Leveraging Government Data Using Unsupervised and Supervised Machine Learning for Firms' Investment Policy-Making in Economic Crises

Loukis Euripidis[1] and Kyriakou Niki[2]

[1] University of the Aegean, Samos, GR
eloukis@aegean.gr
[2] University of the Aegean, Samos, GR
nkyr@aegean.gr

Abstract. In market-based economies economic crises of different geographical scopes and durations are often appearing, and are resulting in economic recessions, which have quite negative consequences for the economy and the society. Governments respond by undertaking large-scale economic stimulation programs, spending vast amounts of financial resources (with orders of magnitude between 3-6% of GDP), in order to mitigate these negative consequences. It is of critical importance to make effective use of these huge financial resources, in order to have high positive impact on the economy and the society in these tough crisis periods. This necessitates careful and rational design and implementation of these large and costly economic stimulation programs. Since one of the most important consequences of economic crises is the decrease of firms' investments, the above economic stimulation programs include investment support actions, which aim to mitigate these crisis-induced firms' investment decreases, and include a wide range of interventions for this reason, such as investment incentives, subsidies, low-interest loans as well as relevant tax rebates. In this paper is presented an integrated methodology for leveraging government data from economic crisis periods, using on one hand Unsupervised Machine Learning techniques, and on the other hand Supervised Machine Learning ones, in order to provide support for the rational design and implementation of firms' investment support actions in economic crises. A first application of the proposed methodology is presented, based on existing data from the Greek Ministry of Finance and the Statistical Authority concerning 363 firms for the economic crisis period 2009-2014, which gave interesting and encouraging results.

Keywords: Economic Crisis, Recession, Stimulation Packages, Artificial Intelligence, Machine Learning, Supervised Learning, Unsupervised Learning

1 Introduction

In market-based economies economic crises of different geographical scope and durations are often appearing, and are resulting in economic recessions, which have quite negative consequences for the economy and the society [1-6]. During the last century numerous economic crises have appeared [2]. A decade ago, we experienced the severe 2007 Global Financial Crisis, while recently we experienced an economic crisis caused by the COVID-19 pandemic [7], and currently the Ukraine war, and the big increases in the prices of oil, gas, wheat and other goods it gives rise to, is expected to spark another economic crisis.

As the negative consequences of economic crises are often severe, governments undertake large-scale economic stimulation programs, and spend vast amounts of financial resources, having orders of magnitude between 3-6% of GDP, in order to mitigate these consequences (such as the recent American Recovery and Reinvestment Act (ARRA) in the United States and the European Economic Recovery Plan (EERP) in the European Union) [8-12]. These increase considerably national debts and cause big macro-economic problems in the post-crisis periods. Therefore, it is of critical importance to design and implement carefully and rationally these large and costly economic stimulation programs, in order to use these huge financial resources effectively, and have high positive impact on the economy and the society in these tough crisis periods. For this purpose, considerable research has been conducted for the assessment of the effects of such economic stimulation programs, which have been designed and implemented for addressing previous economic crises, such as the 2007 Global Financial Crisis, in order to draw useful conclusions, insights and knowledge that can be used for addressing future crises [11-12]. However, there is a lack of research concerning the use and leveraging of the extensive firm-level data from crisis periods that government agencies possess for this highly important purpose: for the support of the rational design and implementation of these large-scale economic stimulation programs, in order to increase their effectiveness and positive economic and social impact. It is therefore a big challenge, of critical economic and social importance, to extract from these extensive government firm-level data from crisis periods as much as possible insight and knowledge, which can be useful in the future for optimizing these economic stimulation programs and using more effectively their huge financial resources.

As many crises have appeared in the last decades, while the economic stability periods have become shorter, governments gradually realize that they have to learn more about how to manage not only normal economic stability periods, but also tough economic crisis periods as well. It is therefore imperative to increase their knowledge about the multiple types of consequences of these economic crises, as well as the possible interventions that can be undertaken in order to mitigate these consequences, and also ways to make these interventions more effective. For these purposes quite useful can be both the macro-economic, and also the micro-economic data as well, which are collected by government agencies during economic crises periods. It is necessary to leverage these valuable data as much as possible, by making the most intensive possible exploitation of them, by using highly sophisticated techniques, especially from the Artificial Intelligence (AI) domain, such as Unsupervised and Supervised Machine Learning techniques, in order to maximize the extraction of useful insights and knowledge from these data, and also to make reliable predictions based on them. These are going to enable governments to design and implement better and more focused and effective programs as well as specific actions for mitigating the negative consequences of economic crises for the economy and the society.

This paper contributes to filling the abovementioned research gap, focusing on one of the most critical components of these economic stimulation programs for addressing economic crises: the firms' investment support actions (which are highly important for their post-crisis competitiveness). In particular, it presents an integrated methodology for leveraging firm-level government data from economic crisis periods, using AI techniques, on one hand Unsupervised Machine Learning techniques, and on the other hand Supervised Machine Learning ones [13-17], in order to provide support for the rational design and implementation by government of firms' investment support actions during economic crises. It includes initially the use of Clustering Analysis techniques (Unsupervised Machine Learning), in order to investigate if we can distinguish some typologies of firms with respect to the impact of economic crisis on the main types of investments they make. If this happens, the proposed methodology includes the use of Analysis of Variance (ANOVA) next, in order to understand better the main characteristics of these typologies of firms (e.g. with respect to personnel, ICT use, processes, strategic directions, innovation, exports, etc.). Finally, it includes the use of Prediction techniques (Supervised Machine Learning) in order to develop prediction models for firm's investment resilience in economic crisis based on the abovementioned individual characteristics of them. Also, a first application of the proposed methodology is presented, based on existing data from the Greek Ministry of Finance and the Statistical Authority concerning 363 firms for the economic crisis period 2009-2014, which gave encouraging and interesting results.

Our research contributes to the growing body of knowledge concerning the use of Artificial Intelligence (AI) in government (briefly reviewed in 2.2), by developing a composite integrated methodology of AI exploitation, which includes of combination of Unsupervised and Supervised Learning techniques, for the design and implementation of policies, programs and actions concerning one of the most severe and difficult problems that governments face: the economic crises.

In the following section 2 the background of our methodology is outlined, while in section 3 the methodology is described, followed by the abovementioned application of it in section 4. The final section 5 summarizes conclusions and proposes directions for future research.

2 Background

2.1 Economic Stimulation Programs

According to relevant literature [1-6] the economic crises usually result in contractions of economic activity, leading to economic recessions, which have two main categories of negative effects on firms: a) decrease of firms' production, procurement, and personnel employment (which increases unemployment, poverty and social exclusion); and b) decrease of the different types of investments they make (e.g. in equipment, buildings, training of personnel, R&D, innovation, etc.). Though the former category of negative effects of the economic crises on firms is more widely and extensively debated, due to their painful short-term consequences, such as the increase of unemployment, and therefore poverty and marginalization, the latter category has equally or even more detrimental medium- or long-term consequences; the most important of them are firm's technological backwardness and obsolescence, loss of important development opportunities, and finally lower competitiveness and growth.

So, governments, in order to mitigate these negative consequences of economic crises, which can give rise to social unrest and political extremism, undertake large-scale economic stimulation programs, spending huge amounts of financial resources [8-12]. These programs vary in size (e.g. the stimulus program of the EU for addressing the recent 2007 Global Financial Crisis amounted to 5% of GDP in the EU [8], while the corresponding program of China was much bigger, reaching an estimated 12.5% of its GDP [10]) as well as in composition (i.e. in the specific actions they include). In general, they include two main categories of actions: i) demand-side oriented ones (aiming to stimulate domestic consumption by citizens, e.g. unemployment assistance, nutritional aid, health and welfare payments, tax cuts, etc.); and ii) supply-side oriented ones (public infrastructure investments, as well as private investment incentives, subsidies, low-interest loans, relevant tax rebates, etc., usually promoting 'green growth', adopting new technologies, innovation, etc.) [10]. The shares of these two categories of actions in the economic stimulation programs vary among countries, but all of them place great emphasis in the mitigating firms' investment decrease during economic crises, through various kinds of actions, such as investment incentives, low-interest loans, subsidies, relevant tax rebates, etc. However, it is widely recognized that these actions should be highly focused on the firms that really need support of their investments (in general, or for specific types of investment, such as the 'soft investments' (e.g. in personnel training, marketing/advertisement) or 'innovation investment' (e.g. in R&D, processes innovation, products/services innovation, etc.).

Our research aims to leverage existing firm-level data from crisis periods possessed by government, by performing highly sophisticated processing of them, using AI techniques, in order to support the design and implementation of this particular highly important supply-side oriented firms' investment support actions in economic crises.

2.2 Artificial Intelligence in Government

Even though AI existed for several decades, its 'real life' exploitation was limited; however recently there has been a high interest in the 'real life' application of AI techniques, initially by private sector firms, for a number of reasons: a) availability of large amounts of data, which enable a more effective training of AI algorithms (and finally the extraction of more reliable models and rules); b) advances in computing power and reduction of its cost; c) substantial improvements of AI algorithms [13-17]. The first AI use initiatives in the private sector have revealed the great potential of AI techniques to offer important benefits, such as improvements in productivity, increase of sale revenue and growth, better decision-making as well as substantial innovations in internal processes, products and services [18].

These first success stories of high beneficial application of AI in the private sector have generated high levels of interest to use AI techniques in the public sector as well, in order to exploit better the huge amounts of data possessed by government agencies, on one hand for supporting decision-making and policy-making, and on the other hand for automating or supporting substantially more sophisticated mental tasks than the simpler routine ones automated or supported by the traditional operational IS of government agencies [21-27]. According to the study described in [26] AI has great potential to support and improve the core government functions:

i) Policy-Making (by enabling/supporting the detection of social issues more quickly, the improvement of public policy decisions, the estimation of potential effects of policy, the monitoring of the implementation of policy as well as the evaluation of existing policy, and the enhancement of citizens' participation in policy making);

ii) Public Services Delivery (by enabling/supporting the improvement of the information services of the organization, as well as the delivery of public service to businesses and citizens, and also the development of new innovative public services);

iii) Internal Management (by enabling/supporting the improvement of the allocation of human resources, the recruitment services of the public organizations, their financial management, the detection of fraud and/or corruption, the maintenance of equipment, the public procurement processes and also organizational (cyber)security).

Using the above typology of AI exploitation in government as an analysis framework, a sample of 250 cases of government use of AI across the European Union were analyzed; it was concluded that AI is used mainly to support the improvement of public service delivery, followed by the enhancement of internal management, but only in a limited number of cases AI was used for the support (directly or indirectly) of policy and decision making.

Some research has been conducted on the development of ways/methodologies of exploiting AI in different public sector thematic domains, for various kinds of problems and tasks, for instance in education, for the prediction of applicants for teacher positions who will be more effective and successful, in order to support making the optimal recruitment decisions [28]; in social policy, for the prediction of higher risk youth concerning criminal activity, in order to target prevention interventions [29]; in restaurant hygiene inspections, for harnessing the social media on-line reviews in order to identify restaurant likely to be severe offenders, for optimizing inspections [30]; in public security, for predictive police patrolling, in order to use more effectively scarce human resources [31], and for the automated analysis and classification of crime reports [32]; in public transportation management in order to predict high crime risk transportation areas [33]; in environmental management and planning for the prediction of ground water levels [34]; in healthcare, for supporting diseases' diagnosis and treatment planning [35].

However, it is widely recognized that only a small part of the great potential of AI use in government has been discovered and exploited; so further research is required in order to exploit this potential to a larger extent: for the development of new innovative ways and methodologies (including combinations of AI techniques, and possibly advanced statistical techniques), for exploiting the potential of AI in different public sector thematic domains, for various kinds of problems and tasks, with main focus on the most severe problems of modern societies and economies. In this direction our research makes a useful contribution, by developing and making a first application of an integrated methodology for leveraging government data from economic crisis periods, using a combination of Unsupervised and Supervised Machine Learning techniques, and also Statistical techniques, in order to provide support for the rational design and implementation of firms' investment policies in economic crises.

2.3 Firm Performance Determinants

Previous economic and management science research has investigated the main elements of a firm that determine its performance. Economic research has concluded that the main production factors of a firm that determine its output and performance are:
a) its capital (meant as the different kinds of production equipment it uses), discriminating between non-computer capital and computer capital,
b) its labor (meant as numbers of personnel of various educational levels and specializations), discriminating between non-computer labor and computer labor,
while recently there is an increasing recognition of the importance also of firm's 'organizational capital' (meant as processes and structures adopted by the firm) as well as 'human capital' (meant as the skills and knowledge of firm's personnel) for its output and performance [36-39].

At the same time management science research has developed several conceptualizations of the main elements of a firm that determine its performance; the most widely recognized and used one is definitely the 'Leavitt's Diamond' framework [40]. According to it the most important elements of a firm that determine its performance are:

i) its task (= the strategies of the firm, as well as the administrative and production processes it follows for implementing these strategies),

b) its people (= the skills of firm's human resources of the firm),

c) its technology (= the technologies used for implementing the above administrative and production processes),

d) its structure (= the organization of the firm in departments, and the communication and coordination patterns them).

An extension of it has been developed subsequently, which analyses the above 'task' element into the 'strategy' and 'processes' elements [41]. We remark that this 'Leavitt's Diamond' framework includes a wider set of firm's elements that determine its performance in comparison with the abovementioned economic framework. Furthermore, there are similarities between these two frameworks: some of the above five main elements of a firm that determine its performance correspond at least to some extent to those determined by economic research. In particular, the 'technology' corresponds to 'capital' (non-computer and computer one), the 'people' correspond to 'labor' and 'human capital', while the 'structure and the 'processes' part of the 'task' correspond to 'organizational capital'.

We can expect that firm's characteristics concerning the above five main elements (strategy, processes, people, technology and structure) will determine to a considerable extent firm's performance during not only normal economic stability periods, but also economic crisis periods as well; these characteristics are expected to determine firm's ability to cope with the difficult economic crisis conditions, minimizing the decrease of sales revenue and profits due to the crisis, and therefore increasing the availability of financial resources for making investments, and also to identify, design and implement successfully highly valuable investments for handling the difficult crisis context.

3 The Proposed Methodology

The proposed methodology uses firm-level data possessed by government from economic crisis periods concerning:
i) the extent of decrease during the economic crisis period of the main types of firm's investments, such as 'basic investments' (e.g. in production equipment, buildings, etc.), 'soft investments' (e.g. in personnel training, marketing/advertisement) and 'innovation investment' (e.g. in R&D, processes innovation, products/services innovation, etc.) (variables INVD1, INVD2, …, INVDN);
ii) various firm characteristics concerning the abovementioned five main elements of a firm that determine its performance, such as personnel, ICT use, processes, strategic directions, innovation, exports, etc. (variables CH1, CH2, …, CHM).
These data are usually collected annually by Ministry of Finance – Taxation Authorities and Statistical Authorities, as firms have legal obligation to provide them, and also there are sanctions in case of not providing these data, or providing inaccurate data; therefore these data possessed by government are of high quality: they are complete and highly reliable, so they are appropriate to be used for extracting insight and knowledge from them, and also for estimating prediction models.

These data undergo advanced processing, which consists of three stages (Fig.1):

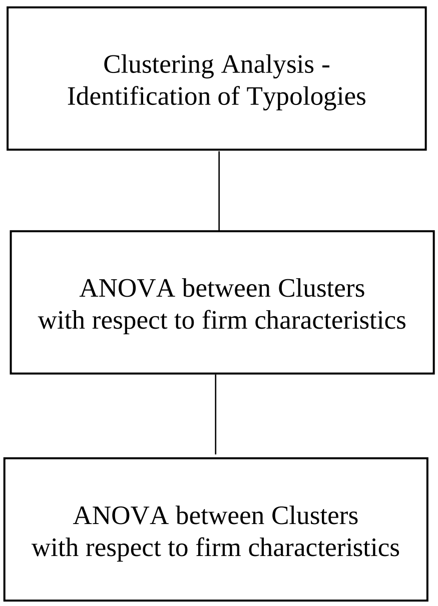

Fig. 1. Structure-stages of the proposed methodology

a) In the first stage we are using Clustering Techniques (Unsupervised Learning) in order to investigate whether we can distinguish some discrete clusters/typologies of firms with respect to the impact of the economic crisis on the main types of investments.

For this purpose, we are using the abovementioned investment decrease variables INVD1, INVD2, …, INVDN (which measure the extent of decrease of the main types of investments made by each particular firm) for performing Clustering Analysis [13-17]:
- initially we perform Hierarchical Clustering, in order to determine the number of clusters (based on the 'gaps' of the dendrogram);
- then using this number of clusters we perform K-means Clustering, in order to determine for each of the firms of our dataset to which if the clusters it belongs (cluster membership);
- based on the cluster memberships of these firms we calculate the center of each cluster (= averages of the above investment decrease variables over all the firms of the cluster);
- and finally perform Analysis of Variance (ANOVA) between the clusters with respect to these investment decrease variables, in order to identify the variables (i.e. types of investment) in which these clusters differ most.

If more than one clusters are identified this indicates that an 'one size fits all' investment policy/set of actions for mitigating the crisis-induced decrease of firms' investments is not the most appropriate approach; it is necessary for each firms' cluster/typology to design and implement a different more focused and specialized investment policy/set of actions. The results of the above Cluster Analysis concerning the center of each cluster (which show the extent of decrease of the main types of investment type in the firms of the cluster) enable an understanding of the cluster (typology) and can provide a sound basis and direction for the design and implementation of a specialized/focused investment policy/set for these firms.

b) In the second stage we are performing ANOVA among the clusters with respect to the firm characteristics' variables CH1, CH2, …, CHM, in order to investigate whether there are differences among the above firms' clusters/typologies with respect to various characteristics concerning the abovementioned five main elements of a firm that determine its performance, such as personnel, ICT use, processes, strategic directions, innovation, exports, etc. (used as independent variables). This enables a better understanding of each of the clusters/typologies identified in the previous stage, concerning the main characteristics of the firms it includes, as well as the firm characteristics in which these clusters/typologies differ most; so, it provides further basis and direction for the design and implementation of a specialized/focused investment policy/set of interventions for each of the identified clusters/typologies of firms.

c) In the third stage we are using Prediction Techniques (Supervised Learning) in order to construct prediction models for an overall index of investment decrease in economic crisis INVD, which is equal to the average of the INVD1, INVD2, …, INVDN variables, using as predictors the abovementioned the firm characteristics' variables CH1, CH2, …, CHM. For this purpose, we can use the existing Supervised Learning algorithms for predicting continuous dependent variables, such as Generalized Linear Modelling, Deep Learning, Decision Trees, Random Forests, Gradient Boosted Trees, Support Vector Machines, etc. [13-17], and select the one that exhibits the highest performance (the lowest absolute error). It is necessary these predictions to be as 'explainable' as possible, so in this direction we have to exploit the research that has been conducted on 'Explainable Artificial Intelligence' [42].

This third stage enables predicting for an individual firm the extent of investment decrease in economic crisis, which can be viewed as its 'investment resilience in economic crisis', based on its particular characteristics, such as personnel, ICT use, processes, strategic directions, innovation, exports, etc. This capability can be very useful for the more effective implementation of investment policies, interventions and programs during economic crisis. In particular, this enables the prediction of this 'investment resilience in economic crisis' for all firms applying for various interventions/programs of investment incentives, subsidies, low-interest loans, tax rebated, etc. implemented in the beginning of future economic crises; this prediction can be taken into account as an additional selection criterion, favoring firms that are predicted to exhibit lower investment resilience, and therefore larger decrease of investments, in economic crisis, so that these interventions/programs can be more focused on such firms, and therefore be more focused and effective.

The first and the second stage of the proposed methodology provide support for gaining a better and deeper understanding of the impact of the particular economic crisis on the main types of investments that firms make, so they provide a sound base for the 'evidence-based' design of appropriate effective firms' investment support policies/actions (and possibly a set of different specialized/focused policies/actions for different groups/clusters of firms, instead of a single 'one size fits all' investment policy/actions that might be less effective). The third stage of our methodology supports relevant internal operations of the government agencies, which are responsible for the implementation of these firms' investment support policies, enabling the effective implementation of them, by focusing on the firms that are expected to have the largest decrease in the investments, so they will be most in need of investment support by government. Therefore, the proposed methodology supports and improves two out of the three core government functions that according to [26] (as mentioned in more detail in 2.2) AI has a great potential to support and improve: policy-making and internal operations/management (which according to this study are the least exploited types of AI use in the governments of the member states of the European Union).

4 Application

A first application of the proposed methodology has been made, using data for 363 firms for the period 2009-2014 from the Ministry of Finance – Taxation Authorities and the Statistical Authority of Greece. These firms cover a wide range of sectors and sizes: 40.2% of them were from manufacturing sectors, 9.4% from constructions, and 50.4% from services sectors; also, 52.6% of them were small, 36.1% medium and 11.3% large ones. In particular, we used data concerning the following variables:

- extent of decrease of firm's investments in production equipment, buildings, personnel training, marketing/advertisement, R&D, processes innovation and products/services innovation (variables INVD1, INVD2, …, INVD7 – five levels ordinal variables: 1 = 'negligible'; 2 = 'small decrease'; 3 = 'moderate decrease'; 4 = 'large decrease'; 5 = 'very large decrease'),

- strategic orientations: extent of adoption by the firm of the main strategies described in relevant strategic management literature [43]: cost leadership, differentiation and innovation (variables STRAT_CL, STRAT_DIF, STRAT-INNOV - five levels ordinal variables: 1 = 'not at all'; 2 = 'to a small extent'; 3 = 'to a moderate extent'; 4 = 'to a large extent'; 5 = 'to a very large extent'), introduction of process and product innovations by the firm in the last three years (INNOV_PROC, INNOV_PRS – binary variables), introduction of innovations by the firm in the production or service delivery processes, in the sales, shipment or warehouse management processes, and in the support processes (e.g. equipment maintenance processes) in the last three years (INN_PRSD, INN_SSWM, INN_SUPP – binary variables), percentage of 2014 firm's sales revenue coming from new products/services introduced during the last three years (NEW_PS – continuous variable), percentage of 2014 firm's sales revenue coming from products/services introduced before 2012 but significantly improved during the last three years (IMPR_PS – continuous variable), existence of Research & Development in the firm (R&D – binary variable) and percentage of exports in firm's sales revenue in 2014 (EXP_P - continuous variable),

- processes: use of 'organic' structural forms of work organization in the firm, such as teamwork and job rotation [44-45] in the last three years (ORG – binary variable),

- personnel: number of firm's employees at the end of 2014 (EMPL – continuous variable), shares of firm's employees having tertiary education, vocational/technical education, high school education, elementary school education (EMPL_TERT, EMPL – VOCT, EMPL_HIGH, EMPL_ELEM – continuous variables), shares of firm's employees using for their work computers, firm's intranet (internal network), Internet (EMPL_COM, EMPL – INTRA, EMPL_INTER – continuous variables) and share of specialized ICT personnel in firm's workforce (EMPL_ICT – continuous variable),

- technology: extent of use of ERP, CRM, SCM, Business Intelligence/Analytics, Collaboration Support systems in the firm (ERP, CRM, SCM, BIBA, CS - five levels ordinal variables: 1 = 'not at all'; 2 = 'to a small extent'; 3 = 'to a moderate extent; 4 = 'to a large extent'; 5 = 'to a very large extent'), conduct of e-sales of products/services (E-SAL – binary variable), extent of use of social media by the firm for sales promotion, collection of customers' opinions, comments and complaints, collections of ideas for improvements and innovations in firm's products/services, finding personnel, supporting the internal exchange of information and co-operation among firm's employees, supporting the external exchange of information and co-operation with other firm (e.g. suppliers, partners, customers, etc.) (SM_SPRO, SM_OPCO, SM_IMINN, SM_PERS, SM_INTCO, SM_EXTCO – three levels ordinal variables: 1 = 'not at all', 2 = 'to a small extent', 3 = 'to a large extent'), use of cloud computing by the firm (CLOUD – binary variable), extent of use of cloud IaaS, PaaS and SaaS services by the firm (CL_IAAS, CL_PAAS, CL_SAAS - five levels ordinal variables: 1 = 'not at all'; 2 = 'to a small extent'; 3 = 'to a moderate extent'; 4 = 'to a large extent'; 5 = 'to a very large extent')

- general firm information: sector (SECT – binary variable: 1 = 'manufacturing or constructions', 2 = 'services'), level of firm's comparative performance in comparison with the other competitor firms in terms of profitability, sales revenue, market share and return on investment (ROI) (COMP_PROF, COMP_SALR, COMP_MS, COMP_ROI - five levels ordinal variables: 1 = ‘much lower than the average’; 2 = ‘lower than the average’; 3 = ‘about at the average’; 4 = ‘higher than the average’; 5 = ‘much higher than the average’).

4.1 Cluster Analysis

Initially, using the investment decrease variables INVD1, INVD2, …, INVD7 we performed Hierarchical Clustering in order to determine the number of clusters – firms’ typologies with respect to investment decrease during the economic crisis. Based on the ‘gaps’ of the dendrogram we can distinguish three clusters of firms. Then we performed K-means Clustering, setting the number of clusters equal to three, in order to determine for each of our dataset the cluster it belongs to (cluster membership), and then calculate the centers of the three clusters with respect to the abovementioned investment decrease variables, which are shown in the second, third and fourth column of Table 1.

Table 1. Center of clusters – ANOVA results

Variable	Cluster 1	Cluster 2	Cluster 3	F-ANOVA	Sig.
INVD1	3.85	2.99	1.70	145.927	0.000
INVD2	3.84	2.88	1.40	138.873	0.000
INVD3	3.68	2.87	1.34	236.410	0.000
INVD4	4.16	3.16	1.69	183.856	0.000
INVD5	4.25	2.47	1.22	380.962	0.000
INVD6	3.92	2.38	1.23	444.162	0.000
INVD7	3.72	2.39	1.23	291.181	0.000

We remark that the firms of the first cluster had medium to large decrease (being closer to the latter) in their basic investments in production equipment and buildings, in their soft investment in personnel training, as well as in their investment in products/services innovation and process innovation, and large to very large decrease in their soft investment in marketing/advertisement, as well as in R&D. Therefore in these firms the economic crisis haD severe negative impact on their investment, especially in their soft investment in marketing/advertisement and in R&D.

The second cluster of firms had slightly lower than moderate decrease in their basic investments in production equipment and buldings, and in their soft investments in personnel training, but only small to moderate (closer to the former) decrease in their innovation-oriented investments in R&D, products/services innovation and process innovation; also, they had slightly higher than moderate decrease in their marketing/advertisement investment. The firms of this second cluster exhibited a quite different behaviour during the economic crisis with respect to their investment than the ones of the first cluster: they maintain the level of investment in innovation, in order to cope with the difficult economic conditions of the crisis (decrease in products/services demand, and therefore in sales revenue) through innovation (innovative products/services, and also innovation in internal processes in order to reduce operating costs). At the same time they make less than moderate reductions in the remaining types of investment, with the only exception of the marketing/advertisement ones.

Finally, the third cluster of firms had negligible to small decrease in all the examined types of investments, being closer to negligible decrease in their innovation-oriented investments in R&D, products/services innovation and process innovation, as well as in their investments in buildings and personnel training, and closer to small decrease in their production equipment and in their marketing/advertisement investments. These firms use innovation as a central strategy for coping with the difficult economic conditions of the crisis, while at the same time they maintain to a good extent they levels of the other types of investment. The above results indicate that we are far from having a homogeneous effect of the economic crisis on the Greek firms, and we can distinguish some discrete typologies of firms with respect to the impact of the crisis on their investment; so in relevant firms' investment policies, interventions and programs more emphasis should be placed on the first cluster/typology firms, in order to mitigate their technological backwardness and obsolescence, and even survival, risks.

Furthermore, we performed ANOVA among these three clusters with respect to the above investment decrease variables, and the results are shown in the fifth and sixth columns of Table 1. We can see that there are statistically significant differences among the three clusters in all seven investment decrease variables. Therefore, the three clusters differ in the decrease they had during the economic crisis in all the examined types of investment; the F-values shown in the sixth column of Table 1 indicate that the highest differences among the three clusters are in the extent of decrease they had in the innovation-oriented investments (in R&D, products/services innovation and process innovation), followed by the personnel training investments.

4.2 Analysis of Variance of Clusters with Respect to Firms' Characteristics

We next performed ANOVA among the three clusters with respect to firms' main characteristics concerning strategic orientations, processes, personnel, technology and comparative performance, which are described in section 4. We found that the three clusters differ mostly (based on the value and the significance of the F) in comparative performance (variables COMP_PROF, COMP_SALR, COMP_MS and COMP_ROI), existence of Research & Development (variable R&D), employment of personnel having tertiary education and vocational/technical education (variables EMPL_TERT, EMPL – VOCT), process innovation (variables INN_PROC, INN_PRSD, INN_SSWM, INN_SUPP), use of cloud (variable CLOUD) and also use of Business Intelligence/Analytics and Collaboration Support systems (variables BIBA and CS). These enable an even better understanding of these three firms' clusters/typologies, and provides further support and direction for the design of investment policies, interventions and programs in economic crisis periods (including incentives, subsidies and support for the employment of tertiary education and vocational/technical education personnel, for making process innovations, as well as use cloud services, business intelligence/analytics and collaboration support systems).

4.3 Prediction of Investment Decrease in Economic Crisis

Finally we construct prediction models for the overall index of investment decrease in economic crisis INVD, which is equal to the average of the INVD1, INVD2, …, INVD7 investment decrease variables, using as predictors the abovementioned firms' main characteristics described in section 4 (concerning strategic orientations, processes, personnel, technology and comparative performance), with six Supervised Learning algorithms for predicting continuous dependent variables: Generalized Linear Modelling, Deep Learning, Decision Trees, Random Forests, Gradient Boosted Trees, Support Vector Machines [8-12]. In Fig. 2 we can see the prediction performance (mean absolute prediction error) of these algorithms. We can see that the Random Forest algorithm exhibits the lowest mean absolute error (0.788). This is a satisfactory prediction performance, given the small size of the dataset we have used (data from 363 firms), so using a larger dataset (this is feasible, as governments have such data for quite large numbers of firms) can result in a smaller mean absolute error, and therefore an even more accurate firm-level prediction of investment decrease during economic crisis.

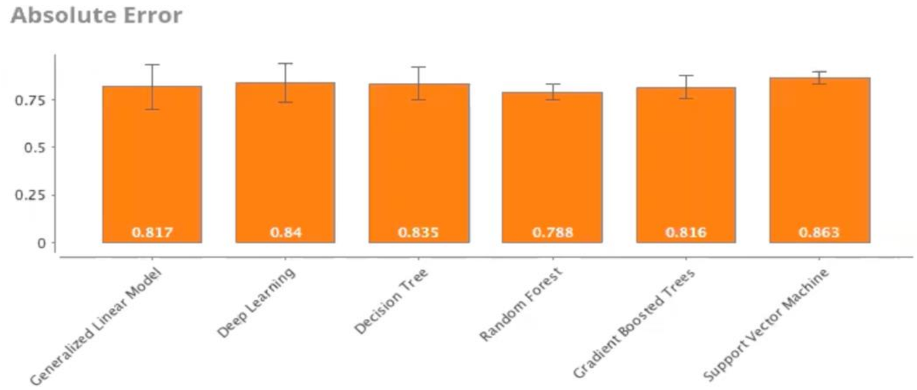

Fig. 2. Mean absolute prediction errors of the six prediction algorithms

In Table 2 we can see the top 10 predictors in terms of weight of the above best performing Random Forest algorithm, which have the highest influence on the predictions it produces of the firm-level investment decrease during economic, providing some level of 'explainability' of the predictions.

Table 2. Most influential predictors (having highest weights)

Predictor Variable	Weight
EMPL_ICT	0.291
EMPL_VOCT	0.116
IMPR_PS_P	0.109
SM_EXTCO	0.108
SM_OPCO	0.100
E_SAL	0.085
EMPL_HIGH	0.075

CL_PAAS	0.071
SM_PERS	0.069
CL_IAAS	0.052

5 Conclusions

In the previous sections has been presented an integrated methodology for leveraging government data from economic crisis periods, using on one hand Unsupervised Machine Learning techniques (clustering analysis), and on the other hand Supervised Machine Learning ones (prediction algorithms), in order to provide support for the rational design and implementation of firms' investment policies/actions for economic crisis periods. Also, a first application - validation of the proposed methodology has been presented, which gave interesting and encouraging results.

The research described in this paper has interesting implications for both research and practice. With respect to research it contributes to the growing body of knowledge concerning the use of AI in government, by developing an integrated multi-stage methodology of AI exploitation, which includes a combination of Unsupervised and Supervised Learning techniques, and also Statistical techniques as well, and providing a more comprehensive support both for the design and for the effective implementation of policies concerning one of the most severe and difficult problems that governments face: the economic crises. With respect to practice the proposed methodology can be useful to central, regional and local government agencies having competences and responsibilities in the area of economic development policies design and implementation, for the tough periods of economic recessionary crises. It can be useful also to the numerous consulting firms undertaking studies and government support in the above areas.

Future research is required towards: i) further application of the proposed methodology using larger datasets, in other national contexts experiencing economic crises of different intensities; ii) investigation of the prediction performance of other algorithms, and especially Deep Learning ones; iii) analysis of the legal aspects of the practical application of the proposed methodology, and especially with respect to the EU GDPR.

References

1. Keeley, B, Love, p.: From Crisis to Recovery - The Causes, Course and Consequences of the Great Recession. OECD Publishing, Paris (2010).
2. Knoop, T. A.: Recessions and Depressions: Understanding Business Cycles - 2nd edition. Praeger Santa Barbara, California (2015).
3. Allen, R. E.: Financial Crises and Recession in the Global Economy - Fourth Edition. Cheltenham, Edward Elgar Publications (2016).
4. Santana, M., Valle, R, Galanb, J. L.: Turnaround strategies for companies in crisis: Watchout the causes of decline before firing people. Business Research Quarterly 20(3), 206---211 (2017).
5. Izsak, K., Markianidou, P., Lukach, R., and Wastyn, A.: The impact of the crisis on research and innovation policies, European Commission, DG Research, Brussels (2013).

6. Arvanitis, S., Woerter, M.: Firm Characteristics and the Cyclicality of R&D Investments. Industrial and Corporate Change 23(5), 1141-1169 (2014).
7. Baldwin, R. and Di Mauro, B. W.: Mitigating the COVID economic crisis: Act fast and do whatever it takes. Center of Economic Policy Research Press, London (2020).
8. European Commission, Directorate-General for Economic and Financial Affairs: Economic Crisis in Europe: Causes, Consequences and Responses. Office for Official Publications of the European Communities, Luxembourg (2009).
9. Khatiwada, S.: Stimulus Packages to Counter Global Economic Crisis: A review. International Institute for Labour Studies, Geneva (2009).
10. Kalinowski, T.: Crisis management and the diversity of capitalism: fiscal stimulus packages and the East Asian (neo-)developmental state. Economy and Society 44(2), 244-270 (2015).
11. Coenen, G., Straub, R., Trabandt, M.: Gauging the Effects of Fiscal Stimulus Packages in the Euro Area. Working Paper 1483, European Central Bank, Frankfurt am Main, Germany.
12. Taylor, J.: Fiscal Stimulus Programs During the Great Recession. Economics Working Paper 18117, Hoover Institution, Stanford, CA (2018).
13. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach - 3rd Edition. Pearson, Essex, UK (2020)
14. Tan, P. N., Steinbach, M., Karpatne, A., Kumar, V.: Introduction to Data Mining – 2nd Edition. Pearson Education, USA (2019).
15. Witten, I. H., Frank, E., Hall, M. A.,, Pal, C. J.: Data mining - practical machine learning tools and techniques. Morgan Kaufmann, Amsterdam, London (2017).
16. Blum, A., Hopcroft, J., Kannan, R.: Foundations of Data Science. Cambridge University Press, Cambridge (2020).
17. Siegel, E.: Predictive Analytics. Wiley, New Jersey (2013).
18. Craglia M. (Ed.), Annoni A., Benczur P., Bertoldi P., Delipetrev P., De Prato G., Feijoo C., Fernandez Macias E., Gomez E., Iglesias M., Junklewitz H, López Cobo M., Martens B., Nascimento S., Nativi S., Polvora A., Sanchez I., Tolan S., Tuomi I., Vesnic Alujevic L.: Artificial Intelligence - A European Perspective, EUR 29425 EN. EU Publications Office, Luxembourg (2018).
19. Duan, Y., Edwards, J. S., Dwivedi, Y. K.: Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management 48, 63–71 (2019).
20. OECD: Artificial Intelligence in Society. OECD Publishing, Paris (2019).
21. Eggers, W. D., Schatsky, D., Viechnicki, P.: AI-augmented government. Using cognitive technologies to redesign public sector work. Deloitte University Press (2017).
22. Desouza, K. C.: Delivering Artificial Intelligence in Government: Challenges and Opportunities. IBM Center for The Business of Government, Washington D.C. (2018).
23. DeSousa, W. G., DeMelo, E. R. P., De Souza Bermejo, P. H., Sous Farias, R. A., Gomes, A. O: How and where is artificial intelligence in the public sector going? A literature review and research agenda. Government Information Quarterly 36(4), 101392 (2019).
24. Misuraca, G., van Noordt, C.: AI watch-Artificial Intelligence in public services. Publications Office of the European Union, Luxembourg (2020).
25. Medaglia, R., Gil-Garcia, R. and Pardo, T. A.: Artificial Intelligence in Government: Taking Stock and Moving Forward. Social Science Computer Review (in-press) (2021).
26. Van Noordt, C., Misuraca, G. (2022). 'Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. Government Information Quarterly (in-press).

27. Manzoni, M., Medaglia, R., Tangi, L., Van Noordt, C., Vaccari, L. and Gattwinkel, D. (2022). 'AI Watch. Road to the adoption of Artificial Intelligence by the public sector. Publications Office of the European Union, Luxembourg.
28. Rockoff, J. E., Jacob, B. A., Kane, T. J., Staiger, D. O.: Can you recognize an effective teacher when you recruit one? Education Finance and Policy 6(1), 43–74 (2010).
29. Chandler, D., Levitt, S. D., List, J. A.: Predicting and preventing shootings among at-risk youth. The American Economic Review 101(3), 288–292 (2011).
30. Kang, J. S., Kuznetsova, P., Luca, M., Choi, Y.: Where not to eat? Improving public policy by predicting hygiene inspections using online reviews. In: Proceedings of Empirical Methods in Natural Language Processing Conference 2013, 1443–1448 (2013).
31. Camacho-Collados, M. Liberatore, F.: A Decision Support System for predictive police patrolling. Decision Support systems 75, 25-37 (2015).
32. Ku, C. H., Leroy, G.: A decision support system: Automated crime report analysis and classification for e-government. Government Information Quarterly 31(4), 534-544 (2014).
33. Kouziokas, G. N.: The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. Transportation Research Procedia 24, 467-473 (2017).
34. Kouziokas, G., Chatzigeorgiou, A., Perakis, K.: Artificial intelligence and regression analysis in predicting ground water levels in public administration. European Water Publications, 57, 361–366 (2017).
35. Sun, T. Q., Medaglia, R.: Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly 36(2), 368-383 (2019).
36. Brynjolfsson, E. and Hitt, L.: Paradox lost? Firm-level evidence on the returns to information systems. Management Science, 42(4), 541-558 (1996).
37. Pilat, D.: The ICT Productivity Paradox: Insights from Micro Data. OECD Economic Studies 2004/1, 38-65 (2005).
38. Arvanitis, S.: Computerization, workplace organization, skilled labour and firm productivity: evidence for the Swiss business sector. Economics of Innovation and New Technology 14 (4), 225–249 (2005).
39. Arvanitis, S., Loukis, E.: Information and communication technologies, human capital, workplace organization and labour productivity: A comparative study based on firm-level data for Greece and Switzerland. Information Economics and Policy 21(1), 43-61 (2009).
40. Leavitt, H. J.: Applied organization change in industry: structural, technical, and human approaches. In Cooper, S., Leavitt, H. J., Shelly, K. (Eds.) New perspectives in organizational research, 55–71. Wiley, Chichester (1964).
41. Scott-Morton, M.S. (1991). The Corporation of the 1990s.: Oxford University Press, New York.
42. Meske, C., Bunde, E., Schneider, J., Gersch, M.: Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. Information Systems Management 39(1), 53-63 (2022).
43. Whittington, R., Regner, P., Angwin, D., Johnson, G., Scholes, K.: Exploring Strategy. 12th edn. Pearson Education Limited, Harlow, UK (2020).
44. Donaldson, L.: The Contingency Theory of Organization. Sage Publications, London, UK (2001).
45. Jones, G. R.: Organizational Theory, Design, and Change. 7th edn. Pearson Education Limited, UK (2013).