., use of various production technologies, digital technologies, etc.)
e) structure (e.g., main characteristics of firm's structure, degree of adoption of 'organic' forms of work organization, such as teamwork, etc.)

and also, a sixth group of independent variables concerning general information about the firm, such as size, sector, comparative performance vis-à-vis competitors, etc.

The structure of the prediction models of firm's crisis-vulnerability dimensions' VEC_i (dependent and independent variables) is shown in Fig. 2.

For the construction of each of these prediction model we can use the main supervised ML algorithms described in relevant literature [17, 18], such as Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Support Vector Machines (SVM), and Multilayer Perceptron (MLP), and then compare the prediction performances of the corresponding prediction models, and finally select the one with the highest prediction performance. For training them we can use relevant OGD for economic crisis periods provided by Statistical Authorities; the available dataset will be divided into two parts:

the 'training dataset', which is used for constructing the prediction model, and the 'test dataset', in which the prediction performance of this model is evaluated, by calculating its prediction accuracy, precision, recall, and F-Score are calculated.

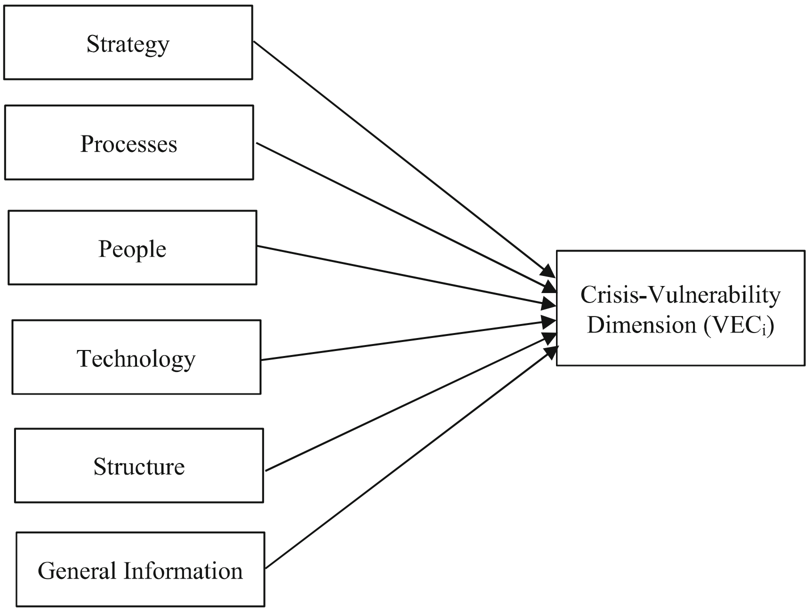

Fig. 2. Structure of the prediction models of firm's crisis-vulnerability dimensions

However, the abovementioned datasets we use for the construction of the models usually a) have missing values, b) their size can be not large enough to train a healthy and accurate ML model, and c) are unbalanced with respect to the classes (they include small numbers of observations/samples for some of the classes, and much larger numbers of observations/samples for some other classes); these can result in prediction models with lower prediction performance and also biased. In order to address problems b) and c) our methodology includes a pre-processing of these datasets using the Synthetic Minority Oversampling Technique (SMOTE) [23]; this technique increases the number of samples of the dataset using the existing samples of the classes (oversampling), balances the dataset with respect to the number of samples of each class, fills missing values, which enable the estimation of better prediction models with higher prediction performance. Furthermore, our methodology includes an initial Exploratory Data Analysis (EDA) in order to get a first insight of the data through visualizations, and make some necessary transformations, and then a Principal Component Analysis (PCA) in order to analyze the importance of the features, and select the eliminate the important ones, and eliminate the ones that are not important, which helps to improve performance and reduce training time.

3 Application

A first application/validation of the proposed methodology was made using a dataset that was released by the Greek Statistical Authority on request by the authors, and after signing an agreement concerning its use, so it constitutes OGD freely available for research purposes. The full dataset was comprised of 363 instances, with each instance being an independent firm. It included for each firm the following features/variables:

a) seven (7) variables concerning firm's vulnerability to the severe economic crisis that Greece experienced between 2009 and 2014 with the main aspects of its economic situation: degree of decrease of domestic sales, foreign sales, employment, traditional investment (e.g. in equipment, buildings, etc.), innovation investment and liquidity, and degree of increase of firm's debt, due to the economic crisis; all these variables were measured in a 5-levels Likert scale (not at all, small, moderate, large, very large), and were then converted to binary ones (with the first three values being converted to 'non-vulnerable' and the other two being converted to 'vulnerable');
b) forty (40) variables concerning various firm's characteristics with respect to strategy, personnel, technology (focusing on the use of various digital technologies), structure (focusing on the use of organic' forms of work organization, such as teamwork) and also some general information about the firm (size measured through the number of employees, sector (services or manufacturing), comparative financial performance in the last three years in comparison with competitors).

The dataset had missing values, so initially we proceeded to filling each missing value with the most suitable value based on the type of the variable (for instance, if the variable was ordinal, then the missing values were filled with the highest relative frequency value of this variable). Then exploratory data analysis (EDA) was applied in order to gain a better insight into the data through visualizations and make some necessary transformations. As a next step, since the size of our data set (which included data for 363 firms as mentioned above) did not allow constructing/training supervised ML prediction models with good prediction performance, we used the abovementioned oversampling and class-balancing algorithm SMOTE. Finally, the dataset was divided into a training dataset including 66% of the samples and a testing dataset including 33% of the samples. The former was used for the training of ML models with the following algorithms: Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Support Vector Machines (SVM), and Multilayer Perceptron (MLP); the latter was used for assessing the prediction performance of these ML models. We can see the results (prediction accuracy, precision, recall and F-score) in Table 1 (with bold are shown for each dimension of crisis-vulnerability the results for the best performing algorithm).

We can see that for two dimensions of firm's vulnerability to economic crisis, concerning foreign sales and employment (degree of decrease of foreign sales and degree of decrease of employment due to the economic crisis), we have a high prediction performance (prediction accuracy 89% and 85% respectively). For two more dimensions of firm's vulnerability to economic crisis, concerning debt and liquidity (degree of increase of debt and degree of decrease liquidity due to the economic crisis), we have a lower – but still high - prediction performance (prediction accuracy 79% for both). For the remaining three dimensions, concerning domestic sales, traditional investment, and innovation

Table 1. Prediction Performance of AI/ML Algorithms for each Crisis-Vulnerability Dimension

Crisis – Vulnerability Dimension	AI/ML Algorithm	Accuracy	Precision	Recall	F-score
Domestic Sales Crisis-Vulnerability	Decision Tree	0.67	0.66	0.67	0.66
	Random Forest	0.76	0.75	0.76	0.75
	Logistic Regression	0.62	0.60	0.60	0.60
	SVM	**0.77**	**0.76**	**0.76**	**0.76**
	MLP	0.68	0.64	0.63	0.64
Foreign Sales Crisis-Vulnerability	Decision Tree	0.76	0.75	0.75	0.75
	Random Forest	**0.89**	**0.90**	**0.87**	**0.88**
	Logistic Regression	0.65	0.64	0.61	0.60
	SVM	0.87	0.88	0.85	0.86
	MLP	0.73	0.72	0.69	0.69
Employment Crisis-Vulnerability	Decision Tree	0.74	0.74	0.74	0.74
	Random Forest	**0.85**	**0.85**	**0.85**	**0.85**
	Logistic Regression	0.65	0.64	0.65	0.65
	SVM	0.77	0.76	0.74	0.75
	MLP	0.69	0.67	0.66	0.67
Traditional Investment Crisis-Vulnerability	Decision Tree	0.71	0.70	0.71	0.70
	Random Forest	0.76	0.78	0.73	0.73
	Logistic Regression	0.62	0.62	0.62	0.62
	SVM	**0.74**	**0.74**	**0.72**	**0.72**
	MLP	0.62	0.61	0.58	0.56
Innovation Investment Crisis-Vulnerability	Decision Tree	0.60	0.62	0.60	0.61
	Random Forest	0.74	0.74	0.74	0.74
	Logistic Regression	0.60	0.61	0.60	0.60
	SVM	**0.76**	**0.75**	**0.76**	**0.75**
	MLP	0.64	0.64	0.64	0.64
Debt Crisis-Vulnerability	Decision Tree	0.68	0.68	0.68	0.68

(*continued*)

Table 1. (*continued*)

Crisis – Vulnerability Dimension	AI/ML Algorithm	Accuracy	Precision	Recall	F-score
	Random Forest	**0.79**	**0.79**	**0.79**	**0.78**
	Logistic Regression	0.64	0.63	0.64	0.62
	SVM	0.77	0.79	0.77	0.76
	MLP	0.64	0.64	0.64	0.62
Liquidity Crisis-Vulnerability	Decision Tree	0.70	0.71	0.70	0.70
	Random Forest	**0.79**	**0.78**	**0.79**	**0.78**
	Logistic Regression	0.63	0.64	0.63	0.63
	SVM	0.75	0.76	0.75	0.76
	MLP	0.56	0.54	0.56	0.55

investment, we had slightly lower prediction performance (prediction accuracy 77%, 76% and 74% respectively).

Overall, the results of this first application of the proposed methodology (prediction performances) can be regarded as satisfactory, taking into account the small size of the dataset we have used (data from 363 firms), and provide a first validation of this methodology; we expect that using a larger dataset (as governments have such data for quite large numbers of firms) will allow training crisis-vulnerability prediction models with higher prediction performances.

4 Conclusion

In the previous sections of this paper has been described a methodology for predicting the whole pattern of vulnerability to economic crisis of individual firms, which respect to the main aspects of their financial situation (such as sales, liquidity, debt, investment, employment, etc.); for this purpose are used AI/ML techniques, in combination with SMOTE in order to increase their performance, which are trained using OGD from Statistical Authorities. The proposed methodology aims to support and enhance/augment (in order to increase its effectiveness) one of the most important and costly kinds of interventions that governments have to make: the interventions they undertake in tough times of economic crises in order to reduce their negative consequences, and especially the large-scale economic stimulus programs.

The research presented in this paper has some interesting implications for research and practice. With respect to research, it makes a significant contribution to two highly important research streams. First, to the growing research stream investigating the use of AI in government, by developing a novel approach for a highly beneficial use of AI/ML for supporting and enhancing/augmenting a critical activity of government with quite high economic/social importance and financial magnitude. Second, to the research stream investigating OGD research stream, by providing an approach for increasing the economic/social value generation form the OGD through advanced processing of them using AI/ML techniques. With respect to practice, it provides a useful tool to government agencies that are responsible for the design and implementation of economic stimulus programs aiming to reduce the negative consequences of economic crises.

However, further research is required in the following directions: a) further application of the proposed methodology, using larger datasets, and in various national and sectoral contexts (experiencing different types and intensities of economic crises); b) investigate the use of other pre-processing algorithms (e.g., oversampling and class-balancing algorithms) as well as AI/ML algorithms (deep learning ones); c) investigate the combination of such OGD with other sources of data about firms (e.g. from other government agencies, from private firms, such as private business information databases), in order to obtain more information about firms that might improve the performance of the prediction of their pattern of vulnerability to economic crises.

Acknowledgments. This research has been conducted as part of the project ODECO that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 955569.

References

1. OECD. Responding to the Economic Crisis Fostering Industrial Restructuring and Re-newal (2009a)
2. Keeley, B., Love, P.: From Crisis to Recovery: The Causes, Course and Consequences of the Great Recession (2010)
3. Knoop, T.A.: Recessions and Depressions: Understanding Business Cycles (2015)
4. Allen, R.E.: Financial Crises and Recession in the Global Economy, 4th edn (2016)
5. Saifi, S., Horowitz, J.: Blackouts and soaring prices: Pakistan's economy is on the brink (2023)
6. Loukis, E., Arvanitis, S., Myrtidis, D.: ICT-related behavior of Greek banks in the economic crisis. Inf. Syst. Manag. **38**, 79–91 (2021)
7. Baldwin, R., di Mauro, B.W.: Mitigating the COVID Economic Crisis: Act Fast and Do Whatever It Takes (2020)
8. Oliveras, L., et al.: Energy poverty and health: trends in the European union before and during the economic crisis, 2007–2016. Health Place **67**, 102294 (2021)
9. Dagoumas, A., Kitsios, F.: Assessing the impact of the economic crisis on energy poverty in Greece. Sustain. Cities Soc. **13**, 267–278 (2014)
10. Stylidis, D., Terzidou, M.: Tourism and the economic crisis in Kavala, Greece. Ann. Tourism Res. **44**, 210–226 (2014)
11. Das, M.: Economic crisis in Sri Lanka causing cancer drug shortage. Lancet Oncol. **23**, 710 (2022)
12. Arvanitis, Sp., Loukis, E.: Factors explaining ICT investment behavior of firms during the 2008 economic crisis. Inf. Syst. Manag. 1–22 (2023). https://doi.org/10.1080/10580530.2023.2213839
13. Khatiwada, S.: Stimulus Packages to Counter Global Economic Crisis
14. Coenen, G., Straub, R., Trabandt, M.: Gauging the effects of fiscal stimulus packages in the Euro area
15. Kalinowski, T.: Crisis management and the diversity of capitalism: fiscal stimulus packages and the East Asian (neo-)developmental state. Econ. Soc. **44**, 244–270 (2015)

16. Taylor, J.B.: Fiscal Stimulus Programs During the Great Recession (201AD)
17. Witten, I.H., Frank, E., Hall, M.A.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann (2011)
18. Russell, S.J., Norvig, P., Davis, E.: Artificial Intelligence: A Modern Approach. Pearson (2015)
19. Gao, Y., Janssen, M., Zhang, C.: Understanding the evolution of open government data research: towards open data sustainability and smartness. Int. Rev. Adm. Sci. **89**, 59–75 (2023)
20. Charalabidis, Y., et al.: The world of open data: Concepts, methods, tools and experiences. In: Public Administration and Information Technology, vol. 28, pp. 1–194 (2018)
21. The Corporation of the 1990s: information technology and organizational transformation. Oxford University Press (1991)
22. Applied organizational change in industry: Structural, technological and humanistic approaches, Handbook of organizations. Routledge (2013)
23. Dudjak, M., Martinović, G.: In-depth performance analysis of SMOTE-based oversampling algorithms in binary classification. Int. J. Electr. Comput. Eng. Syst. **11**, 13–23 (2020)