 doi:10.20944/preprints202202.0359.v1

Angelo Leogrande[1], Nicola Magaletti[2], Gabriele Cosoli[3], Alessandro Massaro[4]

e-Government in Europe. A Machine Learning Approach

Abstract

The following article analyzes the determinants of e-government in 28 European countries between 2016 and 2021. The DESI-Digital Economy and Society Index database was used. The econometric analysis involved the use of the Panel Data with Fixed Effects and Panel Data with Variable Effects methods. The results show that the value of "*e-Government*" is negatively associated with "*Fast BB (NGA) coverage", "Female ICT specialists", "e-Invoices", "Big data"* and positively associated with "*Open Data", "e-Government Users", "ICT for environmental sustainability", "Artificial intelligence", "Cloud", "SMEs with at least a basic level of digital intensity", "ICT Specialists", "At least 1 Gbps take-up", "At least 100 Mbps fixed BB take-up", "Fixed Very High Capacity Network (VHCN) coverage".* A cluster analysis was carried out below using the unsupervised k-Means algorithm optimized with the Silhouette coefficient with the identification of 4 clusters. Finally, a comparison was made between eight different machine learning algorithms using "augmented data". The most efficient algorithm in predicting the value of e-government both in the historical series and with augmented data is the ANN-Artificial Neural Network.

Keywords: Innovation, and Invention: Processes and Incentives; Management of Technological Innovation and R&D; Diffusion Processes; Open Innovation.

JEL Classification: O30; O31, O32; O33; O36.

1. Introduction

E-government is an essential element in a country's digitization process. In fact, using e-government it is possible to offer citizens a set of services that have evolved and that can significantly improve the well-being of the population. Although the goal of creating an e-government system is very widespread in European countries, however, this goal is not easily achievable since the application of an e-government model requires investment in a set of structures that are both physical - such as internet networks - and related to human capital. It follows that the possibility for states to implement economic policies and e-government requires the predisposition of wider interventions that impact on the overall level of digitization of the country. In the case presented, the value of e-government within European countries was analyzed using the DESI-Digital Economy Society Index dataset. Below is a portion of the relevant literature on the subject that highlights the role of e-government as a tool for a country's economic, social and political development.

[1] Assistant Professor at Lum University Giuseppe Degennaro and Researcher at Lum Enterprise s.r.l. Email: leogrande.cultore@lum.it. Strada Statale 100 km 18, 70010 Casamassima BA, Puglia, Italy, European Union.
[2] Chief Operation Officer-COO and Senior Researcher at Lum Enterprise s.r.l. Email: magaletti@lumenterprise.it. Strada Statale 100 km 18, 70010 Casamassima BA, Puglia, Italy, European Union.
[3] Senior IT Specialist and Solutions Architects and Researcher at LUM Enterprise s.r.l. Email: cosoli@lumenterprise.it. Strada Statale 100 km 18, 70010 Casamassima BA, Puglia, Italy, European Union.
[4] Professor at Lum University Giuseppe Degennaro, and Chief Research Officer-CRO at Lum Enterprise s.r.l. Email: massaro@lum.it. Strada Statale 100 km 18, 70010 Casamassima BA, Puglia, Italy, European Union.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

(Kamolov & Konstantinova, 2017) addresses the issue of the applications of e-government within the Russian public administration with the necessary implications in terms of maximizing the human, information, technology, and financial resources of the public economy. (Špaček, et al., 2020) analyzes the case of the application of e-government in the Czech Republic, Hungary, and Romania. The data show that the three countries are lagging in e-government applications even though the Czech Republic has a competitive advantage over the other two countries considered. (Androniceanu & Georgescu, 2021) analyze the role of e-government within the context of European nations. It appears that nations that have implemented e-government more efficiently also experience growth in public governance capacity, a reduction in corruption, greater citizen involvement and faster economic development. (Shkarlet, et al., 2020) make a comparison between different countries in terms of e-government. The authors verified that the countries that have higher levels of e-government are also the countries that have a higher level of development from a political, economic, social point of view and in terms of per capita income. Through the analysis of the e-participation index and the e-government index, the authors verified that the UK, Spain, and France have the highest levels of e-government in Europe which can be used as a reference for the low European countries. income. The most relevant activities from the point of view of e-government are the creation of e-services portals, the involvement of citizens through the dissemination of public administration activities carried out through the internet.

(Androniceanu, et al., 2020) present a cluster analysis to evaluate the diffusion of telecommunication and e-government infrastructures in Europe. The analysis shows that the countries with the highest level of e-government are the Netherlands and Austria which are opposed to Romania and Bulgaria which have, on the contrary, reduced values of the same variable. (Simonofski, et al., 2020) highlight the role of public administration values in the organization of e-government. They offer suggestions on how to create e-government services that can be optimized in three different dimensions, namely: offering better services, enabling better relations, and guaranteeing better democratic quality. (Scholta, et al., 2019) highlight the presence of a contradiction between e-government and federalism in Germany. It appears that federal states have difficulties in applying and administering e-government services. The authors therefore propose economic policy interventions that make it possible to coordinate the needs of e-government with the federal structure of the German state. (Urs, 2018) consider the role of local authorities in the application of e-government in Romania. The authors highlight the possibility of using the municipalities of Romania as a driver for the application of e-government. (Abdullah, 2021) analyzes the case of the application of e-government in Kurdistan. The authors believe that this investment is necessary to provide the country with significant economic growth. However, to implement economic and economic policies aimed at e-government it is necessary that they are made by investments in IT infrastructures, in the training of personnel operating in public structures and it is also necessary to act on the sensitivity of the population to technological innovation promoted by public administration.

(Đurašković, et al., 2021) deal with the case of the use of e-government in the context of public administration in Serbia. In particular, the authors believe that for the Serbian public administration to use e-government services it is necessary to invest in some areas such as: increasing the quality of electronic services, completing digitization, and providing for continuous updating data. (Bokayev, et al., 2021) consider the case of the digitization of public administration in Kazakhstan. The authors highlight the fact that the Kazakh government in realizing e-government has created digital infrastructures without the ability to generate value for citizens. This condition is since in the process of digitization of the public administration the needs of citizens have not been followed, but rather an attempt has been made to provide digital answers to the organizational problems of the public sector. The authors therefore propose to carry out a new phase of e-government in Kazakhstan that takes into consideration the real needs and requirements of citizens. (Pradhan & Shakya, 2018) tackles the issue of the relationship between digitization, big data, and e-government in Nepal. With reference to big data, the authors point out that such data can be both structured and unstructured. The authors believe that the use of big data, including those produced by smart phones, is necessary to create e-government services that are useful to the population. (Thompson, et al., 2020) analyzes the case of applying the degree of security of e-government websites in a comparison between Australia and Thailand. The authors checked 800 pages of 40 websites and found that e-government lacks high levels of security for citizens. In fact, only half of the Australian sites and a third of the Thai sites are equipped with a cryptographic system. From the point of view of vulnerability, it appears that there is no significant difference between the Australian e-government sites and the Thai e-government sites. (Sijabat, 2019) analyze the degree of digital and technological progress of e-government websites in Indonesia. In particular, the authors verified the functioning modalities of 34 provincial e-government portals in Indonesia. The study shows that most of the e-government sites analyzed are at an early stage of technological development and therefore do not offer advanced e-government services for the population.

(Karyono & Agustina, 2019) take into consideration the use of SWOT Analysis to evaluate the efficiency in the implementation of e-government. In particular, the authors conducted qualitative interviews with thirty public employees. The authors thus wanted to use a set of tools that are generally implemented within private companies to verify the best strategies to better promote e-government among the population. The analysis highlighted that the use of an aggressive strategy has significant advantages in terms of e-government penetration. (Sundberg, 2019) analyzes the case of e-government in Sweden considering a long period of time between 1961 and 2018. In the period considered, Sweden went through three different phases of the development of e-government, namely: Automated Data Processing, Information Technology and Digitization. These phases have been deepened in the light of three different criteria, namely: professionalism, efficiency, service, and involvement. The analysis shows the presence of a long-term trend towards the creation of an information and knowledge society in which the development of e-government systems is structured on the relevance of data. The author also suggests regulatory interventions to be implemented to ensure that the digitization process can generate added value for citizens in making the public administration more efficient. (Twizeyimana, 2017) analyzes the case of e-government in Rwanda. The author highlights the fact that the application of e-government in Rwanda can be better realized using a public-private partnership. For the implementation of an efficient e-government model for the population, the possibility of citizen co-participation in the design of e-government models is also envisaged with the aim of increasing trust in technological innovation applied to the public administration. (Jussupova, et al., 2019) analyze the case of the application of e-government in Kazakhstan highlighting the social consequences of technological innovation applied to public administration services. (Kapsa & Musiał-Karg, 2020) analyzed the case of the application of e-government in Poland. Poland was penalized in the application of e-government methodologies since it was an economy of transition from the communist regime to the European market economy. However, the performance in terms of e-government of Poland appears to be good when compared to the performance of Eastern European countries even if these countries have the lowest levels of e-government in Europe. Polish citizens have the possibility to use about 600 digital services of the public administration. Specifically, the Poles have advantages in terms of efficiency that are connected to the reduction of the timeframe for the completion of the practices in public offices. Poles have the following advantages from the use of e-government: increase in electronic services offered by public bodies, improvement of the relationship between citizens and public administration employees, simplification of practices, growth in the number of citizens who use e-government services.

Finally, it must be considered that an essential element for the application of e-government consists in the penetration of broadband (Leogrande, et al., 2021), in whether the large band is fixed (Leogrande, et al., 2022) or mobile, and in the price of broadband (Leogrande, et al., 2022).

The article continues as follows: the second paragraph presents the econometric model, the third paragraph contains a correlation matrix, the fourth paragraph presents clustering, the fifth paragraph contains the machine learning analysis for prediction with augmented data, the sixth paragraph concludes. Finally, the appendix contains the analytical metric results.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

2. The Econometric Model

The estimation of the determinants of e-government was carried out using Panel Data models with Fixed Effects and Panel Data with Variable Effects. The data used were acquired from the DESI-Digital Economy and Society Index database for 28 countries[5] for the period 2016-2021. The variable e-government in the DESI-Digital Economy and Society Index refers to the macro category "*Digital Public Services*". The e-government variable consists of the following sub-variables, namely "*e-Government users*", "*Pre-Filled Forms*", "*Digital Public Services for Citizens*", "*Digital Public Services for Business*", "*Open Data*". It is possible to indicate the structure of the e-government variable in the following extended form:

$$\begin{aligned} \boldsymbol{eGovernment_{it}} &= \boldsymbol{a_1 + b_1(eGovernmentUsers)_{it} + b_2(PreFilledForms)_{it}} \\ &\quad \boldsymbol{+ b_3(DigitalPublicServicesForCitizens)_{it}} \\ &\quad \boldsymbol{+ b_4(DigitalPublicServicesForBusinesses)_{it} + b_5(OpenData)_{it}} \end{aligned}$$

$$\boldsymbol{Where\ i = 28\ e\ t = 6}$$

The model estimated in extended form is shown below:

$$\begin{aligned} \boldsymbol{eGovernment_{it}} &= \boldsymbol{a_1 + b_1(BigData)_{it} + b_2(eInvoices)_{it} + b_3(FemaleICTSpecialists)_{it}} \\ &\quad \boldsymbol{+ b_4(FastBBNGACoverage)_{it}} \\ &\quad \boldsymbol{+ b_5(FixedVeryHighCapacityNetworkVHCNCoverage)_{it}} \\ &\quad \boldsymbol{+ b_6(AtLeast100MbplsFixedBBTakeUp)_{it}} \\ &\quad \boldsymbol{+ b_7(AtLeast1GbpsTakeUp)_{it} + b_8(ICTSpecialists)_{it}} \\ &\quad \boldsymbol{+ b_9(SMEsWithAtLeastABasicLevelOfDigitalIntensity)_{it}} \\ &\quad \boldsymbol{+ b_{10}(Cloud)_{it} + b_{11}(ArtificialIntelligence)_{it}} \\ &\quad \boldsymbol{+ b_{12}(ICTForEnvironmentalSustainability)_{it}} \\ &\quad \boldsymbol{+ b_{13}(eGovernmentUsers)_{it} + b_{14}(OpenData)_{it}} \end{aligned}$$

$$\boldsymbol{Where\ i = 28\ e\ t = 6}$$

The e-Government variable is negatively associated with the following variables that is:

[5] Countries are: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, European Union, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden.

- *Big Data:* it is a variable that measures the percentage of companies that analyze big data. There is a negative relationship between the spread of big data and e-government. Within the DESI database, the value of big data is an essential component of the "*Digital Technologies for Business*" macro-variable. The presence of a negative relationship between the value of Big Data and the value of e-government may seem counterfactual. In fact, one would expect that the growth in the value of big data is positively associated with the value of e-government. However, this negative relationship makes economic, institutional, and operational sense considering that citizens hardly use big data. In fact, big data is a type of digital product-service that is requested and used in a corporate or institutional context rather than by the individual citizen. A negative relationship derives from this as the growth of big data does not impact the value of e-government as big data is generally not used by citizens to use digital services by the public administration.

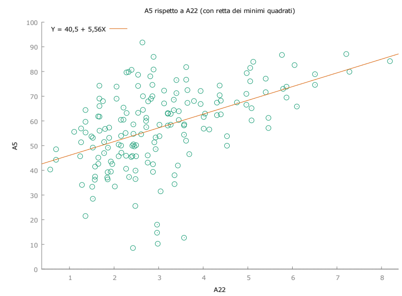


Figure 1. Relationship between the value of e-government and the value of big data.

- *e-Invoices:* it is a variable that considers the percentage of companies that use electronic invoices. There is a negative relationship between the value of e-invoices and the value of e-government. This relationship may seem counterfactual. However, it should be considered that this negative relationship can be better investigated by understanding what the differences between the users of e-government and those are that are the economic organizations that use electronic invoices. In fact, while the user of e-government is made up of citizens, the users of electronic invoices are instead made up of businesses. Furthermore, it should also be considered that while on the one hand the use of electronic invoices mainly concerns the private sector, on the other hand the use of e-government services concerns the public sector. It follows therefore that the fact that private companies use tools such as those of electronic invoicing does not automatically generate a positive impact on the ability of the government to provide e-government services. Therefore, the negative relationship between e-government and e-invoice is due to a double reason: on the one hand, the fact that e-government users are generally citizens while e-invoice users are generally companies and on the other hand, the fact that e-government concerns the organization of public services while e-invoice refers above all to the improvement of exchanges in the private sector.
- *Female ICT specialists:* it is a variable that considers the percentage of women employed as ICT specialists out of the total number of ICT specialists. The professions considered are: ICT service managers, ICT professionals, ICT Technicians, ICT Installers. There is a negative relationship between the value of women in the ICT sector and the value of e-government. Such a negative relationship might seem counterfactual. The motivation could be since the presence of women in the ICT sector is still not widespread, while on the other hand investment in e-government tends to be increasing.
- *Fast BB (NGA) coverage:* It is a variable that considers the percentage of households that use fixed broadband with a download speed of at least 30 Mbps. The technologies that are used to analyze this trend are FTTH, FTTB, Cable Docsis 3.0 and VDS. There is a negative relationship between the trend of e-government and fixed broadband with at least 30 Mbps. This negative relationship can be understood considering that to effectively carry out the e-government activity it is necessary to invest significantly in the band. wide by increasing the download speed and aiming at the limit on mobile technologies. In fact, for the State to invest in the offer of e-government services, it is necessary that the degree of digitization be high,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

the penetration of the Internet significant, and that the population can also count on fast internet services. In fact, the typical operations of e-government such as the request for documentation from the public administration require, as a prerequisite, the presence of fast internet structures. This statement is confirmed by the fact that there are other variables that are positively connected with the development of e-government such as "*At Least 1 Gbps Take Up*" and "*At Least 100 Mbps Fixed BB Take Up*".

The e-Government variable is positively associated with the following variables, namely:

- *At least 1 Gbps take-up*: Percentage of households subscribing to fixed broadband of at least 1 Gbps, calculated as overall fixed broadband take up. There is a positive relationship between the value of e-government and the value of "*At least 1 Gbps take-up*". Specifically, this positive relationship is since the 1 Gbps internet offer constitutes an endowment of the so-called fast internet associated with a very high value of the digitalization of a country. With a faster internet, even the public administration can have greater ease in offering e-government services to the population. The result is therefore a positive relationship between the spread of fast internet - albeit on a fixed network - at 1Gbps and the value of the offer of e-government services. This condition reflects the need for countries to equip themselves with adequate technological infrastructures for the administration of advanced digital services such as those relating to e-government

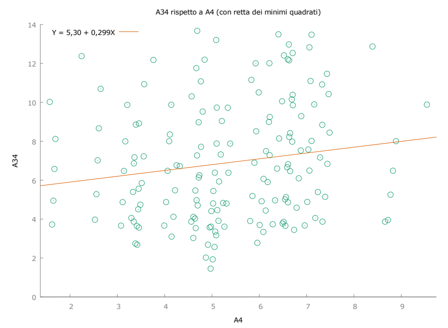


Figure 2. Relationship between "At least 1 Gbps take-up" and the value of "e-government".

- *At Least 100 Mbps fixed BB Take Up:* percentage of households subscribing to fixed broadband of at least 100 Mbps, calculated as overall fixed broadband take-up. Certainly a 100 Mbps internet network cannot be considered fast strictly speaking, however the 100 Mbps limit is generally considered to be the lower limit for applying the definition of fast internet. It follows therefore that even having 100 Mbps even though this value is in any case below 1 Gbps, it is positively associated with the value of e-government. It therefore follows that so that e-government services can be developed effectively only in the presence of significant internet investments.
- *SMEs With at Least a Basic Level of Digital Intensity*: is a variable that calculates the value of the digital intensity. The digital intensity score is based on counting how many out of 12 selected technologies are used by enterprises. A basic level requires usage of at least 4 technologies. There is a positive relationship between the value of small and medium-sized enterprises that have basic levels of digital intensity and the value of e-government. Countries in which small and medium-sized enterprises have at least a basic value of digital intensity also have a tendentially high value of e-government. This relationship may be since the growth of digital skills by companies also increases the demand for digital services from the public administration. Furthermore, companies that have at least a basic level of digital intensity been generally inserted in an economic context characterized by a medium-high level of digitization. It therefore follows that the population of countries characterized by SMEs with at least basic digital intensity can generally have the digital skills necessary to be able to use the e-government services promoted by the public administration.

- *Cloud:* is a variable that considers the percentage of enterprises purchasing at least one of the following cloud computing services: hosting of the enterprise's database, accounting software applications, CRM software, computing power. There is a positive relationship between the value of the cloud and the value of e-government. This relationship can be understood considering that generally in countries where companies use cloud services there is also a higher level of digitization of public administration. In fact, the value of the cloud in 2021 reached a maximum value of 62.09% in Finland, with a value equal to 59.14% in Sweden and an amount equal to 56.61 in Denmark. Furthermore, Denmark and Finland are also respectively second and third in value of e-government in 2021 according to the DESI dataset. Therefore, it follows that where companies have a very significant orientation to the cloud, there is also a growth in the digital culture that allows the supply and demand for e-government services.

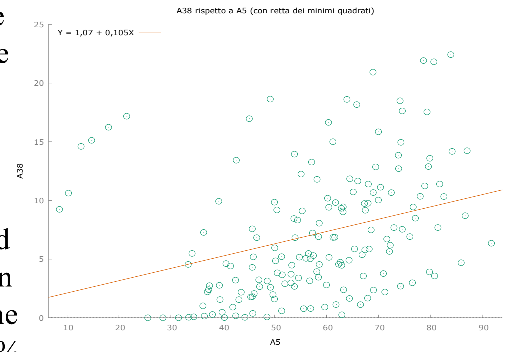


Figure 3. Relationship between e-government and the cloud.

- *Artificial Intelligence:* it is a variable that takes into consideration the percentage of enterprises that use at least 2 artificial intelligence technologies out of the total number of enterprises. There is a positive relationship between the value of artificial intelligence and the value of e-government. Specifically, the growth of companies that use technologies related to artificial intelligence tends to be positively associated with a high degree of digitization. The high degree of digitization allows businesses, citizens, and the state to offer and acquire high digital services. It therefore follows that the value of artificial intelligence tends to grow with the value of e-government. In 2021, the European countries with a high value of artificial intelligence were the Czech Republic with an amount equal to 39.74, followed by Austria with an amount equal to 36.66, and by Greece equal to a value of 33.86. In particular, the average value of artificial intelligence in European countries in 2021 was equal to an amount of 25.13.

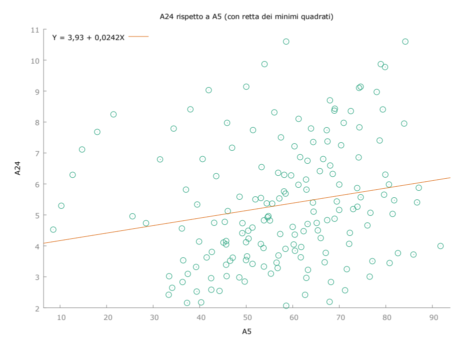


Figure 4. Relationship between the value of artificial intelligence and the value of e-government.

- *ICT For Environmental Sustainability:* It is an indicator that measures the level of support that the adoption of ICT technologies offers companies to engage in more environmentally friendly actions. The level of intensity is measured based on the number of environmental actions (maximum 10) reported by companies as facilitated using ICT. The following categorization was obtained: low intensity (0 to 4 actions), medium intensity (5 to 7 actions) and high intensity (8 to 10 actions). There is therefore a positive relationship between the value of using ICT for environmental sustainability and the value of e-government.

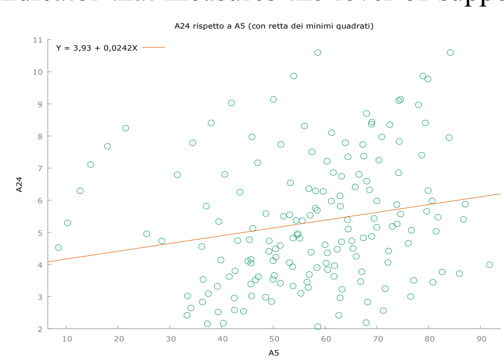


Figure 5. Relationship between the value of artificial intelligence and e-government.

 doi:10.20944/preprints202202.0359.v1

Countries that have a greater digital culture attentive to environmental issues are also able to generate more efficient e-government systems. Specifically, with reference to 2021, the countries that have reached the highest levels of ICT for environmental sustainability are indicated below: Portugal with an amount of 85.53, followed by Luxembourg with an amount of 80.00, and Finland with an amount of 76.67. On average, the value of artificial intelligence in the 28 European countries considered was equal to a value of 67.22.

- *e-Government Users:* It is a variable that considers the percentage of individuals who have used the internet in the last 12 months to interact with public authorities. There is a positive relationship between the ability of the public administration to offer services through the internet and the percentage of individuals who use the internet to interact with the public administration. Obviously, a pre-condition for this relationship consists in the fact that there are significant investments in the internet such as to be able to support both the supply and the demand for public administration services through digital. In addition, a certain degree of digitization of the population is also necessary to be able to use internet services in the sense of public administration. It therefore follows that for the proper functioning of e-government it is also necessary that investments be made both in networks and in the human capital of citizens to increase accessibility to the services offered.

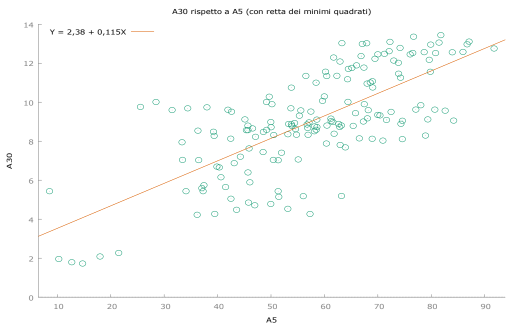


Figure 6. Relationship between "e-Government" and "e-Government users".

- *Open Data:* It is an indicator that measures the presence of a policy aimed at operators at a national level. The indicator also considers the political, economic, and social impact of open data and their characteristics such as the functionality, availability, and use of data. There is a positive relationship between the value of e-government and the value of open data. It should be borne in mind that many data that are generally defined as open data are produced by the public sector in relation to the provision of e-government services. It follows therefore that the growth of e-government generates a positive impact in terms of data production that can be made available to the population in the version of open data. In fact, when citizens use the services of the public administration, both from a legal point of view and from a health or education point of view, a flow of data is generated which, adequately anonymized, constitutes a relevant part of the open data.

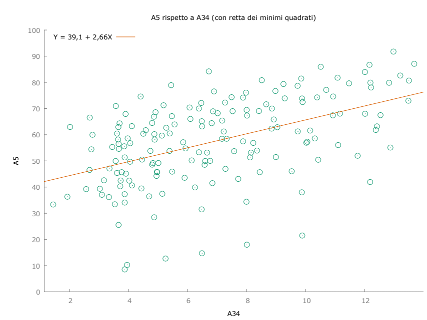


Figure 7. Relationship between Open Data and e-Government.

- *ICT Specialists:* It is a variable that takes into consideration the percentage of ICT specialists employed at national level. Both ICT professionals and ICT technicians are intended for employees in the ICT sector. There is a positive relationship between the value of the distribution of ICT specialists and the value of e-government. This relationship can be better understood considering that where the number of ICT specialists increases there is also a higher degree of digitization with the possibility of also offering complex digital services such as those connected to e-government. It follows therefore that the investment in human capital, through the creation of professional figures who are skilled in the ICT sector, allows to raise

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

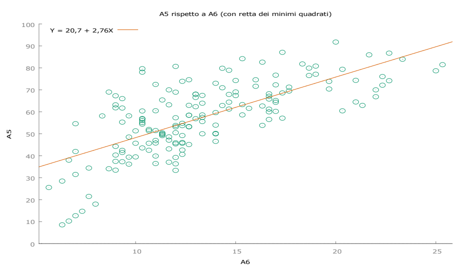


Figure 8. The Relationship between e-Government and ICT Specialist.

the technological content of the presets that are also administered in the e-government. Furthermore, the possibility of creating an efficient e-government model requires the presence of expert professionals who can carry out the digitization of the public administration. In this sense, IT companies are also very important, which can also carry out digitization activities of public administration services through tenders and concessions, increasing the offer of digital e-government products.

- *Fixed Very High-Capacity Network (VHCN) coverage:* is a variable that considers the percentage of households that are covered by a "*Very High Capacity Network*" type network. The technologies that are considered envisage the use of FTTH and FTTB for the period 2015-2018 and FTTH, FTTB and Cable Docsis 3.1 for 2021. There is a positive relationship between the value of "*Very High-Capacity Network*" and the value of the e-government. This relationship can be better understood considering that the possibility of offering e-government services also depends on the development of an efficient and fast internet network. Specifically, the use of high-capacity networks can increase the possibility of offering e-government services, allowing citizens to be able to access effectively. It follows that investment in e-government must be preceded by economic policies that are aimed at providing the country with the adequate internet infrastructures necessary to support the provision of digital services to the population.

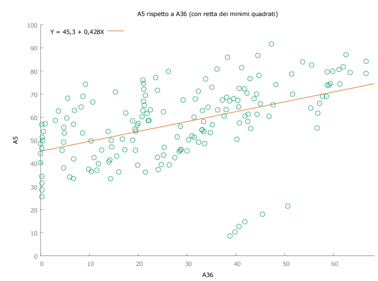


Figure 9. The Relationship between e-Government and Fixed Very High Capacity Network.

Variables	Label	Panel Data Fixed Effects		Panel Data Random Effects		Mean
		Coefficient	P-value	Coefficient	P-value	
Const		26,1545	***	25,4726	***	25,81355
Big data	A22	-0,2583	***	-0,25657	***	-0,25743
Cloud	A23	0,27715	***	0,27517	***	0,2761605
Artificial intelligence	A24	0,38987	***	0,36678	***	0,3783235
ICT for environmental sustainability	A25	0,85831	***	0,79611	***	0,8272105
e-Invoices	A26	-0,21322	**	-0,22425	**	-0,2187375
e-Government Users	A30	0,90465	***	0,92494	***	0,914792
Open Data	A34	1,55806	***	1,53479	***	1,546425
Fast BB (NGA) coverage	A35	-0,06706	***	-0,06156	***	-0,06431155
Fixed Very High Capacity Network (VHCN) coverage	A36	0,01946	***	0,01926	***	0,01936
At least 100 Mbps fixed BB take-up	A38	0,07122	**	0,06891	**	0,070065
At least 1 Gbps take-up	A39	0,11885	**	0,11688	**	0,1178605
SMEs with at least a basic level of digital intensity	A47	0,16187	***	0,18254	***	0,172206
ICT Specialists	A6	0,14971	**	0,16502	**	0,157365
Female ICT specialists	A7	-0,09352	**	-0,09767	**	-0,0955964

Figure 10. Summary of econometric results for estimating the value of e-government.

3. Correlation Analysis

A correlation matrix was then used to verify those relationships present in the interior of the analyzed dataset. The correlation index varies between a value of -1 and a value of 1. Through the analysis of

the correlation matrix, it is possible to identify the most significant relationships from the point of view of e-government. Only those relationships that are particularly significant in positive terms are analyzed below - that is, close to the value +1.
In fact, it appears that the value of e-government is strongly correlated with the following variables, namely:

- *SMEs with at least a basic level of digital intensity with a correlation value equal to an amount of 0.7458;*
- *e-government users with a correlation value equal to an amount of 0.72;*
- *ICT Specialists with a correlation value equal to an amount of 0.70.*

On average, the variables of the model considered are correlated with the value of e-government for an amount equal to a value of 0.481. In the analysis of the correlation matrix, e-government is positively correlated with all the variables of the analyzed model.
Furthermore, there are further particularly significant relationships within the correlation matrix. For example, considering the trend of the Cloud variable, it is possible to verify the presence of the following positive relationships, namely:

- *SMEs with at least a basic level of digital intensity with a correlation index value equal to an amount of 0.8525;*
- *ICT Specialists with a correlation index value equal to an amount of 0.7721.*

Furthermore, there is a positive relationship between the value of the "*SMEs with at least a basic level of digital intensity*" variable and the value of the "*ICT Specialists*" variable with a correlation index value of 0.8228.
In the next part, on the other hand, what are the negative relationships between the variables of the model are analyzed. However, only those correlation index values are analyzed that show minimum values for the distribution considered.
In particular, the variable that presents a greater number of negative correlations is "*Female ICT Specialists*". In fact, this variable is negatively correlated with the following variables, namely:

- *Big data*: with a correlation index equal to -0.0341;
- *Government Users*: with a correlation index equal to -0.0192;
- *Fast BB NGA coverage*: with a correlation index equal to an amount of -0.0932;
- *At Least 1 Gbps take up*: with a correlation index equal to an amount of -0.0623;
- *SMEs with at least a basic level of digital intensity* with an amount of the correlation index equal to a value of -0.1655;
- *ICT specialists* with a correlation index value equal to a value of -0.039;

Furthermore, there is a further negative correlation between the value of "*Artificial intelligence*" and the value of "*e-Government Users*".

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

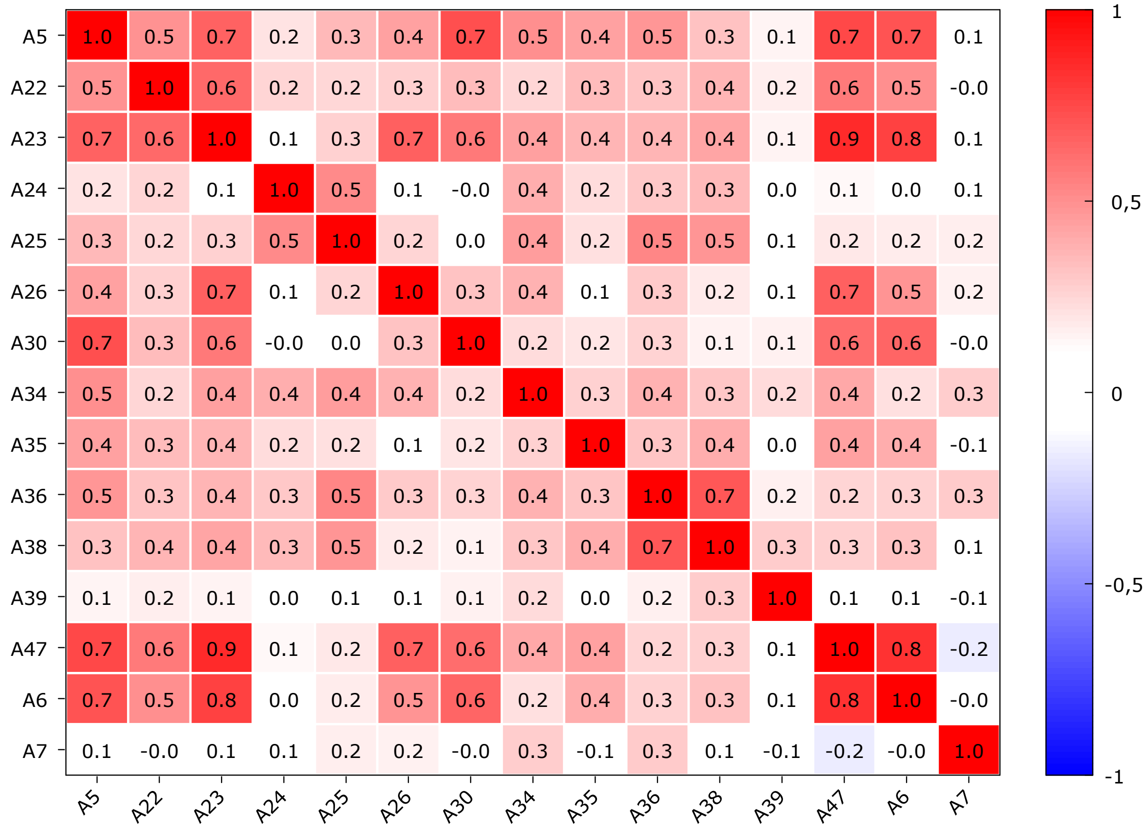


Figure 11. Correlation matrix.

From the analysis of the correlation matrix, it is therefore possible to verify that, contrary to the results that were produced in the regression analysis, the value d of the correlations between e-government and the observed variables tends to be substantially positive. In fact, the average of the correlations between the e-government and the variables is equal to 0.4817. However, there is a further notation that can be added to the analysis of the model with reference to one of the variables or "*Female ICT Specialists*" which is negatively correlated with 6 of the 14 variables of the model.

4. Clusterization

The supervised k-Means noon algorithm was then used to cluster the data. Through clustering, groupings of countries have been highlighted that have the same levels of value due to the diffusion of e-government. However, since the k-Means algorithm is an unsupervised algorithm, it is necessary to use tools to optimize the number of clusters. In this regard, the Silhouette coefficient was used. The Silhouette coefficient is a number that varies between -1 and 1. To optimize the algorithm's capacity to identify the correct number of clusters, it is necessary to choose the positive Silhouette coefficient and the closest to 1. In this regard, 7 clusters have been selected having an increasing number of groups from 2 to 8. The results in terms of Silhouette gave the following orientation:

- 4 clusters with a Silhouette coefficient value of 0.595;
- 2 and 3 clusters with a Silhouette coefficient value of 0.58;
- 5 clusters with a Silhouette coefficient value of 0.576;

- 6 clusters with a Silhouette coefficient value of 0.503;
- 7 clusters with a Silhouette coefficient value of 0.463;
- 8 clusters with a Silhouette coefficient value of 0.423.

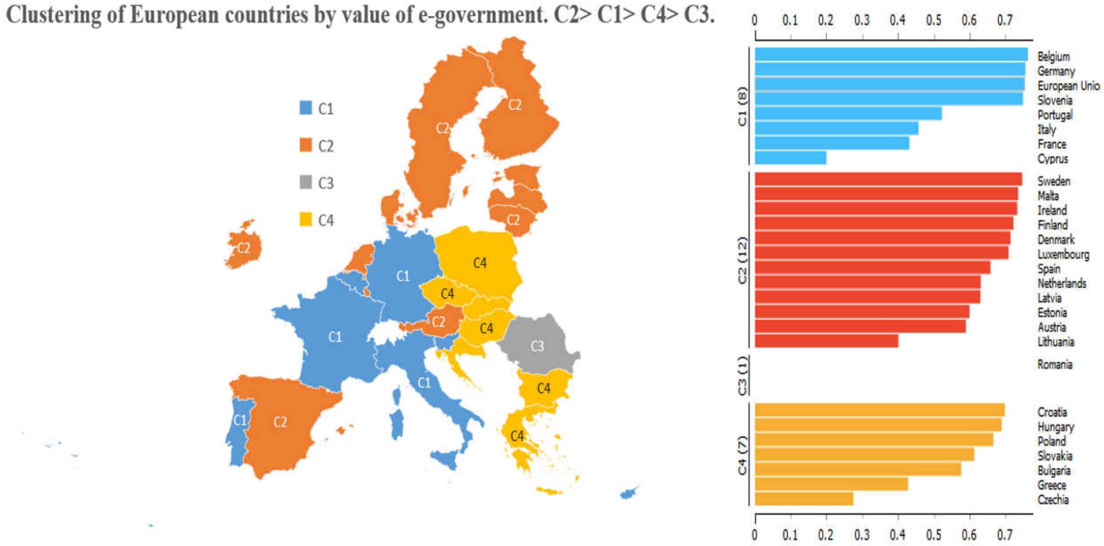


Figure 12. Clustering of European countries by value of e-government. C2> C1> C4> C3.

Therefore, the 4-cluster model was chosen, which optimizes the performance of the k-Means algorithm in terms of the Silhouette coefficient. The following clusters were therefore identified as follows:

- *Cluster 1*: Belgium, Cyprus, European Union, France, Germany, Italy, Portugal, Slovenia;
- *Cluster 2*: Austria, Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Spain, Sweden;
- *Cluster 3*: Romania;
- *Cluster 4*: Bulgaria, Croatia Czechia, Greece, Hungary, Poland, Slovakia

To sort the various clusters, the value of the median of e-government for each single cluster was calculated, generating the following sorting of clusters, that is: cluster 2 with an e-government median value equal to 81.642, followed by cluster 1 with an e-government median value equal to 67.734, cluster 4 with an e-government median value equal to 53.723 and cluster 3 with an e-government median value equal to 21.488. The following order therefore derives from it, namely: C2> C1> C4> C3. As is evident from the analysis of clustering, there is an evident co-relationship between countries that have higher levels of e-government value and countries that have low e-government values. Specifically, the contrast between two areas of Europe is evident: Western Europe which has very high values for the value of e-government and Eastern Europe with low e-government values. This contrast highlights the more general infrastructural gap that Eastern European countries experience in comparison with Western European countries. In fact, the delay in the supply and in the question of e-government services reflects a much wider gap concerning digitization and in a broad sense the endowment of human capital. In this regard, it is necessary that European policy makers invest significantly in Eastern European countries to allow the creation of both networks and human capital, as well as the capacity for action of the public administration that are such as to allow the effective development of e -government.

5. Machine Learning and Prediction with Augmented Data

A prediction was made below by using the machine learning algorithms. Eight different machine learning algorithms were used. The proposed approach is based on the augmented data (Massaro, et al., 2021). As in many applications the initial dataset to process is poor and it is necessary to increase the dataset records to increase predictive algorithm performance (Massaro, 2021). The use of augmented data allows to refine the predictive activity by increasing the available data base (Massaro, et al., 2019). Specifically, in the first stage of data processing, the dataset is partitioned as 70 % of training dataset and 30 % of testing dataset. The second stage of the prediction considers as training dataset the whole dataset (100 % of the records), and as testing dataset the last predicted results (adding 30 % of records). The used method is sketched in the following figure:

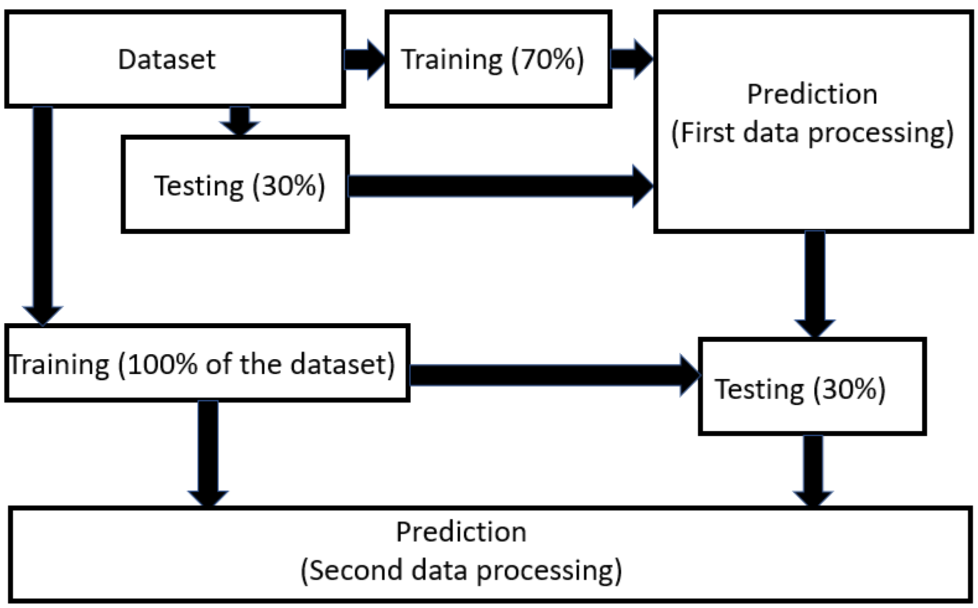


Figure 13. Methodology used for prediction with augmented data.

The algorithms were trained with 70% of the available data while the remaining 30% was used for the actual prediction. The algorithms have been sorted according to their performance through the maximization capacity of the R-squared and through the possibility of minimization of statistical errors or "*Mean Absolute Error*", "*Mean Squared Error*", "*Root Mean Squared Error*", "*Mean Signed Difference"*. Based on the analysis carried out, the following ordering of the algorithms evaluated based on their performance capacity results, i.e.:

- *ANN-Artificial Neural Network* with a payoff value of 6;
- *PNN-Probabilistic Neural Network* with a payoff value of 12;
- *Tree Ensembles and Random Forest* with a payoff value of 26;
- *Gradient Boosted Trees Regression* with a payoff value of 28;
- *Linear Regression* with a payoff value of 37;
- *Polynomial Regression* with a payoff value of 40;
- *Simple Regression* with a payoff value of 41.

In particular, the application of the ANN-Artificial Neural Network produces the following results.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

- *Austria* with an increase from an amount of 79.83 up to a value of 81.63 or equal to an amount of 1.801 units equal to an amount of 2.25%;
- *Belgium* with an increase from a value of 65.827 up to a value of 79.359 or equal to an amount of 13.532 units equal to a value of 20.55%;
- *Bulgaria* with an increase from an amount of 56.048 up to a value of 76.417 or equal to an amount of 20.369 equal to an amount of 36.34%;
- *Croatia* with an increase from a value of 51.97 to a value of 74.635 or equal to a value of 22.663 units equal to an amount of 43.60%;
- *Cyprus* with an increase from an amount of 61.821 up to a value of 77.958 units or equal to a value of 16.137 equal to an amount of 26.10%;
- *Czechia* with an increase from an amount of 58.58 units up to a value of 76.979 or equal to a value of 18,390 units equal to a value of 31.388%;
- *Denmark* with a decrease from an amount of 87,086 units up to a value of 82,485 units equal to a value of -5.28%;
- *Estonia* with a decrease from an amount of 91.76 up to a value of 82.854 or equal to a variation of -8.908 units equal to a variation of -9.70%;
- *European Union* with an increase from a value of 68,053 units up to a value of 79,820 units or equal to a variation of 11,768 units equal to an amount of 17.29%;
- *Finland* with a decrease from an amount of 86.71 units up to a value of 82.582 units or equal to a value of -4, 125 units equal to a variation of -4.7%;
- *France* with an increase from an amount of 72.992 units up to a value of 80.471 units equal to an amount of 7.47 units equal to a value of 10.24%;
- *Germany* with an increase from a value of 67.47 units up to a value of 79.307 units or equal to a variation of 11.833 units equal to a value of 17.53%
- *Greece* with an increase from a value of 41.942 up to a value of 72.211 units or equal to an amount of 30.270 equal to an amount of 72.17%;
- *Hungary* with an increase from an amount of 49,159 up to a value of 74,983 units or equal to an amount of 25,824 units equal to a value of 52.53%;
- *Ireland* with a decrease from an amount of 82.608 up to a value of 81.998 units equal to a change of -0.611 units equal to an amount of -0.739%;
- *Italy* with an increase from an amount of 63,194 units up to a value of 78,328 units or equal to an amount of 15,133 units equal to a variation of 23.94%;
- *Latvia* with an increase from an amount of 79,630 units up to a value of 81,875 units equal to a variation of 2,245 units or equal to a variation of 2.81%
- *Lithuania* with an increase from a value of 78.04 units up to a value of 81.52 units equal to a variation of 3.47 units equal to a variation of 4.44%;
- *Luxembourg* with an increase from an amount of 79.362 units up to a value of 82.164 equal to a change of 2.803 units equal to a change of 3.53%;
- *Malta* with a decrease from an amount of 84.195 units up to a value of 82.319 units equal to a variation of -2.22%;
- *Netherlands* with an increase from a value of 79.902 units up to a value of 81.636 units or equal to a variation of 1.734 units equal to an amount of 2.169%;
- *Poland* with an increase from an amount of 55.10 units up to a value of 75.310 units or equal to a variation of 20.210 units equal to a value of 36.67%;
- *Portugal* with an increase from an amount of 68,948 units up to a value of 80,495 units equal to a variation of 11,547 units equal to a value of 16.74%;

- *Romania* with an increase from a value of 21,488 units up to a value of 69,418 units equal to a variation of 47,390 units equal to a variation of 223.04%
- *Slovakia* with an increase from a value of 53,723 units up to a value of 76,166 units or equal to a variation of 22,443 units equal to a variation of 41.77%;
- *Slovenia* with a variation from 67.994 units up to a variation of 79,066 units or equal to a variation of 11,072 units equal to a variation of 16.28%;
- *Spain* with an incremental variation from a value of 80.676 units or equal to a variation of 81.74 equal to a variation of 1.071 units equal to a variation of 1.327 &,
- *Sweden* with a decrease from a value of 83.945 units up to a value of 82.217 units or equal to a variation of -1.728 units equal to a variation of -2.05%;

On average, the value of e-government in the countries considered + increased from an amount of 68,503 units up to an amount of 79,141 units equal to a variation of 10,638 units equal to 15.52%. Subsequently, a further prediction was made on the augmented data obtained by adding the predicted series to the historical series of data relating to e-government. In particular, the following ranking of the algorithms was obtained based on performance capacity, that is:

- *ANN-Artificial Neural Network* with a payoff value of 5;
- *PNN-Probabilistic Neural Network* with a payoff value of 13;
- *Linear Regression* and *Simple Regression Tree* with a payoff value of 19;
- *Tree Ensemble Regression* with a payoff value of 25;
- *Polynomial Regression* with a payoff value of 29;
- *Gradient Boosted Trees* with a payoff value of 30;
- *Random Forest Regression* with a payoff value of 40.

It follows therefore that using the ANN-Artificial Neural Network algorithm it is possible to realize the following prediction, that is:

- *Belgium* with an increase from a value of 79.36 units up to a value of 81.84 units with a value of 2.48 units up to a value of 3.13%;
- *Bulgaria* with an increase from a value of 76.42 units up to a value of 80.90 units or equal to a value of 4.48 units equal to a value of 5.86%;
- *Estonia* with a decrease from a value of 82.85 units up to a value of 82.46 units or equal to a variation of -0.39 units equal to a variation of -0.47%;
- *European Union* with an increase from a value of 79.82 units up to a value of 81.89 units or equal to a variation of 2.07 units equal to a variation of 2.60%;
- *Germany* with a variation of 79.31 units up to a value of 81.80 units equal to a variation of 2.50 units equal to an amount of 3.15%;
- *Greece* with an increase from an amount of 72.21 up to a value of 79.69 units up to a value of 7.48 units equal to a variation of 10.36%;

On average, the value of the increase increased from an amount of 79.14 units to a value of 81.43 units or equal to a change of 2.29 units equal to an amount of 2.89%. It is possible to make a comparison in terms of performance between the prediction made on the data in historical series and the prediction made with augmented data. In particular, the following considerations arise, namely:

- *R-squared*: with an increase equal to an amount to a value of 0.095%;
- *Mean Absolute Error*: with a statistical error growth of 80.44%;

- *Root Mean Squared Error*: with an increase in the statistical error equal to an amount of 44.44%;
- *Mean Signed Difference*: with a reduction of the statistical error equal to -3.56%;
- *Mean Absolute Percentage Error*: with a statistical error growth of 94.9%.

In summary, in the transition between the prediction made with the values of the historical series and the prediction with augmented data, there is an average increase in statistical errors equal to an amount of 67.84% equivalent to a growth in absolute value equal to a value of 0.0093994344.

6. Conclusion

The following article analyzes the determinants of e-government in 28 European countries between 2016 and 2021. The DESI-Digital Economy and Society Index database was used. E-government is essential as it improves the relationship between citizens and public administration and is a determining element of economic growth as it improves the relationship between businesses and public administration. However, to create an efficient e-government model, it is necessary that there are requirements defined at the country level that refer both to networks and to the formation of human capital. The econometric analysis involved the use of the Panel Data with Fixed Effects and Panel Data with Variable Effects methods. The results show that the value of "*e-Government*" is negatively associated with "*Fast BB (NGA) coverage", "Female ICT specialists", "e-Invoices", "Big data"* and positively associated with "*Open Data", "e-Government Users", "ICT for environmental sustainability", "Artificial intelligence", "Cloud", "SMEs with at least a basic level of digital intensity", "ICT Specialists", "At least 1 Gbps take-up", "At least 100 Mbps fixed BB take-up", "Fixed Very High Capacity Network (VHCN) coverage"*. A cluster analysis was carried out below using the unsupervised k-Means algorithm optimized with the Silhouette coefficient with the identification of 4 clusters. Clustering makes it possible to verify the presence of groupings of countries within the European Union which are defined by value of e-government. The contrast between Western Europe - which also includes southern countries - and Eastern Europe is particularly evident. In fact, while Western Europe has high levels of e-government, Eastern Europe has low levels of e-government. It follows that the European economic policies should invest more to create the conditions for an improvement of e-government in Eastern Europe, both directly and indirectly by acting on the context conditions such as investment in networks and human capital.
Finally, a comparison was made between eight different machine learning algorithms using "augmented data". The most efficient algorithm in predicting the value of e-government both in the historical series and with augmented data is the ANN-Artificial Neural Network. The prediction made both with the historical series and with augmented data highlights the presence of a positive trend in e-government. In fact, in the first prediction made with historical data there is an average forecast of growth of e-government equal to an amount of 15.52% for European countries while in the prediction made with augment given the average growth is equal to 2.89%. However, it should be considered that between the prediction made with historical data and the prediction made with augmented data, even though there is a growth in the R-square, there is also a growth in the average of statistical errors equal to a value of 67.8% equivalent to a growth in absolute value equal to a value of 0.0093994344. In summary, the analysis shows that to increase the presence of e-government at the country level, it is necessary to invest in the infrastructures of the Internet network and in human capital. Furthermore, it is necessary to take action to remove the gap between Eastern Europe and Western Europe in terms of e-government. Finally, the prediction suggests an increasing trend for e-government in European countries.

7. Bibliography

Abdullah, N. N., 2021. Factors Affecting the Adoption of e-Government in Public Sector Organizations of Kurdistan Region. *Public administration aspects,* 9(2), pp. 15-27.

Androniceanu, A. & Georgescu, I., 2021. E-Government in European Countries, a Comparative Approach Using the Principal Components Analysis.. *NISPAcee Journal of Public Administration and Policy,* 14(2), pp. 65-86.

Androniceanu, A., Kinnunen, J. & Georgescu, I., 2020. E-Government clusters in the EU based on the Gaussian Mixture Models. *Administratie si Management Public,* Volume 35, pp. 6-20.

Bokayev, B. et al., 2021. Transforming E-government in Kazakhstan: A Citizen-Centric Approach. *The Innovation Journal,* 26(1), pp. 1-21.

Đurašković, J., Viduka, D. & Gajić-Glamočlija, M., 2021. The Use of E-government from the Perspective of Biggest Business Entities in Serbia. *Journal of Information and Organizational Sciences,* 42(1), pp. 39-53.

Jussupova, G., Bokayev, B. & Zhussip, D., 2019. Digital government maturity as a technologically new e-government maturity model: Experience of Kazakhstan. *Proceedings of the 2019 3rd International Conference on E-commerce, E-Business and E-Government,* pp. 10-14.

Kamolov, S. & Konstantinova, A., 2017. E-government: way of modernization and efficiency enhancement of public governance. *Право и управление,* 1(XXI век,), pp. 13-21.

Kapsa, I. & Musiał-Karg, M., 2020. E-Government in Poland in public data and opinions of Poles: empirical analysis. *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance,* pp. 419-429.

Karyono, O. & Agustina, K., 2019. Determining the Priority Strategy in the Implementation of E-Government Through Swot Analysis Model. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal),* 2(2), pp. 66-74.

Leogrande, A., Costantiello, A. & Laureti, L., 2021. The Broadband Penetration in Europe. *Journal of Applied Economic Sciences,* 16(3).

Leogrande, A., Magaletti, N., Cosoli, G. & Massaro, A., 2022. Broadband Price Index in Europe. *SSRN Electronic Journal,* p. 10.2139/ssrn.4038261.

Leogrande, A., Magaletti, N., Cosoli, G. & Massaro, A., 2022. Fixed Broadband Take-Up in Europe. *Available at SSRN 4034298.*

Massaro, A., 2021. Implementation of a Decision Support System and Business Intelligence Algorithms for the Automated Management of Insurance Agents Activities. *International Journal of Artificial Intelligence and Applications (IJAIA),* Volume 12.

Massaro, A. et al., 2019. LSTM DSS Automatism and Dataset Optimization for Diabetes Prediction. *Appl. Sci. ,* 9(https://doi.org/10.3390/app9173532), p. 3532.

Massaro, A., Panarese, A., Giannone, D. & Galiano, A., 2021. Augmented Data and XGBoost Improvement for Sales Forecasting in the Large-Scale Retail Sector. *Appl. Sci.,* https://doi.org/10.3390/app11177793(11), p. 7793.

Pradhan, P. & Shakya, S., 2018. Big Data Challenges for e-Government Services in Nepal. *Journal of the Institute of Engineering,* 14(1), pp. 216-222.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

Scholta, H. et al., 2019. Fast and Federal—Policies for Next-Generation Federalism in Germany. *Proceedings of the 52nd Hawaii International Conference on System Sciences.*

Shkarlet, S. et al., 2020. Comparative analysis of best practices in e-Government implementation and use of this experience by developing countries. *Administratie si Management Public,* Volume 34, pp. 118-136.

Sijabat, R., 2019. Digitalization of Local Government in the Decentralized Era: An Insight of the Stage of E-government Across Provinces in Indonesia. *Proceedings of the 2nd International Conference on Inclusive Business in the Changing World (ICIB 2019),* pp. 287-296.

Simonofski, A. et al., 2020. The Influence of Public Values on User Participation in e-Government: An Exploratory Study. *Proceedings of the 53rd Hawaii International Conference on System Sciences.*

Špaček, D., Csótó, M. & Urs, N., 2020. Questioning the real citizen-centricity of e-government development: Digitalization of G2C services in selected CEE countries. *NISPAcee Journal of Public Administration and Policy,* 13(1), pp. 213-243.

Sundberg, L., 2019. From Automatic Data Processing to Digitalization: What is Past is Prologue. *International Conference on Electronic Government,* Issue Springer, Cham., pp. 31-42.

Thompson, N., Mullins, A. & Chongsutakawewong, T., 2020. Does high e-government adoption assure stronger security? Results from a cross-country analysis of Australia and Thailand. *Government Information Quarterly,* 1(101408), p. 37.

Twizeyimana, J. D., 2017. User-centeredness and usability in e-government: A reflection on a case study in Rwanda. *Proceedings of the Internationsl Conference on Electronic Governance and Open Society: Challenges in Eurasia,* pp. 172-178.

Urs, N., 2018. E-government development in Romanian local municipalities: A complicated story of success and hardships. *Transylvanian Review of Administrative Sciences,* Volume 55, pp. 118-129.

8. Appendix

Modello 20: Effetti fissi, usando 168 osservazioni
Incluse 28 unità cross section
Lunghezza serie storiche = 6
Variabile dipendente: A5

	Coefficiente	*Errore Std.*	*rapporto t*	*p-value*	
const	26,1545	1,03464	25,28	<0,0001	***
A22	−0,258295	0,0886981	−2,912	0,0042	***

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

A23	0,277149	0,0745037	3,720	0,0003	***
A24	0,389868	0,130066	2,997	0,0033	***
A25	0,858313	0,179110	4,792	<0,0001	***
A26	−0,213221	0,103557	−2,059	0,0416	**
A30	0,904645	0,0720240	12,56	<0,0001	***
A34	1,55806	0,0557284	27,96	<0,0001	***
A35	−0,0670605	0,0205198	−3,268	0,0014	***
A36	0,0194589	0,00705432	2,758	0,0067	***
A38	0,0712196	0,0284054	2,507	0,0134	**
A39	0,118846	0,0479132	2,480	0,0144	**
A47	0,161868	0,0370269	4,372	<0,0001	***
A6	0,149706	0,0641767	2,333	0,0212	**
A7	−0,0935193	0,0419809	−2,228	0,0277	**

Media var. dipendente	57,42698	SQM var. dipendente	16,27955
Somma quadr. residui	38,44075	E.S. della regressione	0,552345
R-quadro LSDV	0,999131	R-quadro intra-gruppi	0,995590
LSDV F(41, 126)	3535,239	P-value(F)	5,5e-175
Log-verosimiglianza	−114,4946	Criterio di Akaike	312,9892
Criterio di Schwarz	444,1957	Hannan-Quinn	366,2392
rho	0,081337	Durbin-Watson	1,391625

Test congiunto sui regressori -
Statistica test: F(14, 126) = 2031,69
con p-value = P(F(14, 126) > 2031,69) = 4,66301e-141

Test per la differenza delle intercette di gruppo -
Ipotesi nulla: i gruppi hanno un'intercetta comune
Statistica test: F(27, 126) = 986,525
con p-value = P(F(27, 126) > 986,525) = 1,47306e-133

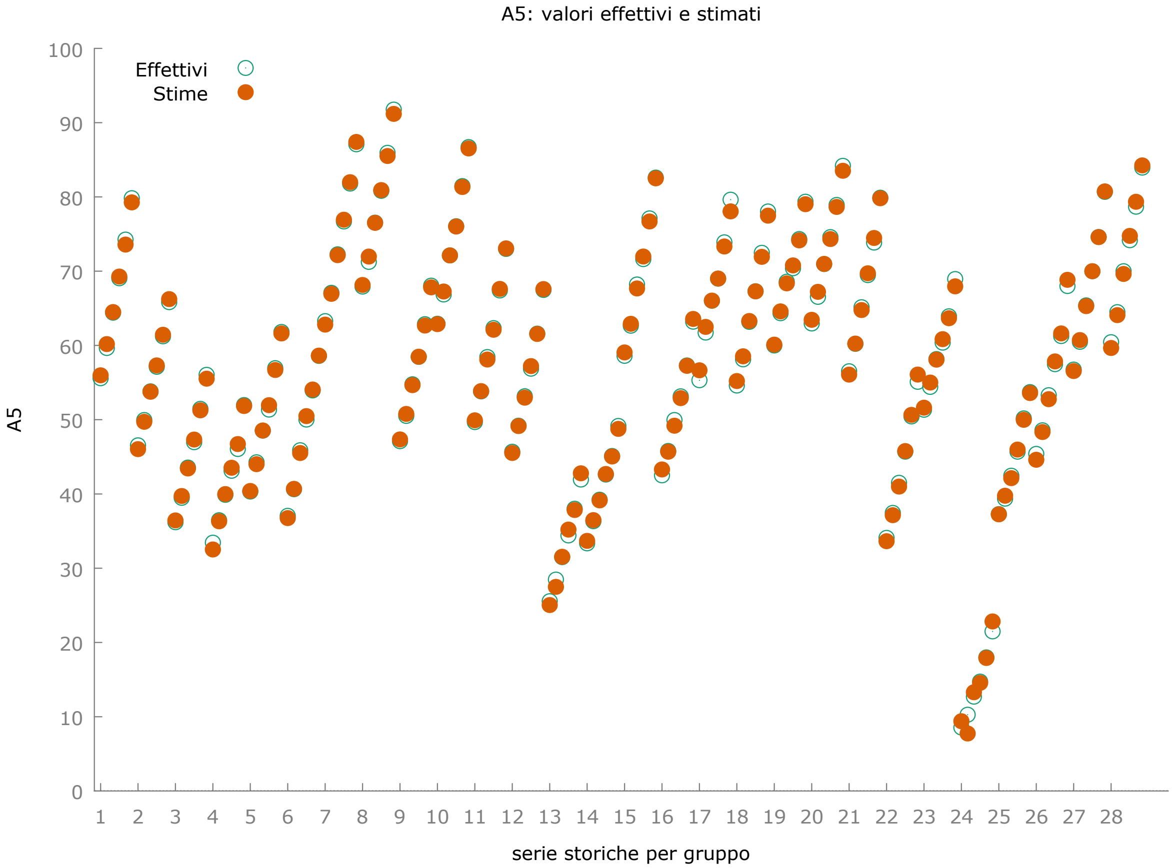
Test non-parametrico di Wald per l'eteroschedasticità -
Ipotesi nulla: le unità hanno in comune la varianza dell'errore
Statistica test asintotica: Chi-quadro(28) = 1254,6
con p-value = 1,41027e-246

Test per la normalità dei residui -
Ipotesi nulla: L'errore è distribuito normalmente
Statistica test: Chi-quadro(2) = 45,8514
con p-value = 1,10534e-010

Test di Wooldridge per l'autocorrelazione in dati panel -

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

Ipotesi nulla: Non c'è autocorrelazione del prim'ordine (rho = -0.5)
Statistica test: F(1, 27) = 23,771
con p-value = P(F(1, 27) > 23,771) = 4,25535e-005
Test CD di Pesaran per dipendenza fra unità -
Ipotesi nulla: Assenza di dipendenza fra unità
Statistica test asintotica: z = 4,54727
con p-value = 5,43458e-006



Modello 21: Effetti casuali (GLS), usando 168 osservazioni
Incluse 28 unità cross section
Lunghezza serie storiche = 6

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

Variabile dipendente: A5

	Coefficiente	*Errore Std.*	*z*	*p-value*	
const	25,4726	1,97338	12,91	<0,0001	***
A22	−0,256565	0,0924663	−2,775	0,0055	***
A23	0,275172	0,0774755	3,552	0,0004	***
A24	0,366779	0,134715	2,723	0,0065	***
A25	0,796108	0,183249	4,344	<0,0001	***
A26	−0,224254	0,107820	−2,080	0,0375	**
A30	0,924939	0,0748648	12,35	<0,0001	***
A34	1,53479	0,0575874	26,65	<0,0001	***
A35	−0,0615626	0,0213104	−2,889	0,0039	***
A36	0,0192611	0,00734549	2,622	0,0087	***
A38	0,0689104	0,0295855	2,329	0,0198	**
A39	0,116875	0,0499996	2,338	0,0194	**
A47	0,182544	0,0373074	4,893	<0,0001	***
A6	0,165024	0,0666673	2,475	0,0133	**
A7	−0,0976735	0,0436734	−2,236	0,0253	**

Media var. dipendente	57,42698	SQM var. dipendente	16,27955
Somma quadr. residui	17111,10	E.S. della regressione	10,54092
Log-verosimiglianza	−626,7572	Criterio di Akaike	1283,514
Criterio di Schwarz	1330,374	Hannan-Quinn	1302,532
rho	0,081337	Durbin-Watson	1,391625

Varianza 'between' = 70,915
Varianza 'within' = 0,305085
Theta usato per la trasformazione = 0,973232

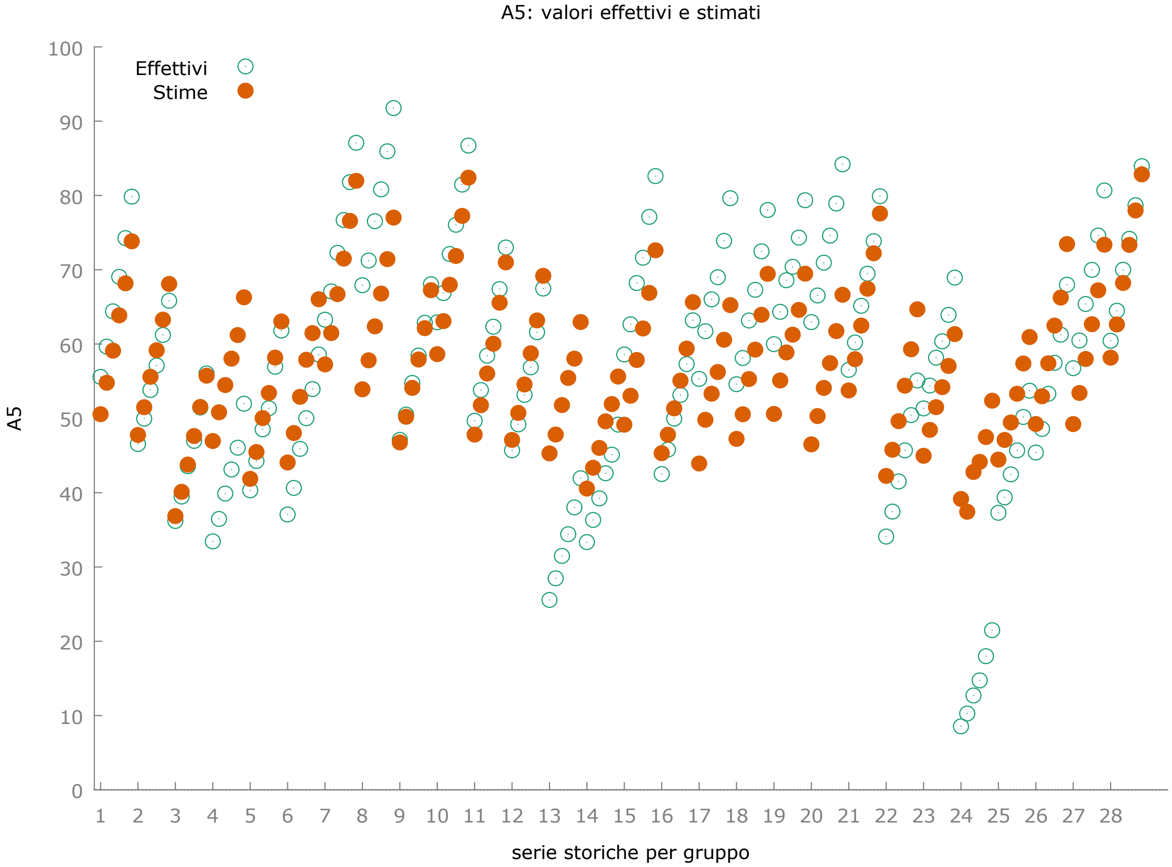
Test congiunto sui regressori -
Statistica test asintotica: Chi-quadro(14) = 26109,5
con p-value = 0

Test Breusch-Pagan -
Ipotesi nulla: varianza dell'errore specifico all'unità = 0
Statistica test asintotica: Chi-quadro(1) = 261,82
con p-value = 6,88245e-059

Test di Hausman -
Ipotesi nulla: le stime GLS sono consistenti
Statistica test asintotica: Chi-quadro(14) = 33,7544
con p-value = 0,00223984

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

Test per la normalità dei residui -
Ipotesi nulla: L'errore è distribuito normalmente
Statistica test: Chi-quadro(2) = 24,6949
con p-value = 4,34084e-006
Test CD di Pesaran per dipendenza fra unità -
Ipotesi nulla: Assenza di dipendenza fra unità
Statistica test asintotica: z = 4,31571
con p-value = 1,59089e-005

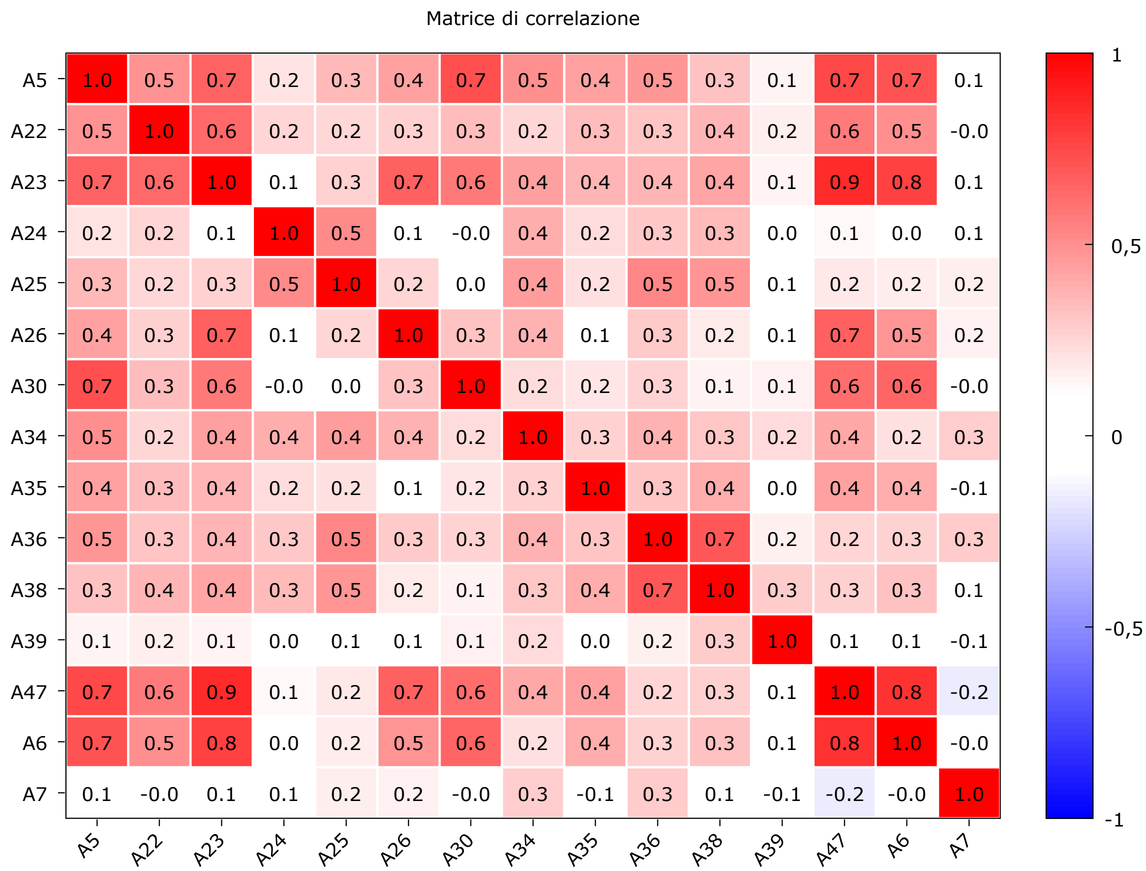


2. Correlation analysis

Coefficienti di correlazione, usando le osservazioni 1:1 - 28:6

Valore critico al 5% (per due code) = 0,1515 per n = 168

A5	A22	A23	A24	A25	
1,0000	0,4846	0,6525	0,1990	0,3471	A5
	1,0000	0,6297	0,2461	0,2413	A22
		1,0000	0,0588	0,2618	A23
			1,0000	0,5125	A24
				1,0000	A25
A26	A30	A34	A35	A36	
0,4312	0,7209	0,4946	0,4317	0,4712	A5
0,2649	0,3346	0,2436	0,3353	0,3078	A22
0,6619	0,5719	0,4343	0,3612	0,3608	A23
0,0656	-0,0312	0,3862	0,2240	0,2939	A24
0,2473	0,0208	0,4453	0,2095	0,5251	A25
1,0000	0,3128	0,3697	0,0899	0,2854	A26
	1,0000	0,2215	0,1901	0,2563	A30
		1,0000	0,2627	0,3722	A34
			1,0000	0,3044	A35
				1,0000	A36
A38	A39	A47	A6	A7	
0,3144	0,1372	0,7458	0,7083	0,0866	A5
0,3622	0,1596	0,5704	0,4952	-0,0341	A22
0,4171	0,1414	0,8525	0,7721	0,0722	A23
0,3437	0,0134	0,1230	0,0381	0,0555	A24
0,4732	0,0805	0,1807	0,1656	0,1578	A25
0,1740	0,0754	0,6600	0,4762	0,1505	A26
0,1449	0,1485	0,6150	0,6448	-0,0192	A30
0,3014	0,2257	0,4067	0,2099	0,2743	A34
0,3885	0,0353	0,4297	0,3897	-0,0932	A35
0,6940	0,1513	0,2483	0,2569	0,2840	A36
1,0000	0,2795	0,2652	0,3133	0,0877	A38
	1,0000	0,0784	0,0941	-0,0623	A39
		1,0000	0,8228	-0,1665	A47
			1,0000	-0,0399	A6
				1,0000	A7



3. Principal component analysis

Analisi delle componenti principali			
n = 168			
Auto-analisi della matrice di covarianza			
Componente	Autovalore	Proporzione	Cumulata
1	647,2956	0,6141	0,6141
2	263,6997	0,2502	0,8643
3	66,2093	0,0628	0,9272
4	31,1207	0,0295	0,9567
5	13,0924	0,0124	0,9691
6	10,0681	0,0096	0,9787
7	7,9249	0,0075	0,9862
8	3,4579	0,0033	0,9895
9	2,9818	0,0028	0,9923
10	2,3246	0,0022	0,9945
11	1,9494	0,0018	0,9963
12	1,3341	0,0013	0,9976
13	1,2417	0,0012	0,9988
14	0,7431	0,0007	0,9995

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

15	0,5388	0,0005	1,0000

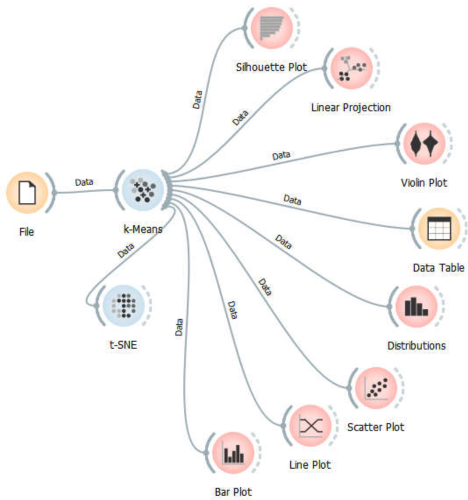
Autovettori (pesi della componente)

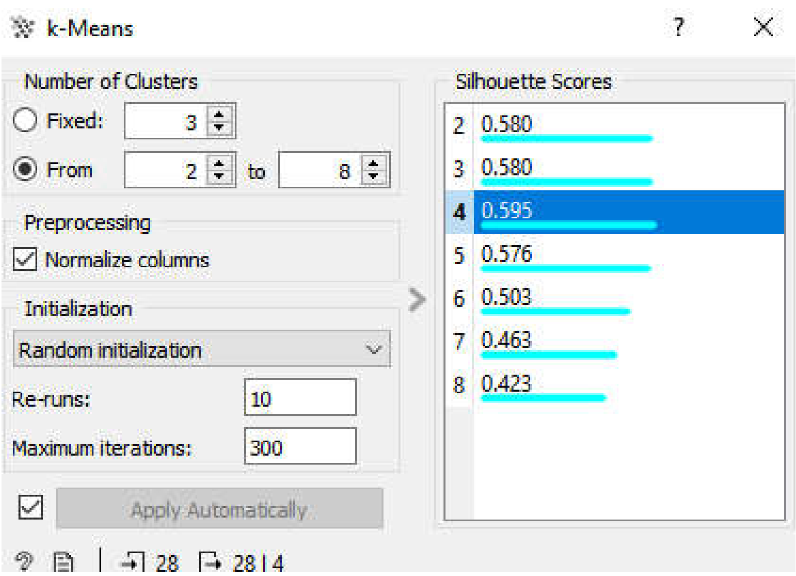
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
A5	0,584	0,121	0,781	-0,018	0,089	0,044	-0,070
A22	0,032	0,009	-0,026	-0,015	0,053	-0,002	0,010
A23	0,103	0,049	-0,108	0,046	0,135	-0,129	0,272
A24	0,019	-0,024	-0,003	-0,061	0,088	-0,204	-0,268
A25	0,023	-0,032	0,005	-0,013	0,068	-0,118	-0,103
A26	0,038	0,018	-0,060	0,092	-0,037	-0,136	0,056
A30	0,068	0,035	0,077	0,076	0,026	0,196	0,147
A34	0,062	-0,012	0,029	-0,006	0,046	-0,687	-0,270
A35	0,124	0,004	-0,066	-0,939	-0,270	-0,041	0,119
A36	0,461	-0,828	-0,190	0,125	-0,208	0,066	-0,007
A38	0,109	-0,164	-0,185	-0,236	0,888	0,019	-0,060
A39	0,007	-0,007	0,002	-0,002	0,104	-0,007	-0,056
A47	0,617	0,505	-0,531	0,135	-0,103	-0,046	-0,082
A6	0,127	0,086	-0,045	0,002	0,169	0,266	0,527
A7	0,006	-0,068	0,090	0,090	0,003	-0,564	0,661

	PC8	PC9	PC10	PC11	PC12	PC13	PC14
A5	-0,020	0,110	0,061	-0,026	0,009	0,002	0,046
A22	-0,063	0,039	-0,102	-0,559	-0,518	-0,020	-0,146
A23	0,264	-0,029	0,053	-0,709	0,235	-0,122	0,281
A24	-0,648	-0,239	-0,490	-0,134	0,021	0,153	0,295
A25	-0,258	-0,087	0,133	-0,239	0,444	0,020	-0,788
A26	-0,001	0,150	0,095	-0,003	0,562	0,444	0,313
A30	0,423	-0,384	-0,705	0,074	0,202	-0,001	-0,177
A34	0,277	-0,497	0,209	0,125	-0,072	-0,185	0,055
A35	0,058	-0,014	-0,033	-0,014	0,052	0,047	-0,004
A36	-0,001	-0,037	0,012	-0,001	-0,024	-0,003	0,010
A38	0,072	0,196	-0,045	0,162	0,023	-0,054	0,001
A39	0,230	-0,158	0,104	-0,059	-0,293	0,848	-0,136
A47	-0,025	0,109	-0,062	0,117	-0,071	-0,002	-0,087
A6	-0,327	-0,593	0,331	0,106	-0,038	-0,006	0,076
A7	-0,120	0,274	-0,218	0,183	-0,143	0,067	-0,160

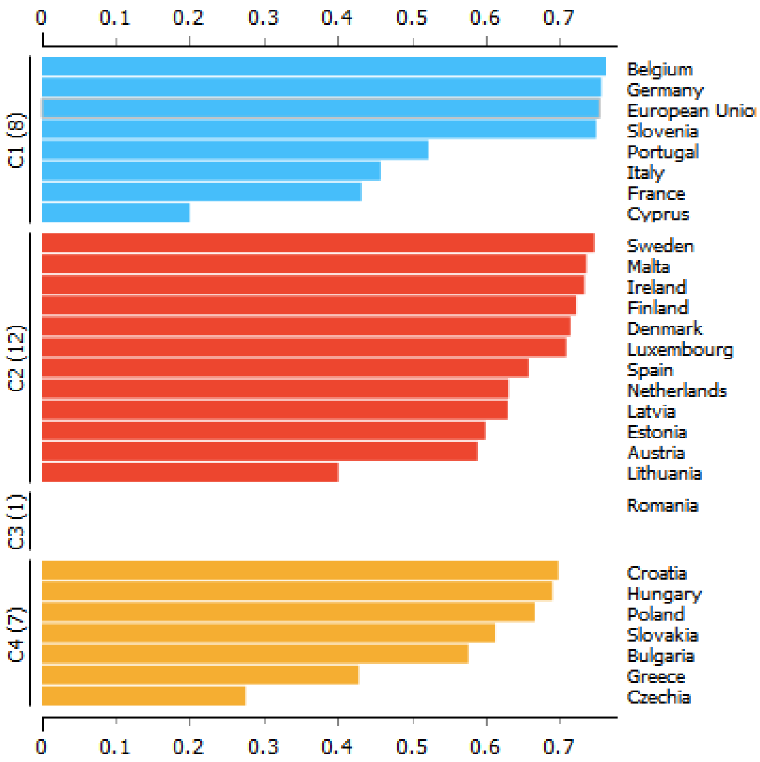
	PC15
A5	-0,010
A22	0,614
A23	-0,373
A24	-0,168
A25	-0,034
A26	0,566
A30	0,143
A34	0,150
A35	0,021
A36	-0,014
A38	0,061
A39	-0,264
A47	-0,042
A6	0,108
A7	-0,062

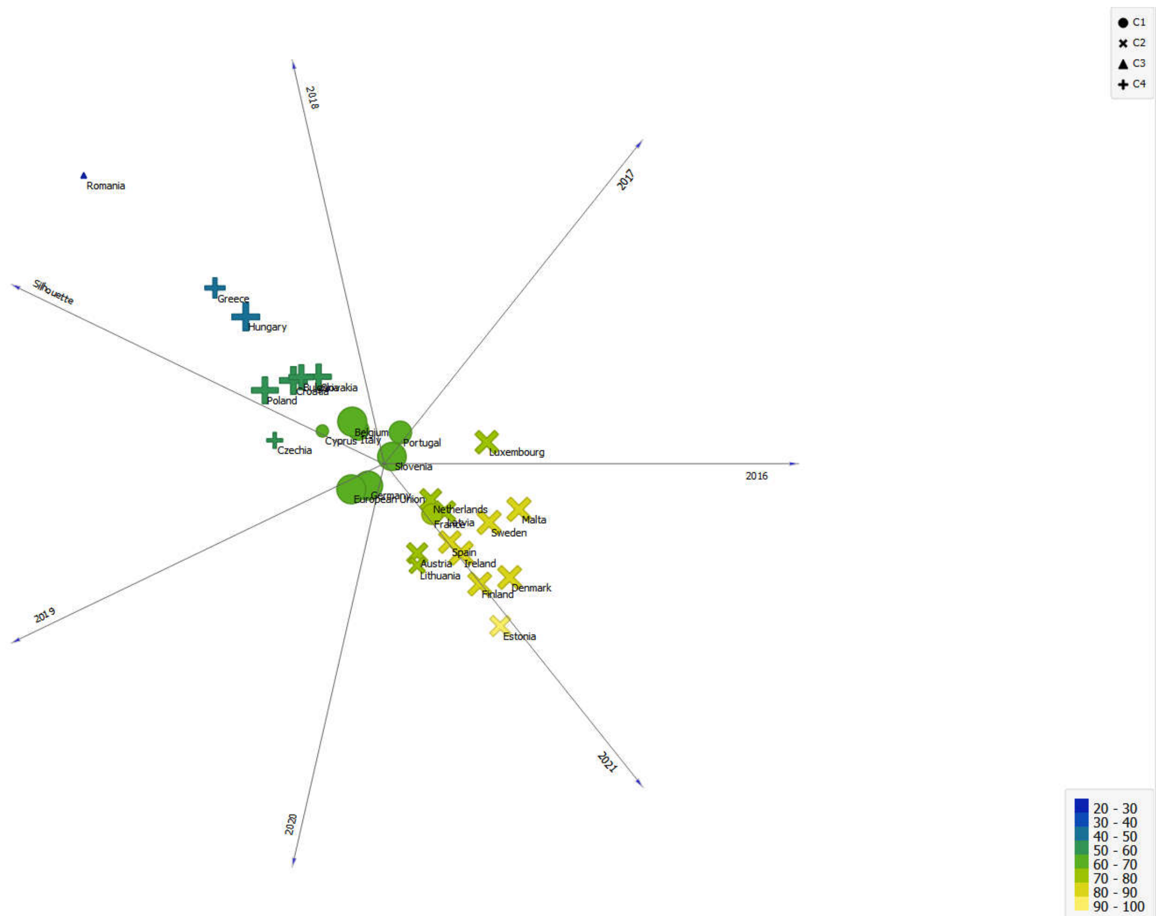
4. Clusterization



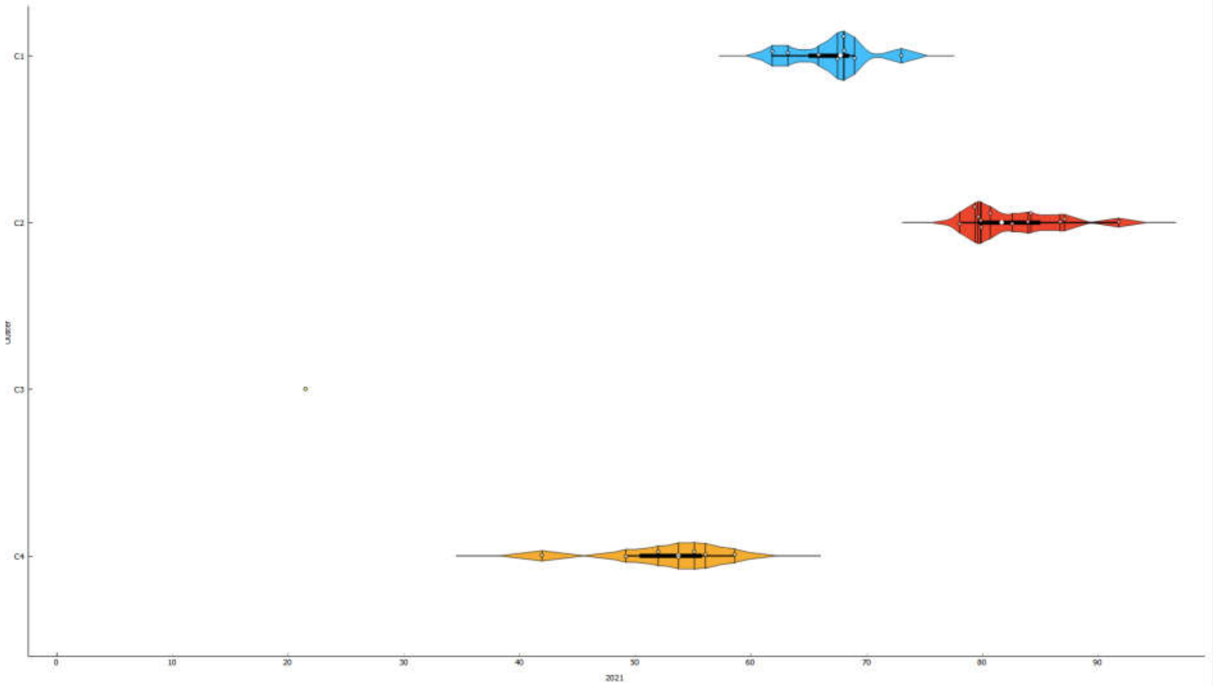


Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

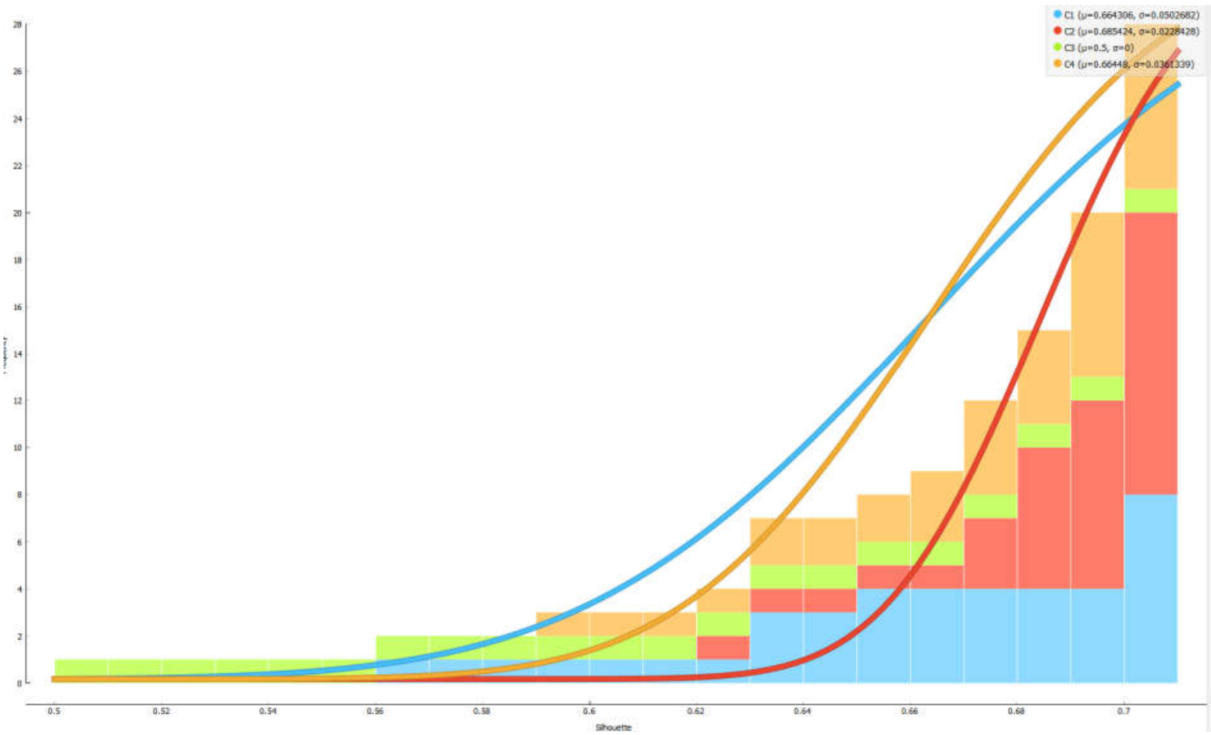


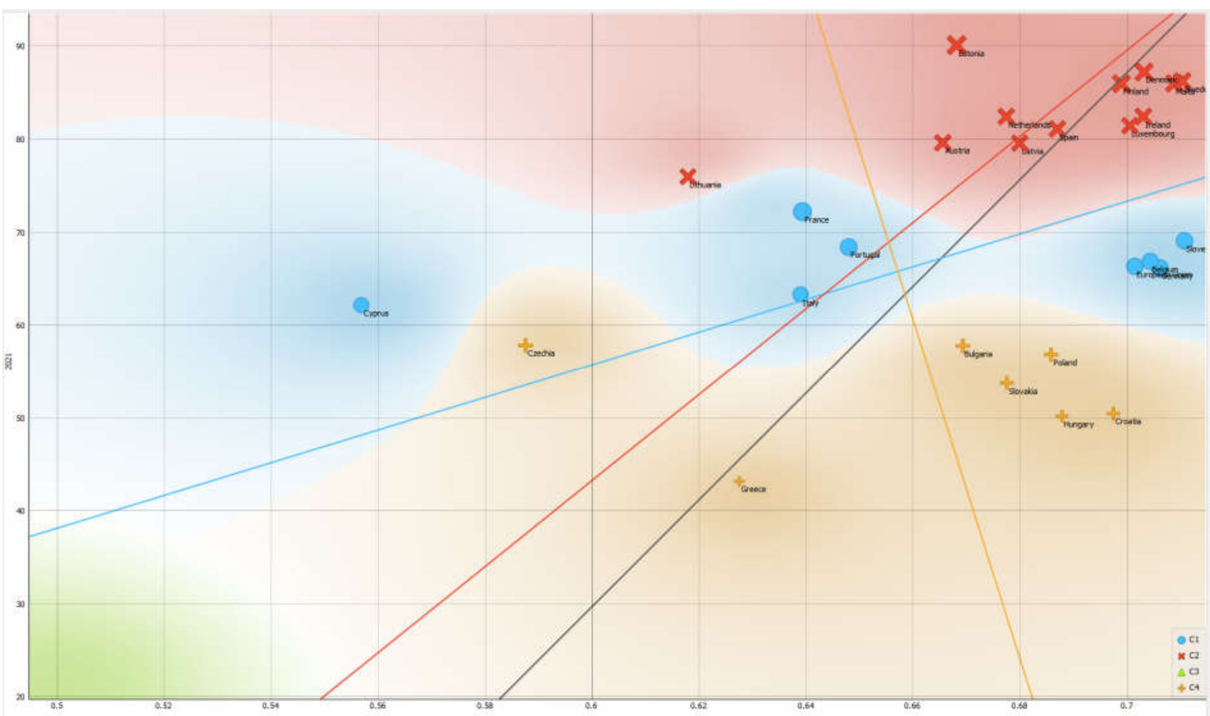


Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1

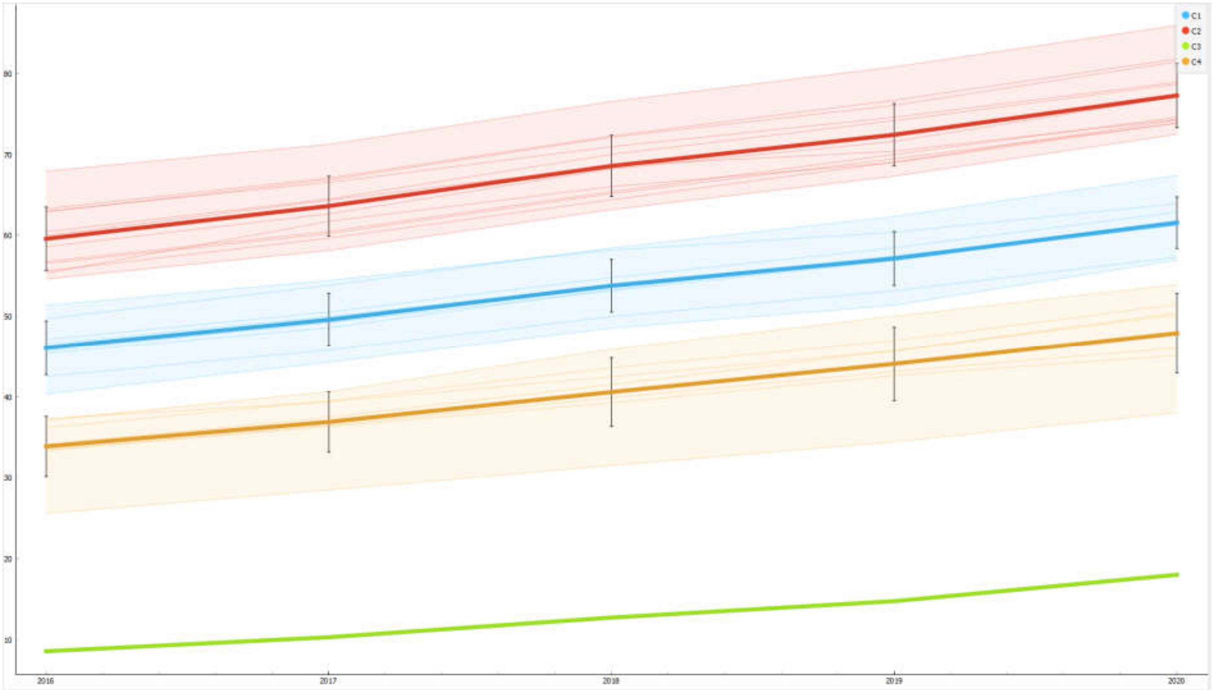


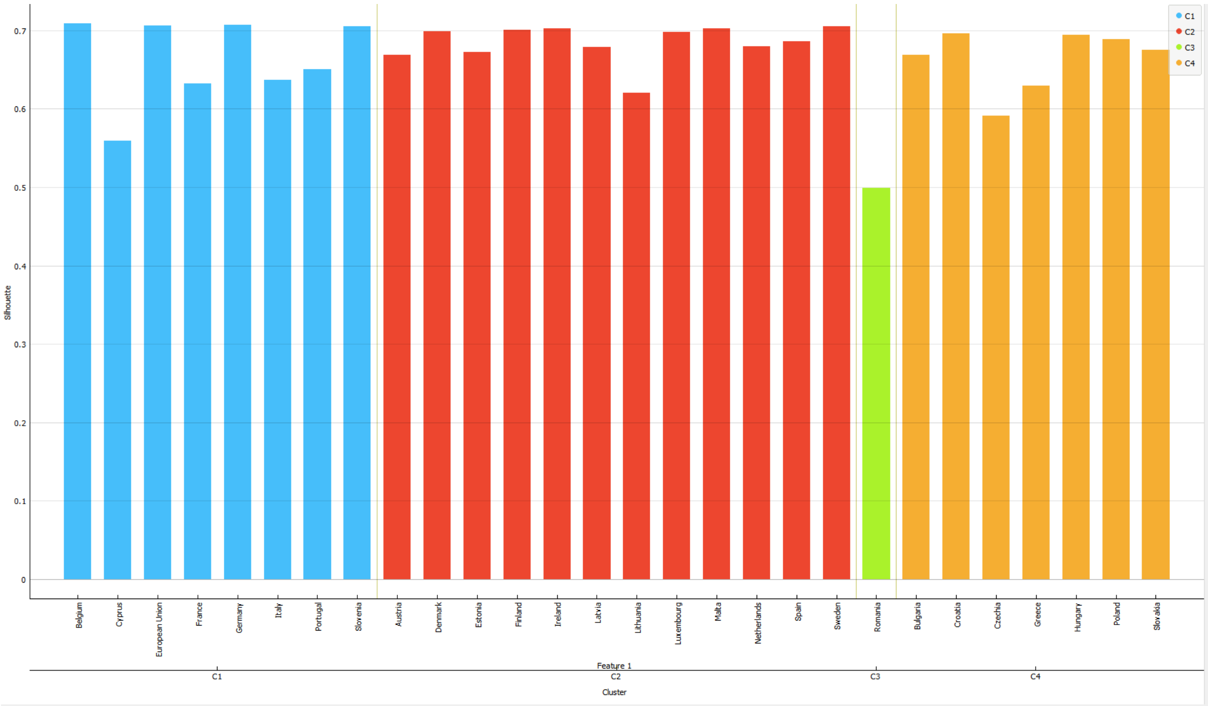
	2021	Feature 1	Cluster
2	65.8265	Belgium	C1
5	61.8211	Cyprus	C1
9	68.0527	European Union	C1
11	72.992	France	C1
12	67.4736	Germany	C1
16	63.1941	Italy	C1
23	68.9478	Portugal	C1
26	67.9937	Slovenia	C1
1	79.8336	Austria	C2
7	87.0858	Denmark	C2
8	91.7629	Estonia	C2
10	86.7161	Finland	C2
15	82.6082	Ireland	C2
17	79.6304	Latvia	C2
18	78.0492	Lithuania	C2
19	79.3615	Luxembourg	C2
20	84.1947	Malta	C2
21	79.9021	Netherlands	C2
27	80.6759	Spain	C2
28	83.9453	Sweden	C2
24	21.4884	Romania	C3
3	56.0482	Bulgaria	C4
4	51.9722	Croatia	C4
6	58.5889	Czechia	C4
13	41.9415	Greece	C4
14	49.1589	Hungary	C4
22	55.0999	Poland	C4
25	53.723	Slovakia	C4

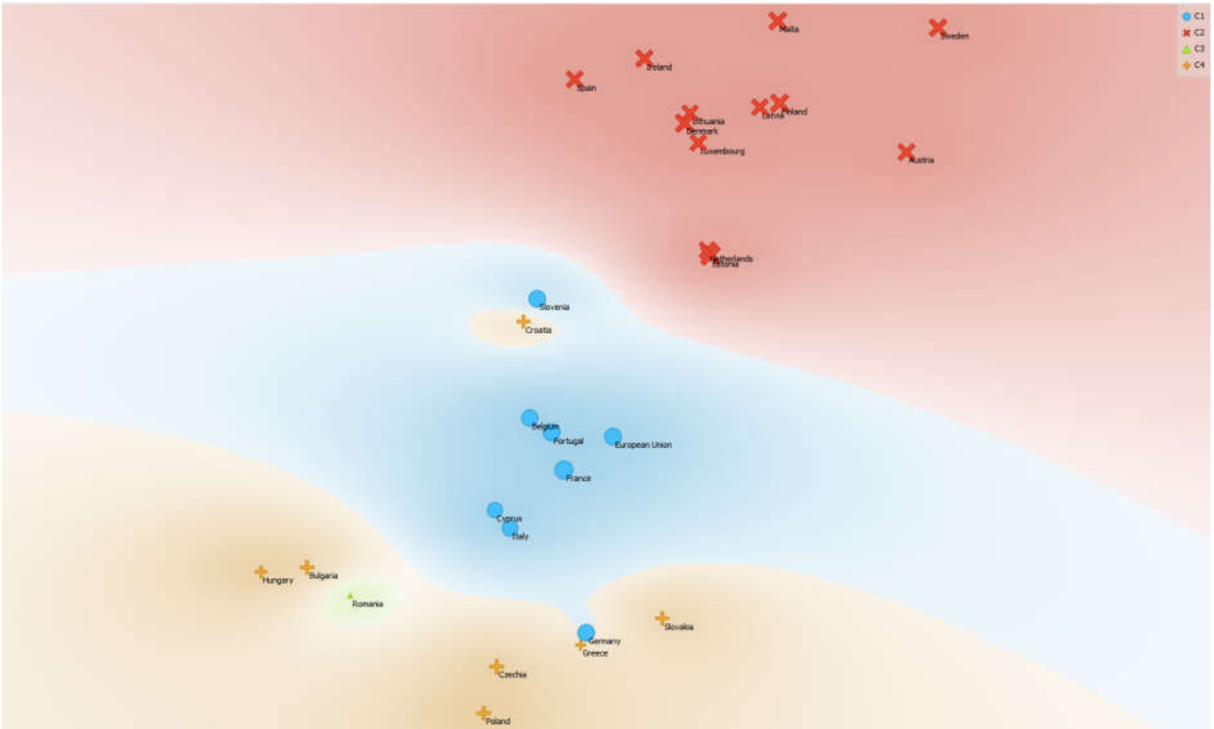

C1 (μ=0.664306, σ=0.0502682)
C2 (μ=0.685424, σ=0.0228428)
C3 (μ=0.5, σ=0)
C4 (μ=0.66448, σ=0.0361339)
Silhouette


Estonia
Finland
Austria
Netherlands
Latvia
Spain
Ireland
Luxembourg
Lithuania
France
Portugal
Italy
Cyprus
Czechia
Bulgaria
Poland
Slovakia
Hungary
Croatia
Greece
2021
Silhouette
C1
C2
C3
C4

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 February 2022 doi:10.20944/preprints202202.0359.v1







5. Machine learning and prediction

Country	2021	Prediction 1	Var assoluta	Var percentuale
Austria	79,834	81,634	1,801	2,256
Belgium	65,827	79,359	13,532	20,557
Bulgaria	56,048	76,417	20,369	36,342
Croatia	51,972	74,635	22,663	43,606
Cyprus	61,821	77,958	16,137	26,103
Czechia	58,589	76,979	18,390	31,389
Denmark	87,086	82,485	-4,601	-5,284
Estonia	91,763	82,854	-8,908	-9,708
European Union	68,053	79,820	11,768	17,292
Finland	86,716	82,582	-4,135	-4,768
France	72,992	80,471	7,479	10,246
Germany	67,474	79,307	11,833	17,537
Greece	41,942	72,211	30,270	72,172
Hungary	49,159	74,983	25,824	52,532
Ireland	82,608	81,998	-0,611	-0,739
Italy	63,194	78,328	15,133	23,948
Latvia	79,630	81,875	2,245	2,819
Lithuania	78,049	81,522	3,473	4,450
Luxembourg	79,362	82,164	2,803	3,532
Malta	84,195	82,319	-1,876	-2,228

Netherlands	79,902	81,636	1,734	2,170
Poland	55,100	75,310	20,210	36,678
Portugal	68,948	80,495	11,547	16,747
Romania	21,488	69,418	47,930	223,049
Slovakia	53,723	76,166	22,443	41,776
Slovenia	67,994	79,066	11,072	16,284
Spain	80,676	81,747	1,071	1,327
Sweden	83,945	82,217	-1,728	-2,059
Media	68,503	79,141	10,638	15,529

6. Augmented Data Machine Learning and Prediction

Ranking of the algorithms created through augmented data							
Rank	Algoritmo	R^2	Mean absolute erro	Mean squared erro	Root mean squared error	Mean signed difference	Totale
1	*ANN*	★ 1	★ 1	★ 1	★ 1	★ 1	☆ 5
2	*PNN*	☆ 3	★ 2	★ 2	★ 2	☆ 4	☆ 13
3	*Linear Regression*	★ 2	☆ 5	☆ 3	☆ 3	☆ 6	☆ 19
3	*Simple Regression Tree*	☆ 5	☆ 3	☆ 4	☆ 4	☆ 3	☆ 19
4	*Tree Ensemble Regression*	☆ 6	☆ 4	☆ 5	☆ 5	☆ 5	☆ 25
5	*Polynomial Regression*	☆ 4	☆ 6	☆ 6	☆ 6	☆ 7	★ 29
6	*Gradient Boosted Trees*	☆ 7	☆ 7	☆ 7	☆ 7	★ 2	★ 30
7	*Random Forest Regression*	☆ 8	☆ 8	☆ 8	☆ 8	☆ 8	★ 40

	Prediction 2-Prediction 1			
Errors	*Prediction 1*	*Prediction 2*	*Var Ass*	*Var Per*
	ANN	**ANN**		
R^2	0,9847057062	0,98564299	0,0009372806	0,095183826
mean absolute error	0,0143025109	0,02580836	0,0115058474	80,44634599
mean squared error	0,0000000000	0,00100000	0,0010000000	
root mean squared error	0,0205465921	0,02967856	0,0091319637	44,44515032
mean signed difference	0,0074363346	0,00717097	-0,0002653641	-3,568480277
mean absolute percentage error	0,0269878962	0,05261262	0,0256247247	94,94895241
Mean	0,0138546668	0,0232541011	0,0093994344	67,84309224

 doi:10.20944/preprints202202.0359.v1

Country	Prediction 1	Prediction 2	Var Assoluta	Var Per
Belgium	79,36	81,84	2,48	3,13
Bulgaria	76,42	80,90	4,48	5,86
Estonia	82,85	82,46	-0,39	-0,47
European Union	79,82	81,89	2,07	2,60
Germany	79,31	81,80	2,50	3,15
Greece	72,21	79,69	7,48	10,36
Media	79,14	81,43	2,29	2,89