The Determinants of Internet User Skills in Europe

Angelo Leogrande (leogrande.cultore@lum.it)
LUM University Giuseppe Degennaro https://orcid.org/0000-0003-1381-4006

Nicola Magaletti
LUM Enterprise s.r.l.

Gabriele Cosoli
LUM Enterprise s.r.l.

Vito Giardinelli
LUM Enterprise s.r.l.

Alessandro Massaro
LUM University Giuseppe Degennaro https://orcid.org/0000-0003-1744-783X

Research Article

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

Posted Date: May 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1669068/v1

Angelo Leogrande[1], Nicola Magaletti[2], Gabriele Cosoli[3], Vito O. Giardinelli[4], Alessandro Massaro[5]

The Determinants of Internet User Skills in Europe

Abstract

The following article indicates the determinants of "*Internet User Skills*" among European countries based on the application of the database deriving from the DESI-Index. The data were analyzed using the following econometric models, namely: Panel Data with Fixed Effects, Panel Data with Random Effects, Pooled OLS, WLS, WLS corrected for heteroskedasticity. The Elbow method and the Silouette coefficient method were compared for the optimization of the number of clusters obtained by the k-Means algorithm. The result shows the presence of 5 clusters. A network analysis was carried out using the Euclidean distance with the result of identifying two network structures between some analyzed countries. subsequently a comparison was made between six different machine learning algorithms for the prediction of the future value of the variable of interest. The result shows that the best predictor algorithm is Gradient Boosted Tree Regression with an expected value of the predicted variable increasing by a value of 1.75%. Later a further comparison was made by comparing 6 algorithms with the increased data. The result shows that the best predictor is Simple Regression Tree. The interest variable is predicted to decrease by an amount equal to -6.099%. Statistical errors improve on average by 32.43% in the transition between the original data and the increased data.

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-Research Question

In the context of digitization, it is absolutely necessary to refer to the ability of the population, and in particular of citizens, to develop internet skills. Although internet skills can certainly be considered as basic compared to more advanced technological and IT skills such as programming and creating software, it is also true that they play a central role both for companies and for the public administration. In fact, it is not possible for companies that want to offer digital products and services, nor for governments interested in the dynamics of digital public administration, to operate on the supply side if there is a lack of consumers and citizens capable of using the internet. However, the internal skills do not only and exclusively concern human capital, and therefore the need to organize adequate training courses for the population at national level, but rather also concern the capital of

[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] Business Developer and Researcher at Lum Enterprise s.r.l. Email: giardinelli@lumenterprise.it Strada Statale 100 km 18, 70010 Casamassima BA, Puglia, Italy, European Union.
[4]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.
[5] 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

network and IT infrastructures. The price of access to the internet, and the same accessibility and coverage of the broadband at national level, are factors that significantly affect the ability to increase the level of digitalization and the ability to use internet.
Specifically, we wanted to address the issue of internet user skills in the European continent. This choice was made not only thanks to the abundance of data made available by the European Union through the DESI-Index, but also for the curiosity to verify how the socio-economic and cultural diversities present among European countries could have an impact in terms of internet user skills. Europe is in fact a geographical area of hundreds of millions of inhabitants characterized by a substantial dichotomy which generally takes the form of the contrast between Northern and Southern Europe, and which however, occasionally, can also be manifested as a gap between Eastern and Western Europe.
These socio-economic differences among European countries have enormous relevance in terms of digitization and contribute to determining a scenario in which a part of the population, the one with the least digital literacy, can be effectively subjected to exclusion from the world of work, from political participation and from the consumption of digital products and goods. In this context of general digital divide, there is that particular gap which is the gender digital divide which sees women particularly discriminated in the acquisition of internet skills.
Finally, it is necessary to consider that in the context of the tech-war that pits the US against China and which is all based on computer-digital skills and science, Europe risks being excluded. In fact, Europe in this context, with the exception of some areas of excellence, shows difficulties in contributing to the development of the knowledge economy, the information economy, and risks being absolutely marginal with respect to the Sino-American tech-war.
Last but not least, it is necessary to consider the distribution of the European population from a territorial point of view. In fact, unlike the US and especially China, Europe is made up of an city sparsely populated and vast rural areas, in which access to the internet and digital culture are scarcely present.
The article continues as follows: the second paragraph presents a literature review, the third paragraph shows the results of the econometric model, the fourth paragraph presents the results of clustering, the fifth paragraph illustrates the network analysis, the sixth paragraph refers the use of machine learning techniques for prediction, the seventh paragraph uses increased data for a further predictive analysis with machine learning, the eighth paragraph concludes. Finally, the appendix shows the details of the analytical results.

2. Literature Review

[1] analyze the case of the digital divide between men and women in Europe. The authors used two indicators of which one measures the use of the internet and the other the purchase of goods and services through the internet in the last 3 months. The results of the analysis show that the digital divide in terms of internet use is lower than the digital divide in terms of internet purchases. The nations that show a higher level of digital gender divide are Croatia and Italy. The most egalitarian nation is Cyprus in terms of the digital gender divide. The data also show a significant convergence between European nations in terms of the digital divide. Similarly, [2] show how the presence of women graduated in STEM disciplines reduces digital divide between male and female in Europe. [3] address the issue of the relationship between digital literacy and the economic conditions of the context favorable to the establishment of a business culture, in consideration of the role of human capital. According to the authors, digital literacy can be achieved through three elements, namely the

improvement of efficiency, the encouragement of innovation and the promotion of inclusiveness. There is therefore a positive relationship between digital literacy and the ability to conclude digital deals and in this relationship an important role is played by the training of human capital in the use of IT and internet tools. [4] consider the issue of digital culture and digitization in general in the Baltic countries, underlining how, especially Estonia, has managed to obtain significant advantages in terms of digitization by reaching the standards of European countries - and in some cases even exceeding the average level of EU28 countries.

The use of the internet and the acquisition of internet skills is also a requirement for optimizing the use of public data. In fact, the digitization of public services, the preparation of public documents online, can only be useful to the population in the presence of a digital culture and a literacy that allows citizens to have access to the opportunities of e-government [5]. The acquisition of IT skills is relevant to allow countries to optimize the benefits that digitization generates in terms of economic growth and social development. [6] shows that low-income European countries are more capable of translating internet skills and digitalisation into opportunities for economic growth and employment than high-income European countries. [7] consider the relationship between digitization and the impact on logistics systems at country level in Europe. The results of the analysis show that the countries that have greater capacity to develop the population's internet and IT skills also have greater opportunities to trigger a virtuous relationship between the efficiency of national logic systems and the degree of digitalization of the economy.

The use of information technologies, the acquisition of skills also related to the use of the internet, is absolutely necessary for participation in the European labor market. Indeed, workers with low technological and IT skills are also more likely to be excluded from the labor market [8]. The acquisition of digital skills is therefore an absolutely essential element in both the labor market and active citizenship. In this sense, the authors of [9] refer to a set of participation models between the public and the private sector which have the possibility of increasing the digital skills of the population. The ability to use the internet, and to increase digital skills, at the country level, also allows to increase the offer in terms of e-commerce. However, the authors [10] did not find a positive relationship between the development of e-commerce at the national level and the economic growth due to the development of digital. [11] highlight the role of digitization in connection with entrepreneurship. Digitization is positively associated with the increase of the Human Development Index at a national level and also positively contributed to the development of an environment generally oriented towards technological innovation favorable to entrepreneurship.

However, to effectively evaluate the use of the internet, it is necessary to refer to the provision of internet services within European countries. Specifically, Covid19 has significantly increased the use of the internet by increasing the size of the services and goods produced through the digital economy and consumed by citizens. It follows that there is a positive relationship between the development of internet skills and the development of market operators who are able to reduce the price of internet services at country level [12].

[13] analyzes the role of the use of the internet for the socio-economic development of European countries. Digital training is a useful tool for increasing the digital knowledge capital of Europeans. There are also positive relationships between the development of the digital economy, the impact on the GDP growth rate and the impact on unemployment. This positive report highlights the fact that the countries that are growing the most in Europe are precisely the countries that have the greatest capacity to use the internet and also to effectively apply artificial intelligence and data analytics. [14] address the issue of digitization as an economic factor and as a strategic factor from an educational point of view in the formation of human capital by focusing their attention on the analysis of the Spanish case. Analyzing the DESI data, the authors verify the presence of a set of social, economic

and educational gaps due to the digital divide. The elements that more than others reduce Spain's orientation towards technology consist in the lack of human capital equipment capable of having specialized skills in ITC and in the use of software. Furthermore, it is possible to identify the presence of a digital gender divide that can be effectively resolved through the increase of STEM education.
[15] take into consideration the case of technological innovation and digitization in the Baltic countries. The analysis shows the presence of a positive relationship between digitization and socio-economic performance. However, the potential that exists in the digitalisation economy is slow to translate into real increases in gross domestic product due to the lack of adequate IT skills among employees. [16] afford the question of the relationship between internet use and learning in educational institutions in Ghana. The use of the internet was facilitated by the presence of internet cafes and by the relationships that the students developed with the teachers. Furthermore, the use of the internet, and in particular of social networks, has improved communication between students. The authors believe that an improvement in internet services may also lead to an improvement in learning ability. [17] present an analysis of the relationship between the digital skills required by the market and the actual level of digital skills of a group of young people and social workers. The analysis shows that both young people and social workers are not actually able to improve their digital skills. [18] analyze the relationship between digitization and entrepreneurial capacity with attention to the female issue. The authors show that the growth of digital skills especially among the female population can increase entrepreneurial capital nationwide. [19] highlight the role that the acquisition of IT and internet skills have played in allowing workers access to smartworking during the Covid 19 pandemic. [20] show the presence of a positive relationship between the increase in digital skills including the ability to use the internet to fight corruption in European countries. The most digitized countries also have reduced levels of corruption. [21] consider the degree of digitization by comparing the case of Greece with the case of Europe using data from the DESI Index. The analysis shows that Greece's low performance in terms of digitization is related to the low use of the internet both by the population and by public institutions and administrations. [22] consideration the relationship between the Digital Divide and digital literacy in the production of Big Data for the optimization of Supply Chain Management. The analysis shows that the increase in digital literacy has a positive impact in terms of reducing the digital divide and increasing the efficiency of the Supply Chain optimized with the use of big data. The development of internet skills together with the ability to increase basic and applied computer knowledge within the DESI Index has positive effects in terms of reducing unemployment and increasing the number of employed [23]. [24] analyze the determinants of the ability of digital natives to use the internet. Contrary to the common opinion that believes that digital natives are naturally able to use the internet, authors find elements of discrimination in internet access that depend on the socio-economic condition of the families they belong to. The digital natives who have greater internet and computer skills also have parents who appear to have a high level of education, they are generally white or Asian. Therefore the authors conclude that the use of socio-economic variables is necessary for the prediction of the internet, computer and technological skills of digital natives. [25] analyze the role of internet use by men and women. The results show that men and women do not differ significantly in their online behavior except for self-assessment. In fact, the self-assessment of women is lower than that of men. This different perception can influence the behavior of women online and therefore have an impact in terms of social inequality in accessing new technologies. Finally the presence of internet user skills is also necessary to promote the diffusion of ICT Specialists [26] and for e-government [27]. Furthermore, there are external market forces, such as the price of internet subscriptions [28] and access to broadband [29] , which can have an impact in favoring or redirecting the acquisition of internet user skills by citizens. A relevant part of the article is devoted to the relationship between clusterization with the k-Means algorithm optimized with the

Silhouette coefficient and clustering with the Elbow method. In this regard, reference is made to the following titles for the literature concerning clustering [30] with particular application to images [31], [32], [33], [34].

3. The Econometric Model

We have estimated the following econometric model

$$\begin{aligned}\boldsymbol{InternetUserSkills_{it}} &= \boldsymbol{a_1} + \boldsymbol{b_1}(\boldsymbol{FixedBroadbandTakeUp})_{it} + \boldsymbol{b_2}(\boldsymbol{FixedBroadbandCoverage})_{it} \\ &+ \boldsymbol{b_3}(\boldsymbol{MobileBroadband})_{it} + \boldsymbol{b_4}(\boldsymbol{BroadbandPriceIndex})_{it} \\ &+ \boldsymbol{b_5}(\boldsymbol{eGovernment})_{it} + \boldsymbol{b_6}(\boldsymbol{AdvancedSkillsAndDevelopment})_{it} \\ &+ \boldsymbol{b_7}(\boldsymbol{DigitalIntensity})_{it} + \boldsymbol{b_8}(\boldsymbol{DigitalTechnologiesForBusiness})_{it} \\ &+ \boldsymbol{b_9}(\boldsymbol{eCommerce})_{it} + \boldsymbol{b_{10}}(\boldsymbol{IntegrationOfDigitalTechnologies})_{it} \\ &+ \boldsymbol{b_{11}}(\boldsymbol{DESIScoreIndex})_{it} + \boldsymbol{b_{12}}(\boldsymbol{ElectronicInformationSharing})_{it} \\ &+ \boldsymbol{b_{13}}(\boldsymbol{5GCoverage})_{it} + \boldsymbol{b_{14}}(\boldsymbol{AtLeastBasicDigitalSkills})_{it} \\ &+ \boldsymbol{b_{15}}(\boldsymbol{AboveBasicDigitalSkills})_{it} + \boldsymbol{b_{16}}(\boldsymbol{AtLeastBasicSoftwareSkills})_{it} \\ &+ \boldsymbol{b_{17}}(\boldsymbol{SMEsWithAtLeastABasicLevelOfDigitalIntensity})_{it}\end{aligned}$$

Where $\boldsymbol{i} = \boldsymbol{28}$[6] and $\boldsymbol{t} = [\boldsymbol{2016}; \boldsymbol{2021}]$

The variable denominated "*Internet User Skills*" is the sum of three different variables that are "*At Least Basic Digital Skills*", "*Above Basic Digital Skills*", "*At Least Basic Software Skills*". The variable can be better understood in explicit form:

$$\begin{aligned}\boldsymbol{InternetUserSkills_{it}} &= \boldsymbol{AtLeastBasicDigitalSkills_{it}} + \boldsymbol{AboveBasicDigitalSkills_{it}} \\ &+ \boldsymbol{AtLeastBasicoftwareSkills_{it}}\end{aligned}$$

Furthermore, the variable "*Internet User Skills*" added up with "*Advanced Skills and Development*" are part of the macro-variable "*Human Capital*" in the Desi Index Database.

$$\boldsymbol{InternetUserSkills_{it}} = \boldsymbol{HumanCapital_{it}} - \boldsymbol{AdvacendSkillsAndDevelopment_{it}}$$

We found that "*Internet User Skills*" is positively associated to the following variables:

- *5G Coverage:* is a variable that takes into consideration the "% of populated areas with coverage by 5G". There is a positive relationship between the percentage of the populated area covered by 5G and the value of the Intenet User Skills. Since at present 5G represents a very advanced technology, it follows that its diffusion tends to be widespread iitially in countries that have a greater presence of human capital skilled in the use of the internet. In fact, the geographical areas that are most sensitive to investment in 5G are precisely the areas that have a population more skilled in technical-IT terms.

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

- *Electronic Information Sharing*: is a variable that takes into consideration the number of companies that use ERP-Enterprise Resource Planning models to share useful information in carrying out the business management activity. Among the most widespread applications of ERP systems are accounting, production and marketing systems. There is a positive relationship between the presence of companies that have the ability to use ERP-type models and the presence of people who have skills in using the internet. This relationship derives from the fact that ERP is a software model used for management adopted by companies which tend to be innovative and which tend to be more widespread in the presence of qualified human capital.
- *Digital Intensity:* is a composite variable that measures the use of different digital technologies in companies. The technologies selected for the construction of the Digital Intensity Index are 12 and include for example the fact of having a website, the use of electronic invoices, the fact of employing ICT expert employees, e-commerce, the use of of big data, the use of robots and industrial services, the use of 3D printers. There is therefore a positive relationship between the value of the Digital Innovation Index and the presence of "Internet USer Skills". Evidently, the fact that companies are able to use a composite number of enabling information technologies also depends on the fact that there are digital skills that are partially summarized in the "Internet User Skills" indicator.
- *Desi Aggregate Scores:* it is a complex variable made up of a set of elements including Human Capital, Connectivity, Integration of Digital Technology and Digital Public Services. It is possible to represent the variable in extented form as follows: $\boldsymbol{DesiAggregateScore_{it} = HumanCapital_{it} + Connectivity_{it} + IntegrationOfDigitalTechnology_{it} + DigitalPublicServices_{it}}$. There is a positive relationship between the value of the Desi Index and the value of the "Internet User Skills" variable. This report highlights how important it is for countries that want to increase their level of digitization, the fact of investing in the training of the population so that they can acquire growing skills in the use of the internet. Digitization requires offering goods and services over the internet. If the population has a high ability to use the internet then it also has greater opportunities both to be an active consumer and also to offer content, services and thus become an online producer. In fact, many of the opportunities that are offered online require the involvement of the consumer who also becomes a producer by configuring the new case of the prosumer.
- *At Least Basic Digital Skills:* is a variable that considers the low-level digital skills of the population. In particular, digital skills are evaluated through the following categories, namely: information, communication, problem solving and software for creating content. There is a positive relationship between the dissemination of basic knowledge in IT and the dissemination of digital skills among internet users. Also in this case it is necessary to underline how important is the training of the population in the use of IT tools with the positive effects in economic, political and social terms.
- *Above basic digital skills:* is a variable that considers the percentage of individuals who are above basic computer knowledge through the analysis of four characteristics: information, communication, problem solving, and software for creating content. The percentage of people with above-average IT skills is positively associated with the presence of digital skills of internet users. Obviously, the presence of a growing percentage of the population having IT skills in itself also indicates the presence of digital skills of internet users.
- *At least basic software skills:* is a variable that takes into account the ability of people to use basic software features, such as word processing and the use of spreadsheets. The variable also considers the ability of individuals to use programming languages. There is a positive

relationship between the percentage of people with basic knowledge skills and the presence of internet skills among users. Obviously, individuals who have basic software skills are also equipped with internet skills.

We also find that the level of "Internet Use Skills" is negatively associated to:

- *Integration of Digital Technology:* is a composite variable consisting of the following subvariables, namely "SMEs with a basic level of digital intensity", "AI", "Cloud", "Big Data". This variable can be expressed explicitly as indicated below, that is: $\boldsymbol{IntegrationOfDigitalTechnology_{it} = SMEsWithABasicLevelOfDigitalIntensity_{it} + AI_{it} + Cloud_{it} + BigData_{it}}$. There is a negative relationship between the value of the Integration of Digital Technology variable and the value of the Internet User Skills variable. This negative relationship may seem paradoxical and contradictory but it can be better understood by considering the basis of the statistical analysis. In fact, while the Integration of Digital Technology variable is aimed at companies, the Internet User Skills variable refers to individuals. Hence the consideration that an increase in the Digital Integration of Digital Technology is possible even in the presence of a reduction in the level of internet user skills among the population. This relationship can be better understood considering that the fact that in a certain nation the percentage of companies able to use cutting-edge technologies increases does not automatically guarantee that there is a growing percentage of the population of those countries that is able to acquire skills. in the use of the internet. This is the case, for example, of low-digitization countries where hubs of companies that have evolved from a technological and digital point of view have been established.
- *Fixed Broadband Take Up:* is a composite variable that considers broadband access. The variable can also be expressed explicitly through the following formula: $\boldsymbol{FixedBroadbandTakeUp_{it} = OverallFixedBroadbandTakeUP_{it} + AtLeast100MbpsFixedBroadbandTakeUp_{it} + AtLeast1GBPSATake_{it}}$. the analysis shows the presence of a negative relationship between the "Fixed Broadband Take Up" variable and the presence of internet user skills. This negative relationship indicates that the growth in broadband connections can also be associated with a reduction in users' internet skills. This relationship may appear to be counterfactual. However, it must be considered that a large part of the population uses internet services for entertainment and communication purposes and therefore does not necessarily develop those characteristics of internet skills that would allow, for example, to activate positive effects from an economic, social and political point of view.
- *Fixed Broadband Coverage:* is a variable that considers the spread of the internet fixed band. The value of this variable can also be expressed through the explicit form, that is: $\boldsymbol{FixedBroadbandCoverage_{it} = FastBroadbandNGACoverage_{it} + FixedVeryHighCapacityNetworkVHCNCoverage}$. The econometric analysis shows the presence of a negative relationship between the value of broadband coverage and the value of the users' internet skills. This relationship stems from the fact that broadband coverage does not necessarily imply an increase in digital capacities in the population. In fact, very often the population buys broadband not so much to develop digital or advanced technological skills, to start-ups and newcos, but rather to access forms of digital consumption.
- *Mobile Broadband:* It is a variable made up of various sub-variables including "4G Coverage", "5G Readiness", "5G coverage", "Mobile Broadband Take Up". This indicator can also be expressed explicitly in the following formula, that is: $\boldsymbol{MobileBroadband_{it} =}$

$\boldsymbol{4GCoverage_{it} + 5GReadiness_{it} + 5GCoverage_{it} + MobileBroadbandtakeUp_{it}}$. The result shows the presence of a negative relationship between the value of "Mobile Broadband" and the value of Internet User Skills. This relationship appears to be counterfactual. However, it is clear that the diffusion of the network in its various components alone is not sufficient to guarantee that the population acquires digital knowledge that perhaps becomes real hard skills. As demonstrated in the previous points, it is only training, or the acquisition of real skills that involves the growth of internet skills by the population. The simple distribution of the network has no capacity to generate in itself a growth in the users' internet skills.

- *Broadband price index:* is a variable that considers the price of internet subscriptions on the basis of a set of fixed, mobile and convergent broadband internet offers. The indicator varies between 0 and 100. Also in this case there is a negative relationship between the value of the price of broadband and the value of the digital skills of internet users. This condition means that as the price index increases, the value of users' internet skills decreases. This condition may be due to the fact that the increase in the price of broadband shows the presence of a monopolistic internet services market which therefore turns out to be very distant from a competitive internet services market that characterizes the matured markets also consisting of presence of a growing percentage of the population with internet skills.
- *E-Government:* is a variable that considers e-government as constituted by a set of sub-elements such as: "eGovernment Users", "Pre-Filled Forms", "Digital Public Services For Citizens", "Digital Public Services for Business" , "Open Data". It is also possible to represent this variable in extended form as indicated below: $\boldsymbol{eGovernment_{it} = eGovernmentUsers_{it} + PrefilledForms_{it} + DigitalPublicServicesForCitizens_{it} + DigitalPublicServicesForCitizens_{it} + DigitalPublicServicesForBusinesses_{it} + OpenData_{it}}$ There is a negative relationship between the value of e-government and the value of internet user skills at country level. This negative relationship can be better understood considering that the presence of e-government services does not in itself guarantee the fact that the population of those same countries has the digital characteristics such as to be able to use the Internet effectively.
- *Advanced Skills and Development:* is a variable consisting of the following elements: "ICT Specialists", "Female ICT Specialists", "Enterprise Providing ICT Training", "ICT graduates". It is a variable that can also be represented in extend form as follows: $\boldsymbol{AdvancedSkillsAndDevelopment_{it} = ICTSpecialists_{it} + FemaleICTSpecialists_{it} + EnterprisesProvidingICTTraining_{it} + ICTGraduates_{it}}$. There is a negative relationship between the value of this variable and the value of "Internet User Skills". However, this negative relationship must be understood above all as an accounting element. In fact, within the DESI Index there is a sort of zero-sum game between the value of Internet User Skills and the value of Advanced Skills and Development. In fact, the sum of Interner User Skill and Advanced Skills and Development constitutes the total value of Human Resources. It therefore follows that in the internal dynamics of the determination of Human Resources there is a negative relationship between Internet User Skills and Advanced Skills and Development as they both add to unity.
- *Digital Technologies for Businesses:* is a variable made up of the following sub-variables, namely: "Electronic information sharing", "Social Media", "Big Data", "Cloud", "AI", "ICT for Environmental Sustainability", "e-Invoice". It is possible to represet the same variable in an extended form as follwors: $\mathbf{DigitalTechnologiesForBusinesses_{it} = ElectronicInformationSharing_{it} + SocialMedia_{it} + BigData_{it} + Cloud_{it} + AI_{it} +}$

$\mathbf{ICTForEnvironmentalStability_{it} + eInvoices_{it}}$. This variable is negatively associated with the Internet User Skills value. The negative relationship between Internet User Skills and Digital Technologies for Business underlines the fact that the digital orientation of companies does not necessarily imply a capacity of the population to be able to develop significant IT and digital skills. However, it is clear that the degree of innovativeness and digitization of the population is absolutely necessary to ensure that companies are able to monetize their digital investments.

- *E-Commerce:* e-commerce is a composite variable consisting of a set of sub-varibials or "SMEs Selling Online", "eCommerce Turnover", "Selling Online Crossboder". This variable can also be expressed explicitly as indicated below: $\boldsymbol{eCommerce_{it} = SMEsSellingOnline_{it} + eCommerceTurnover_{it} + SellingOnlineCrossBorder_{it}}$. There is a negative relationship between eCommerce and the value of Internet User Skills. This report indicates that increasing digital commerce skills does not automatically lead to an increase in the digital skills of the population.
- *SMEs with at least a basic level of digital intensity:* is an indicator that detects the minimum number of basic technologies that are used by companies in the context of digitization. The DEsi Index takes into consideration 12 enabling technologies of which 4 are considered as basic. This variable is negatively associated with the value of the Internet User Skills. Also in this case the negative relationship between these variables may be due to the fact that while on the one hand the variable in question takes small and medium-sized enterprises as a reference, on the other hand the Internet User Skills variable considers people. It follows that the fact that companies improve in the use of information technologies does not necessarily imply an increase in the internet skills of individuals.

Statistical Results of the Econometric Analysis

Variables		Panel Data with Fixed Effects		Panel Data with Random Effects		WLS		WLS corrected for heteroskedasticity		Pooled OLS		Mean
		Coefficient	P-Value	Coefficient	P-Value	Coefficient	P-Value	Coefficient	P-Value	Coefficient	P-Value	
Internet User Skills	A10											
Constant		-6,17E-05		3,98E-05		4,91E-05	*	2,50E-05		4,78E-05		2,00E-05
Fixed broadband take-up	A1	-0,0478424	**	-0,0482625	**	-0,0350607	*	-0,0496907	***	-0,0517111	**	-4,65E-02
Fixed broadband coverage	A2	-0,0478432	**	-0,0482629	**	-0,0350607	*	-0,0496932	***	-0,0517122	**	-4,65E-02
Mobile broadband	A3	-0,0478419	**	-0,0482632	**	-0,0350624	*	-0,0496945	***	-0,0517133	**	-4,65E-02
Broadband price index	A4	-0,0478474	**	-0,0482623	**	-0,0350609	*	-0,0496902	***	-0,0517115	**	-4,65E-02
e-Government	A5	-0,0478363	**	-0,0482622	**	-0,0350607	*	-0,0496924	***	-0,0517118	**	-4,65E-02
Advanced Skills and Development	A11	-0,0478438	**	-0,0482612	**	-0,0350594	*	-0,0496911	***	-0,0517104	**	-4,65E-02
Digital intensity	A12	8,83E-05	***	6,92E-05	**	6,70E-05	**	5,05E-05	***	6,22E-05	**	6,75E-05
Digital technologies for businesses	A13	-2,18E-05	***	-1,78E-05	**	-1,74E-05	**	-1,10E-05	*	-1,59E-05	**	-1,68E-05
e-Commerce	A14	-1,80E-05	*	-1,71E-05	**	-1,79E-05	**	-1,01E-05	*	-1,58E-05	**	-1,58E-05
Integration of Digital Technology	A17	-0,191285	**	-0,192985	**	-0,140178	*	-0,198733	***	-0,20679	**	-1,86E-01
Aggregate score	A19	0,191362	**	0,193049	**	0,140242	*	0,19877	***	0,206847	**	1,86E-01
Electronic Information Sharing	A20	1,07E-05	*	6,67E-06	**	5,54E-06	**	5,80E-06	**	5,73E-06	**	6,88E-06
5G coverage	A42	2,12E-06	*	1,75E-06	*	2,11E-06	***	2,26E-06	***	2,08E-06	**	2,06E-06
At least Basic Digital Skills	A44	0,476092	***	0,475876	***	0,482482	***	0,47516	***	0,474151	***	4,77E-01
Above basic digital skills	A45	0,476082	***	0,475873	***	0,482472	***	0,475158	***	0,474149	***	4,77E-01
At least basic software skills	A46	0,476058	***	0,475847	***	0,48244	***	0,475134	***	0,474123	***	4,77E-01
SMEs with at least a basic level of digital intensity	A47	-1,83E-05	***	-1,22E-05	**	-1,15E-05	**	-8,42E-06	**	-1,08E-05	**	-1,22E-05

Figure 1. Statistical Results of the Econometric Analysis.

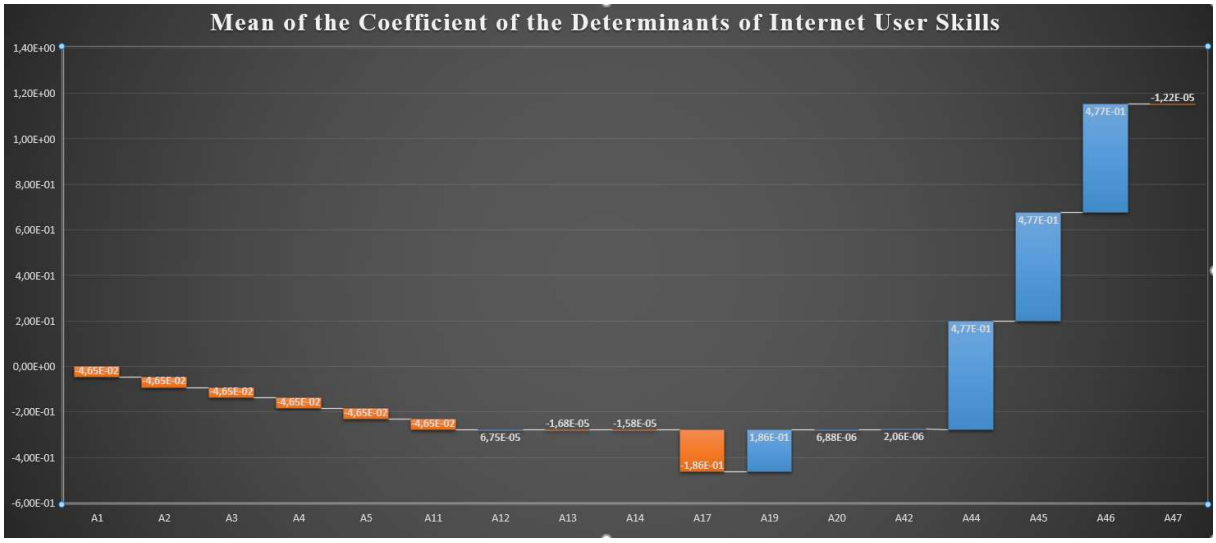

Figure 2. Mean of the Coefficient of the Determinants of Internet User Skills.

In summary, an average of the values of the coefficients of the various econometric models that were analyzed was achieved. Based on the following formula:

$$\boldsymbol{Mean_{coefficient}} = \boldsymbol{OLS_i} + \boldsymbol{RandomEffect_i} + \boldsymbol{FixedEffects_i} + \boldsymbol{WLS_i} + \boldsymbol{WLSWithHeteroskedasticity_i}$$

Where $\boldsymbol{i = 17}$

However, as is evident from the analysis carried out, the effects that have an impact more than others in terms of Internet User Skills are indicated below: "*At Least Basic Digital Skills*", "*Above Basic Digital Skills*", "*At Least Basic Software Skills*". These variables are positively correlated to "*Internet User Skills"* and highlight the fact that the elements that have an impact more than others in terms of digital skills of internet users consist in training and in the acquisition of digital knowledge by Internet users. The variable that negatively affects the value of "*Internet User Skills*" is "*Integration of Digital Technology*". Integration of Digital Technology concerns the ability of companies to make investments in digital areas such as AI, Big Data, Cloud, Social Media. The negative relationship between Integration of Digital Technology and Internet User Skills may seem paradoxical, but it is also necessary to consider how the indicators are calculated. In fact, while on the one hand Integration of Digital Technology concerns companies and their ability to innovate, on the other hand Internet User Skills concerns the technological skills of individuals. As a result, the fact that companies invest significantly in digitization says nothing about the ability of individuals to perfect their technological and digital skills.

4. Clusterization with k-Means Algorithm Optimized with Silhouette Coefficient and Elbow Method

The k-Means algorithm was then applied to analyze the presence of groupings by value of the observed variable or "*Internet User Skills*" among the various European countries between 2016 and 2021. To choose an optimal number of clusters, we used initially the Silhouette coefficient model.

The Silhouette coefficient varies between -1 and 1. The optimal number of clusters is chosen in relation to the maximum value of the Silhouette coefficient. Based on the application of this criterion, the optimal number of clusters is 5. The clusters are analyzed as follows:

- C1: *Cyprus, Poland, Italy, Greece, Latvia, Ireland, Hungary;*
- C2: *Austria, Luxembourg, Estonia, Germany, Belgium;*
- C3: *Romania, Bulgaria;*
- C4: *Finland, Netherlands, Denmark, Sweden;*
- C5: *Lithuania, Czechia, Croatia, Spain, France, Slovenia, Slovakia, Malta, Portugal.*

Clusters can be evaluated and placed on a hierarchical scale based on the median value of "Internet User Skills". In first place for median value of "*Internet User Skills*" is the cluster 4-C4 with a value pair of 37.12; in second place is the 2-C2 cluster with a value equal to 31.3987; in third place is the 5-C5 cluster with a value of 27.40; in fourth place there is cluster 1-C1 with a value equal to 21.8220 and in fifth place there is cluster 3-C3 with a value equal to 13.71. Therefore, the following ordering derives from it, that is: C4> C2> C5> C1> C3.

However, the Elbow method was used to obtain a better confirmation of the correctness of the clustering procedure. The Elbow method allows to identify the optimal number of clusters through a graphical method. In a system of Cartesian axes the number of clusters k is reported on the x axis while the value of the SSq is placed on the y axis. Using the Elbow method it is therefore possible to verify a number of clusters equal to 4. According to this method, the value of the cluster is lower than that indicated with the Silhouette method. Since the Silhouette method has indicated an optimal number of clusters equal to 5 and the Elbow method has indicated an optimal number of clusters equal to 4, then the clustering is repeated with the 4-cluster Silhouette method to verify whether clustering with k = 5 to a clustering with k = 4. We must also consider the fact that the value of the Silhouette coefficient for cluster 5 is equal to 0.558 while the value of the Silhouette coefficient for cluster 4 is equal to 0.555 units. In other words, in the passage between a clustering with k = 4 and a clustering with k = 5 there is a reduction of 0.003 units equal to 0.54%. Thus realizing the clustering with the k-Means algorithm with a value of k = 4. The clusters obtained are composed as follows:

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

It is possible to create a translation matrix between the cluster with k = 4 and the cluster with k = 5, indicating the countries that in the transition to the cluster with k = 4 and the cluster with k = 5 are differently attributed. To analyze the comparison between the various clustering models, the following method is proposed, that is: CN_n with C which summarizes the cluster value, N which indicates the Cluster number and n which represents the value of k in the clustering. Therefore, the following correlation between the various clusters derives, namely:

- $C3_5 = C1_4$
- $C4_5 = C4_4 + Germany\ and\ Austria$
- $C1_5 = C3_4$
- $C5_5 = C2_4 + Belgium, Estonia\ and\ Luxembourg$

Therefore, from a strictly quantitative point of view, the transition from an optimized clustering model with the Silhouette coefficient with k = 5 and an optimized clustering model with the Elbow method with k = 4 involves an absolutely marginal reduction of the Silhouette coefficient 0.54 %. However, this marginal variation between the Elbow method and the Silhouette method should not be understood as a universal law but rather as an application case due to the presence of particular data or as a result of a data-driven analytical procedure. However, there are cases in which the choice between the Elbow method and the Silhouette method generates results that are particularly divergent as indicated in [35]. Finally, from a strictly geographical point of view, it is possible to note that the countries that have most evolved from the point of view of Internet User Skills are the Scandinavian countries followed by the countries of central Europe such as Germany, Austria, Luxembourg, Estonia, Belgium. Hence the presence of a significant divergence within Europe between the countries that have a very high value of Internet User Skills, essentially coinciding with the countries of Central Europe and Scandinavia, and the countries that are more backward among the what are the nations of southern and eastern Europe.

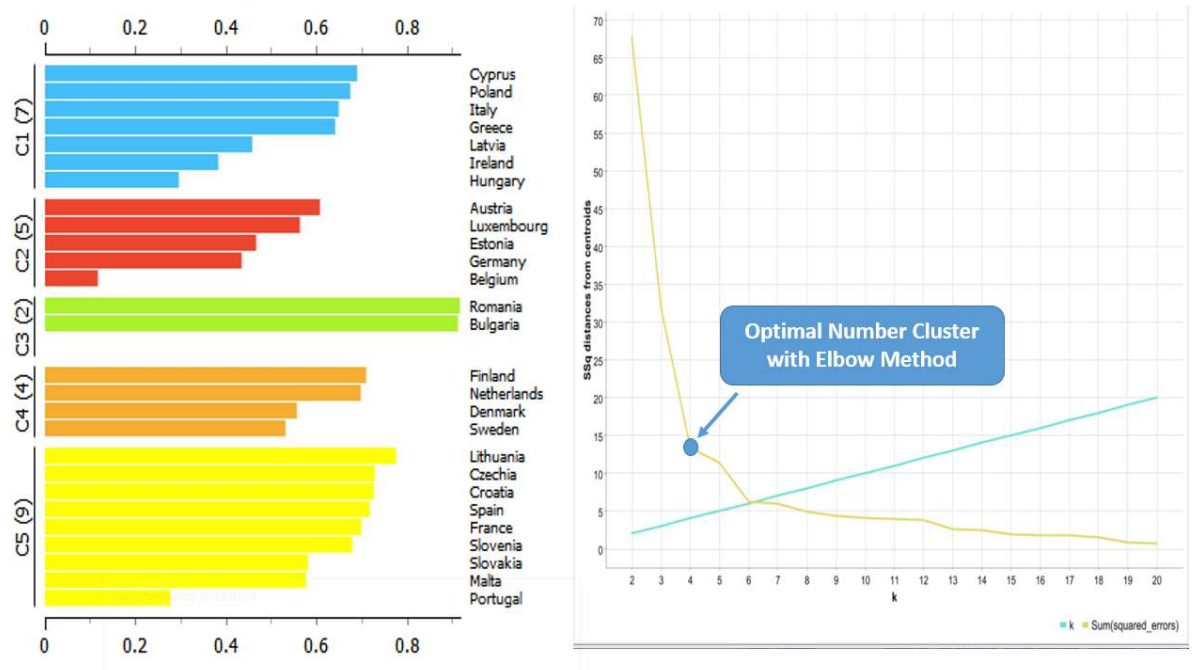

Figure 3. Comparison between k-Means Clustering with the Silhouette Coefficiente and the k-Means Clustering with the Elbow Method.

5. Network Analysis

A network analysis was carried out below to verify the presence of particular links between the countries considered. Among the results that have been identified, the network structures between the various countries considered are proposed first and then the specific links between the various countries within the networks. Specifically, two network structures have been identified, namely a

"Main Network" consisting of a set of multiple links and a "Minor Network" consisting of two countries.

Main Network: the main network consists of a set of 7 nodes between the following countries, namely: Malta, Czechia, France, Spain, Lithuania, Croatia, Slovenia. Specifically, there are the following linkages between countries, namely:

- Malta has a bond with Czechia of an amount equal to 0.23;
- Czechia has a bond with Malta equal to a value of 0.23, with France equal to an amount of 0.27, with Lithuania with an amount equal to 0.19; with Spain with an amount equal to 0.22;
- France has a link with Czechia with an amount of 0.27, with Lithuania with a value of 0.25, and with Spain with a value of 0.19, and with Croatia with an amount equal to 0.22;
- Lithuania has a link with the Czech Republic equal to an amount of 0.19 units, with France equal to 0.25, with Spain for an amount equal to 0.21, with Croatia equal to an amount of 0.24 units, and with Slovenia with an amount equal to 0.16 units;
- Spain has a link with the Czech Republic equal to 0.22 units, with France for a value equal to 0.19, with Lithuania with a value equal to 0.21, with Croatia with an amount equal to 0.25 units;
- Croatia has a link with Spain for an amount of 0.25 units, with Lithuania with an amount equal to 0.24 units, with France for an amount of 0.22 units.
- Slovenia has a link with Lithuania equal to an amount of 0.16 units, with Croatia for an amount equal to 0.26 units.

Minor Network: consists of a basic structure with only two countries, namely Austria and Germany with a link value of 0.22 units.

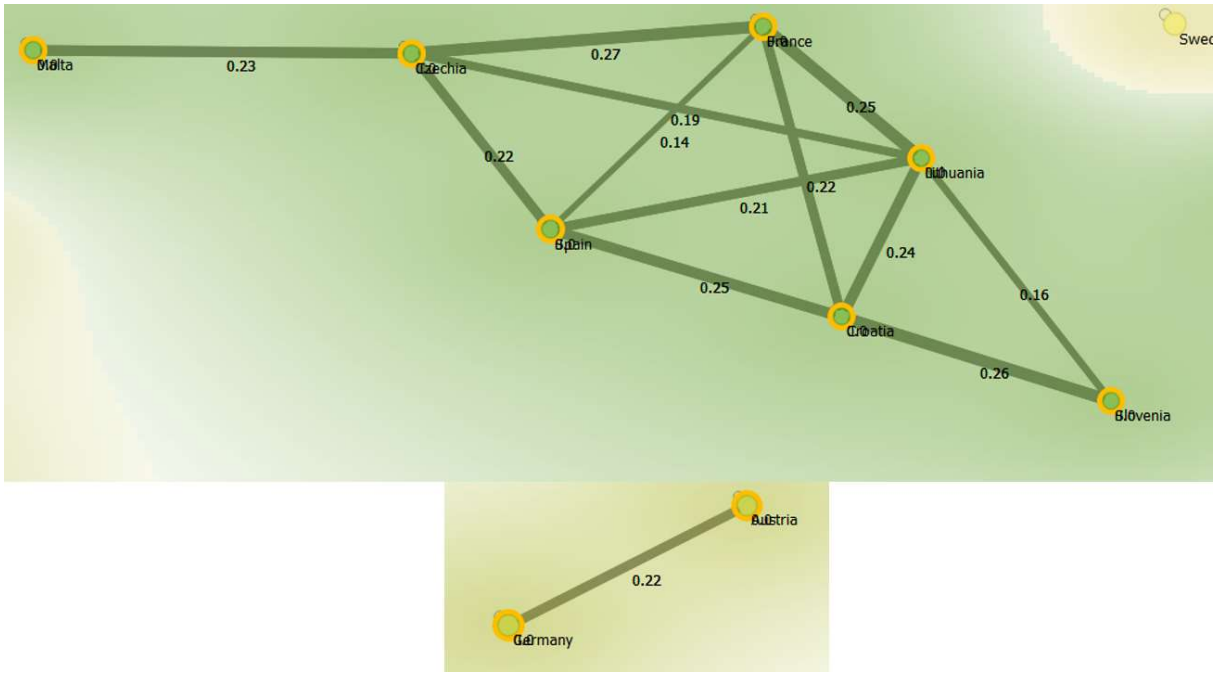

Figure 5. Network Analysis.

6. Machine Learning and Predictions with Original Data

An analysis is carried out below through the use of eight different machine learning algorithms used to predict the value of Internet User Skills in European countries. The algorithms were trained using 70% of the available data while the remaining 30% was used for the actual prediction. The comparison between the results produced by the various algorithms is carried out through a set of indicators that is R-squared, "*Mean Absolute Error*", "*Mean Squared Error*", "*Root Mean Squared Error*", "*Mean Absolute Percentage Error*". On the basis of these indicators, the following ranking of the algorithms by predictive capacity was obtained, that is:

- *Gradient Boosted Tree Regression* with a payoff value of 7;
- *ANN-Artificial Neural Network* with a payoff value of 13;
- *PNN-Probabilistic Neural Network* with a payoff value of 17;
- *Tree Ensemble Regression* with a payoff value of 25;
- *Simple Regression Tree* with a payoff value of 28;
- *Random Forest Regression* with a payoff value of 36.

Therefore, by applying the best performing algorithm or the Gradient Boosted Tree Regression it is possible to obtain the following predictive results, namely:

- *Austria* with a decrease from an amount of 32.38 to a value of 31.40 or equal to a variation of -0.98 units equal to a value of -3.03%;
- *Belgium* with an increase from an amount of 27.69 up to a value of 28.46 units or equal to a change of 0.77 units equal to an amount of 2.80%;
- *Bulgaria* with an increase from an amount of 13.35 units up to a value of 14.09 or equal to a change of 0.73 units equal to a value of 5.50%;
- *Croatia* with an increase from an amount of 27.04 units up to a value of 27.41 units or equal to an amount of 0.37 units equal to a value of 1.38%;
- *Cyprus* with a decrease from an amount of 21.82 units up to a value equal to 20.22 units or equal to an amount of -1.60 units equal to a value of -7.35%;
- *Czechia* with a decrease from an amount of 28.43 units up to a value of 28.39 units equal to an amount of -0.04 units equal to a value of -0.15%;
- *Denmark* with an increase from an amount of 35.48 up to a value of 38.23 units or equal to a change of 2.75 units equal to an amount of 7.74%;
- *Estonia* with an increase from an amount of 30.22 units up to a value of 31.70 units or equal to a variation of 1.48 units equal to a change of 4.91%;
- *Finland* with an increase from an amount of 38.23 units up to a value of 39.25 units up to a value of 1.02 units equal to a growth of 2.68%;
- *France* with an increase from an amount of 27.69 units up to a value of 28.04 units or equal to a change of 0.35 units equal to an amount of 1.26%;
- *Germany* with an increase from an amount of 33.87 units up to a value of 36.01 units or equal to a change of 2.15 units equal to an amount of 6.35%;
- *Greece* with a decrease from an amount of 24.04 units up to a value of 21.82 units or equal to a variation of -2.21 units equal to an amount of -9.21%;
- *Hungary* with a variation from an amount of 23.30 units up to a value of 25.90 units or equal to a variation of 2.59 units equal to 11.13%;
- *Ireland* with a decrease from an amount of 26.67 units up to a value of 26.47 with a decrease equal to -0.20 units equal to a variation of -0.74%;

- *Italy* with an increase from an amount of 20.22 units up to a value of 21.82 units or equal to a variation of 1.60 units equal to an amount of 7.93%;
- *Latvia* with an increase from an amount of 20.86 up to a value of 25.90 or a variation of 5.04 units equal to a variation of 24.15%;
- *Lithuania* with a variation from an amount of 27.41 units up to a value of 26.89 units or equal to a variation of -0.51 units equal to a variation of -1.88%;
- *Luxembourg* with a decrease from an amount of 31.39 units up to a value of 29.46 units or equal to an amount of -1.93 units or equal to an amount of -6.14%;
- *Malta* with a variation from an amount of 28.84 units up to a value of 28.43 units or with a variation of -0.41 units or equal to an amount of -1.41%;
- *Netherlands* with a variation from an amount of 39.24 units up to a value of 38.23 units or equal to a variation of -1.01 units equal to an amount of -2.56%;
- *Poland* with an increase from an amount of 20.91 units up to a value of 21.77 units or equal to a change of 0.87 units or equal to a change of 4.15%;
- *Portugal* with an increase from an amount of 25.90 units up to an amount of 21.77 units or equal to a variation of 0.87 units or equal to an amount of 4.15%;
- *Romania* with a decrease from an amount of 14.08 units up to a value of 13.35 units or equal to a variation of -0.73% or equal to an amount of -5.20:
- *Slovakia* with an increase from an amount of 25.54 units up to a value of 26.51 units or equal to a variation of 0.97 units or equal to an amount of 3.78%;
- *Slovenia* with a decrease from an amount of 28.55 units up to a value of 28.43 units or equal to a variation of -0.12 units equal to an amount of -0.41 units;
- *Spain* with a decrease from an amount of 28.55 units up to a value of 28.37 units or equal to an amount of -0.18 units equal to an amount of -0.63%;
- *Sweden* with an increase from a value equal to 36.01 units up to a value of 38.50 units or equal to a variation of 2.49 units, equal to an amount of 6.91%;

Overall, the average value of Internet User Skills in European countries is predicted to grow from an amount of 27.32 units up to a value of 27.80 units or equal to an amount of 0.48 units equal to an amount of 1.75%.

7. Machine Learning and Predictions with Increased Data

Later the data obtained from the prediction were added to the original dataset generating a new dataset with increased data. The new dataset was thus analyzed through the use of six different machine learning algorithms used for data prediction. In particular, the algorithms were analyzed on the basis of their ability to maximize the R-square and minimize a set of statistical errors including "*Mean Absolute Error*", "*Mean Squared Error*", "*Root Mean Squared Error*". 70% of the data was used for training the algorithms while the remaining 30% was used for the actual prediction. The algorithms thus evaluated on the basis of performance constitute the following order, namely:

- *Simple Regression Tree* with a payoff value of 4;
- *PNN-Probabilistic Neural Network* with a payoff value of 8;
- *Tree Ensemble Regression* with a payoff value of 12;
- *Random Forest Regression* with a payoff value of 16;

- *ANN-Artificial Neural Network* with a payoff value of 20;
- *Gradient Boosted Tree Regression* with a payoff value of 24.

Therefore, through the use of the best performer algorithm or the Simple Regression Tree it was possible to make the following predictions, namely:

- *Austria* with an increase forecast from an amount of 0.697 units up to a value of 0.708 units or equal to an amount of 0.011 units equal to 1.578%;
- *Croatia* with an increase from an amount of 0.543 to an amount of 0.582 units or equal to a change of 0.039 units equal to an amount of 7.182%;
- *Denmark* with a value of 0.961 also confirmed in the prediction;
- *Finland* with a diminutive value of the prediction from a value equal to 1 up to a value of 0.961 or equal to a value of -0.039 units equal to an amount of -3.90%;
- *Germany* with a value of the change from an amount of 0.875 units up to a value of 0.708 units or a change equal to an amount of -0.167 units equal to an amount of -19.086%;
- *Hungary* with a value of 0.484 also confirmed in the prediction with the increased data;
- *Ireland* with a decrease from an amount of 0.507 units up to a value of 0.327 units or equal to a change of -0.180 units equal to an amount of -35.503%;
- *Luxembourg* with an increase of 0.028 units in absolute value.

On average, the value of "*Internet User Skills*" is predicted to decrease from an amount equal to 0.632 to a value of 0.594 units or equal to a value of -0.039 units equal to a value of -6.09%. Finally, in making a comparison between the value of the prediction made with original data and the value of the prediction made with the increased values, a significant improvement is evident both in terms of R-squared and also in terms of minimization of statistical errors or "*Mean Absolute Error*", "*Mean Squared Error*", "*Root Mean Squared Error*". The data were normalized or included in the interval [0.1]. Overall, in the transition between the prediction with the original data made with the Gradient Boosted Tree Regression algorithm and the Simple Regression Tree algorithm, the following improvements were found in terms of predictive capacity, namely:

- *R-squared:* growth from an amount of 0.96 up to a value of 0.98 or equal to an amount of 0.022 units equal to an amount of 2.3%;
- *Mean Absolute Error*: decrease from a value of 0.037 units up to a value of 0.0260 units or equal to a variation of -0.011 or equal to a value of -30.81%;
- *Mean Squared Error*: decrease from an amount of 0.0021 to a value of 0.001 or equal to an amount of -0.0011 units equal to a value of -54.045;
- *Root Mean Squared Error*: with a variation from an amount of 0.04664 up to a value of 0.0313 or equal to a variation of -0.0152 equal to an amount of -32.72.

Finally, by averaging the variation of the three most relevant statistical errors, namely "*Mean Absolute Error*", "*Mean Squared Error*", "*Root Mean Squared Error*", it is possible to verify a reduction in errors equal to -32.42% in the transition from prediction with the original data to the prediction with the increased data.

8. Conclusions

In the following article, the socio-economic determinants of the ability of individuals to acquire skills in terms of internet use in Europe have been investigated. The analysis of the literature highlights the positive relationship between the growth of the population's digitalization skills and economic

performance at the country level. A particularly important role is dedicated to the analysis of gender difference in the context of the digital divide. The scientific debate highlights how women suffer particular discrimination in terms of access to the internet and use of the internet. The growth in the number of women graduates in STEM disciplines is seen as a solution to the issue of the gender digital divide. The econometric analysis conducted through the panel data technique with 5 different models shows that the elements that have the greatest positive impact in terms of "*Internet User Skills*" are the following variables, namely "*At Least Basic Digital Skills*", "*Above Basic Digital Skills*","*At Least Basic Software Skills*". On the other hand, the variable that most of all negatively affects "*Internet User Skills*" is "*Integration of Digital Technology*" that measure the ability of a firm to implement in its business model some digital technologies such as AI, Big Data, Cloud, and Social Media. A clustering was then carried out using the optimized k-Means algorithm through a comparison between the Silhouette coefficient and the Elbow method. The Silhouette coefficient indicated an optimal number of clusters equal to 5, while the Elbow method indicated an optimal number of clusters equal to 4. Clustering with k = 5 was preferred to clustering with k = 4 for the ability to better represent the differentiated performances of the various European areas by value of "*Internet User Skills".* The network analysis applied with the Manhattan distance reveals the presence of two network structures, one of which presents a complex structure of linkages and the other is simplified with a single link between two countries.
Finally, eight different algorithms were used to predict the value of "*Internet User Skills*". The best performing algorithm is Gradient Boosted Tree Regression. The result shows an increase in the value of "*Internet User Skills*" on average for the countries considered equal to an amount of 1.75%. The prediction data with the Gradient Boosted Tree Regression was added to the data from the original dataset for further comparison and prediction. The best predictor algorithm with increased data is Simple Regression Tree with an improved predictive capacity, compared to prediction with the original data, on average of 32.42%.

9. Bibliography

[1] M. López-Martínez, O. García-Luque e M. Rodríguez-Pasquín, «Digital Gender Divide and Convergence in the European Union Countries,» *Economics,* vol. 15, n. 1, pp. 115-128, 2021.

[2] F. Damiani e P. R. Mondroño, «Measuring Women's Digital Inclusion. A poset-based approach to the Women in Digital Scoreboard,» *University of Modena and Reggio Emilia, Department of Economics" Marco Biagi",* n. 0210, 2022.

[3] Z. Wang, X. Li, J. Li e C. Yuan, «Theoretical Research on the Mechanism of Improving Digital Literacy for Optimizing Doing-Digital-Business Environment,» *Proceedings of the 4th International Conference on Economic Management and Green Development ,* vol. Springer, n. Singapore, pp. 487-498, 2021.

[4] J. Masso, K. Espenberg e I. Mierina, «Social dialogue and the new world of work: The case of the Baltic state,» in *The New World of Work*, Edward Elgar Publishing., 2021.

[5] R. Zdjelar, N. Ž. Hrustek e N. Vrček, «Usage and Role of Open Government Data and Public Policies of 54+ Citizens e-Inclusion Issues,» *Central European Conference on Information and Intelligent Systems,* pp. 121-129, 2020.

[6] E. C. Yalçın, «Efficiency Measurement of Digitalization on EU Countries: A Study Based on Data Envelopment Analysis,» *International Journal of Management, Knowledge and Learning,* vol. 10, n. 1, pp. 323-333, 2021.

[7] A. Moldabekova, R. Philipp, H. E. Reimers e B. Alikozhayev, «Digital technologies for improving logistics performance of countries,» *Transport and Telecommunication,* vol. 22, n. 2, pp. 207-216, 2021.

[8] I. Czaja e M. Urbaniec, «Digital exclusion in the labour market in European countries: causes and consequences,» *European Journal of Sustainable Development,* vol. 8, n. 5, pp. 324-324, 2019.

[9] G. Lang e T. Triantoro, «Preparing for the Future of Work: Towards a Typology of Digital Skills Initiatives,» *Proceedings of the EDSIG Conference,* n. 2473, p. 4901, 2021.

[10] B. Jaković, T. Ćurlin e I. Miloloža, «Enterprise digital divide: Website e-commerce functionalities among European Union enterprises,» *Business Systems Research: International journal of the Society for Advancing Innovation and Research in Econo,* vol. 12, n. 1, pp. 197-215, 2021.

[11] B. I. Andreea e L. A. Elena, «THE RELATIONSHIP BETWEEN ENTREPRENEURSHIP AND DIGITALIZATION-SPOTLIGHT ON THE EU COUNTRIES,» *Studies in Business & Economics,* vol. 15, n. 3, 2020.

[12] V. M. Dumitrache, M. L. Popescu, M. D. Oancea-Negescu e A. Diaconu, « Use of Internet-Based Services in EU in 2019 and 2020».

[13] J. Cifuentes-Faura, «Digital Agenda, New Technologies and Education for the Integration of Europe: an Economic Study,» *European Integration Studies,* vol. 15, pp. 55-62, 2021.

[14] I. M. G. Vidal e A. G. Barujel, «Socioeducational gaps derived from the impact of digitization in Spain 2020,» *Red-Revista De Educacion a Distancia,* pp. 24-24, 2021.

[15] D. Vasilevska e B. Rivza, «Factors of The Effectiveness of Innovative Development of Baltic States in the Context of Digitalization,» *International Multidisciplinary Scientific GeoConference: SGEM,* vol. 20, n. 2.1, pp. 153-160, 2020.

[16] P. O. Brafi e C. Arthur, «Internet Use among Students in Tertiary Institutions in the Sunyani Municipality, Ghana.,» *Library Philosophy & Practice.,* 2013.

[17] A. López Peláez, A. Erro-Garcés e E. J. Gómez-Ciriano, «Young people, social workers and social work education: the role of digital skills,» *Social Work Education,* vol. 39, n. 6, pp. 825-842, 2020.

[18] Z. Brixiová e M. &. I. S. Genčev, «The Digital Gender Gap and Entrepreneurship in Emerging Europe».

[19] P. Checcucci, «Recovery 4.0. Ageing labour markets, digitalization of the economy and Covid-19,» *2022.*

[20] H. Le Thanh, «Accelerating Digital Transformation Implementation in the Fight Against Corruption?: Evidence From European Countries Before and During the COVID-19 Pandemic.,» *International Journal of Electronic Government Research (IJEGR),* vol. 18, n. 2, pp. 1-2, 2022.

[21] E. Laitsou, A. Kargas e D. Varoutas, «Digital competitiveness in the European Union era: The Greek case,» *Economies,* vol. 85, n. 8, p. 4, 2020.

[22] G. Gravili, M. Benvenuto, A. Avram e C. Viola, «Big Data generation within supply chain management,» *The International Journal of Logistics Management,* vol. 2, n. 29, pp. 592-628, 2018.

[23] O. Başol e E. C. Yalçın, «Does the digital economy and society index (DESI) affect labor market indicators in EU countries?,» *Human Systems Management,* vol. 40, n. 4, pp. 503-512, 2021.

[24] E. Hargittai, «Digital Na (t) ives? Variation in Internet Skills and Uses among Members of the ''Net Generation'',» *Sociological Inquiry,* vol. 80, n. 1, pp. 92-113, 2009.

[25] E. Hargittai e S. Shafer, «Differences in actual and perceived online skills: The role of gender.,» *Social Science Quarterly,* vol. 87, n. 2, pp. 432-448, 2006.

[26] A. Leogrande, N. Magaletti, G. Cosoli, V. Giardinelli e A. Massaro, «ICT Specialists in Europe,» *Available at SSRN.,* 2022.

[27] A. Leogrande, N. Magaletti, G. Cosoli e A. Massaro, «e-Government in Europe. A Machine Learning Approach.,» *University Library of Munich, Germany.,* 2022.

[28] A. Leogrande, N. Magaletti, G. Cosoli e A. Massaro, «Broadband Price Index in Europe,» *University Library of Munich, Germany,* 2022.

[29] A. Leogrande, N. Magaletti, G. Cosoli e A. Massaro, «Fixed Broadband Take-Up in Europe,» *University Library of Munich, Germany,* 2022.

[30] A. Massaro, «Information Technology Infrastructures Supporting Industry 5.0 Facilities,» in *Electronics in Advanced Research Industries: Industry 4.0 to Industry 5.0 Advances*, 2022, pp. 51-101.

[31] A. Massaro, G. Dipierro, E. Cannella e A. M. Galiano, «Comparative analysis among discrete fourier transform, K-means and artificial neural networks image processing techniques oriented on quality control of assembled tires,» *Information,* vol. 11, n. 5, 2020.

[32] A. Massaro, A. Galiano, G. Meuli e S. F. Massari, «Overview and Application of Enabling Technologies Oriented on Energy Routing Monitoring, on Network Installation and on Predictive Maintenance,» *SSRN,* 2018.

[33] A. Massaro, I. Manfredonia, A. Galiano e N. Contuzzi, «Inline Image Vision Technique for Tires Industry 4.0: Quality and Defect Monitoring in Tires Assembly,» *2019 II Workshop on Metrology for Industry 4.0 and IoT,* pp. 54-57, 2019.

[34] A. Massaro e A. Galiano, in *Handbook of Research on Intelligent Data Processing and Information Security Systems*, 2020, pp. 117-146.

[35] A. Et-taleby, M. Boussetta e M. Benslimane, «Faults detection for photovoltaic field based on k-means, elbow, and average silhouette techniques through the segmentation of a thermal image,» *International Journal of Photoenergy,* 2020.

Figure Index

10. Appendix

The Econometric Model		
Variable	**Label**	**Regressor**
Internet User Skills	A10	y
Fixed broadband take-up	A1	x_1
Fixed broadband coverage	A2	x_2
Mobile broadband	A3	x_3
Broadband price index	A4	x_4
e-Government	A5	x_5
Advanced Skills and Development	A11	x_6
Digital intensity	A12	x_7
Digital technologies for businesses	A13	x_8
e-Commerce	A14	x_9
Integration of Digital Technology	A17	x_{10}
Aggregate score	A19	x_{11}
Electronic Information Sharing	A20	x_{12}
5G coverage	A42	x_{13}
At least Basic Digital Skills	A44	x_{14}
Above basic digital skills	A45	x_{15}
At least basic software skills	A46	x_{16}
SMEs with at least a basic level of digital intensity	A47	x_{17}

Pooled OLS, using 168 observations					
Including 28 units of cross section					
Time series length = 6					
Variable: A10					
	Coefficient	*Standard Error.*	*t*	*p-value*	
Const	4,78304e-05	3,08664e-05	1,550	0,1233	
A1	−0,0517111	0,0216093	−2,393	0,0179	**
A2	−0,0517122	0,0216096	−2,393	0,0179	**
A3	−0,0517133	0,0216095	−2,393	0,0179	**
A4	−0,0517115	0,0216096	−2,393	0,0179	**
A5	−0,0517118	0,0216095	−2,393	0,0179	**
A11	−0,0517104	0,0216095	−2,393	0,0180	**

A12	6,22267e-05	2,87284e-05	2,166	0,0319	**
A13	−1,59075e-05	7,46066e-06	−2,132	0,0346	**
A14	−1,58119e-05	7,59374e-06	−2,082	0,0390	**
A17	−0,206790	0,0864403	−2,392	0,0180	**
A19	0,206847	0,0864381	2,393	0,0179	**
A20	5,72759e-06	2,79269e-06	2,051	0,0420	**
A42	2,08094e-06	9,37100e-07	2,221	0,0279	**
A44	0,474151	0,0108050	43,88	<0,0001	***
A45	0,474149	0,0108045	43,88	<0,0001	***
A46	0,474123	0,0108045	43,88	<0,0001	***
A47	−1,07546e-05	5,28576e-06	−2,035	0,0436	**

Mean Dependent Variabile	26,80277	Standard Deviation Dependend Variable	6,213459
Residual Sum of Squares	1,70e-07	Standard Error of the Regression	0,000034
R-squared	1,000000	Adjusted R-Squared	1,000000
F(17, 150)	3,34e+11	P-value(F)	0,000000
Log-Likelihood	1501,232	Akaike Criterion	−2966,464
Schwarz Criterion	−2910,233	Hannan-Quinn	−2943,643
Rho	0,224540	Durbin-Watson	1,274138

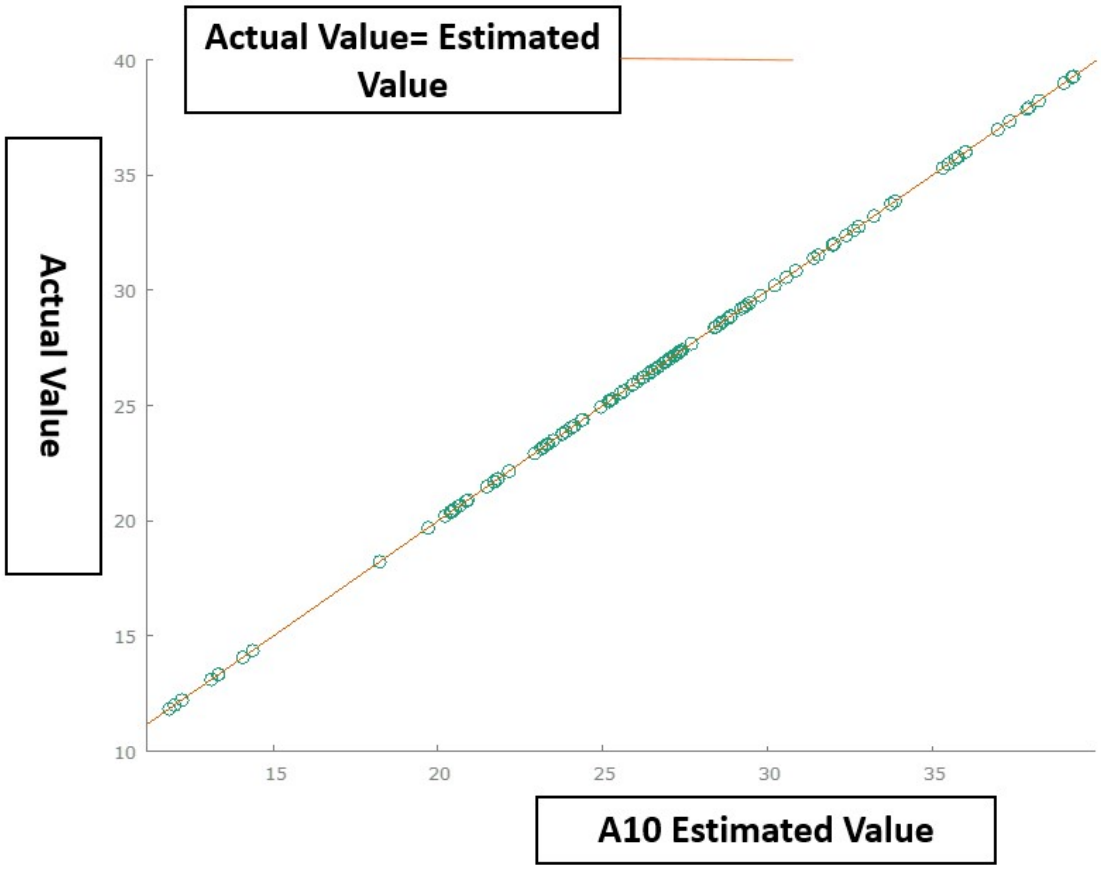

Figure 4. Actual Value and Estimated Value with POOLED OLS.

Fixed Effects Using 168 observations					
Including 28 cross section units					
Time series length = 6					
Dependent variable: A10					
	Coefficient	*Std. Error*	*T*	*p-value*	
const	−6,16572e-05	9,27773e-05	−0,6646	0,5076	
A1	−0,0478424	0,0222529	−2,150	0,0335	**
A2	−0,0478432	0,0222531	−2,150	0,0335	**
A3	−0,0478419	0,0222533	−2,150	0,0335	**
A4	−0,0478474	0,0222536	−2,150	0,0335	**
A5	−0,0478363	0,0222530	−2,150	0,0335	**
A11	−0,0478438	0,0222530	−2,150	0,0335	**
A12	8,83219e-05	3,06574e-05	2,881	0,0047	***

A13	−2,18245e-05	8,01318e-06	−2,724	0,0074	***
A14	−1,79883e-05	1,01133e-05	−1,779	0,0778	*
A17	−0,191285	0,0890155	−2,149	0,0336	**
A19	0,191362	0,0890129	2,150	0,0335	**
A20	1,06507e-05	5,78387e-06	1,841	0,0680	*
A42	2,11688e-06	1,07370e-06	1,972	0,0509	*
A44	0,476092	0,0111259	42,79	<0,0001	***
A45	0,476082	0,0111268	42,79	<0,0001	***
A46	0,476058	0,0111275	42,78	<0,0001	***
A47	−1,82564e-05	5,83969e-06	−3,126	0,0022	***

Mean Dependent Variabile	26,80277	*Standard Deviation Dependend Variable*	6,213459
Residual Sum of Squares	1,23e-07	*Regression Standard Error*	0,000032
R-squared LSDV	1,000000	*R-Squared Among Groups*	1,000000
LSDV F(44, 123)	1,46e+11	*P-value(F)*	0,000000
Log-likelihood	1528,320	*Akaike Criterion*	−2966,640
Schwarz Criterion	−2826,062	*Hannan-Quinn*	−2909,587
Rho	−0,003259	*Durbin-Watson*	1,743037

Joint regressor test -

Test statistic: F (17, 123) = 1.16576e + 10

with p-value = P (F (17, 123)> 1.16576e + 10) = 0

Group Intercept Difference Test -

Null hypothesis: groups have a common intercept

Test statistic: F (27, 123) = 1.73359

with p-value = P (F (27, 123)> 1.73359) = 0.0231889

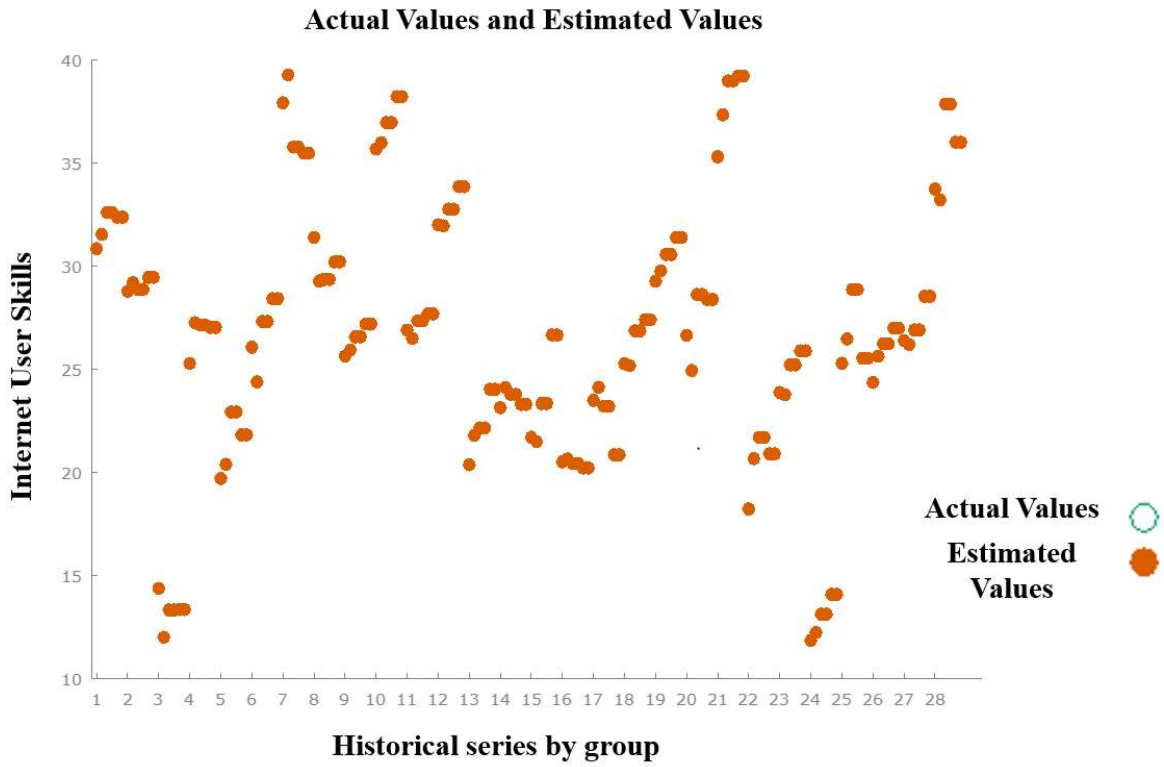

Figure 5. Actual Values and Estimated Values with Fixed Effects.

Random Effects (GLS), using 168 observations					
Including 28 cross section units					
Time series length = 6					
Dependent variable: A10					
	Coefficient	*Std. Error*	*z*	*p-value*	
Const	3,97879e-05	3,90486e-05	1,019	0,3082	
A1	−0,0482625	0,0210199	−2,296	0,0217	**
A2	−0,0482629	0,0210201	−2,296	0,0217	**
A3	−0,0482632	0,0210201	−2,296	0,0217	**
A4	−0,0482623	0,0210202	−2,296	0,0217	**
A5	−0,0482622	0,0210201	−2,296	0,0217	**
A11	−0,0482612	0,0210200	−2,296	0,0217	**
A12	6,92157e-05	2,78975e-05	2,481	0,0131	**
A13	−1,78195e-05	7,28023e-06	−2,448	0,0144	**
A14	−1,71356e-05	7,55204e-06	−2,269	0,0233	**
A17	−0,192985	0,0840829	−2,295	0,0217	**
A19	0,193049	0,0840805	2,296	0,0217	**
A20	6,66946e-06	3,34175e-06	1,996	0,0460	**

A42	1,75214e-06	9,26744e-07	1,891	0,0587	*
A44	0,475876	0,0105102	45,28	<0,0001	***
A45	0,475873	0,0105099	45,28	<0,0001	***
A46	0,475847	0,0105098	45,28	<0,0001	***
A47	−1,22220e-05	5,15065e-06	−2,373	0,0176	**

Mean Dependent Variabile	26,80277	Standard Deviation Dependent Variable	6,213459
Residual Sum of Squares	1,72e-07	Standard Error of Regression	0,000034
Log-Likelihood	1500,220	Akaike Criterion	−2964,439
Schwarz Criterion	−2908,208	Hannan-Quinn	−2941,618
Rho	−0,003259	Durbin-Watson	1,743037

Variance 'between' = 2.26076e-010

Variance 'within' = 1.00289e-009

Theta used for transformation = 0.348026

Joint regressor test -

Asymptotic test statistic: Chi-square (17) = 2.83246e + 12

with p-value = 0

Breusch-Pagan Test -

Null hypothesis: variance of unit-specific error = 0

Asymptotic test statistic: Chi-square (1) = 0.164076

with p-value = 0.685431

Hausman test -

Null hypothesis: GLS estimates are consistent

Asymptotic test statistic: Chi-square (17) = 22.4226

with p-value = 0.169009

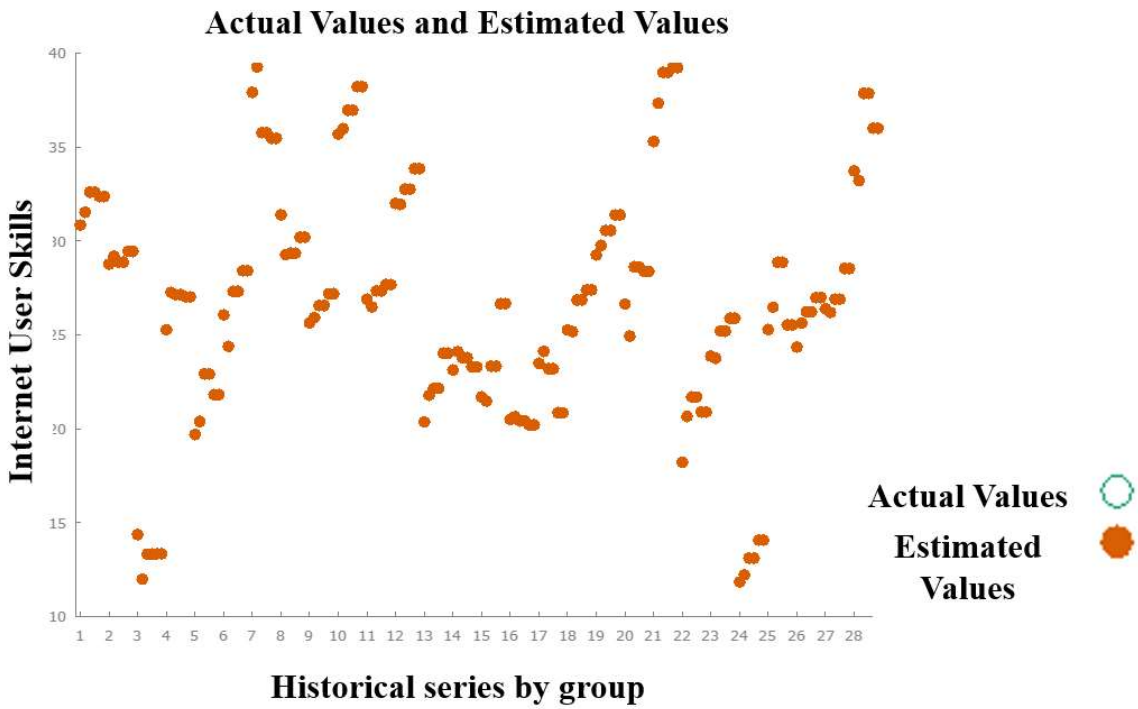

Figure 6. Actual Values and Estimated values of Panel Data with Random Effects.

WLS, using 168 observations					
Including 28 cross section units					
Dependent variable: A10					
Weights based on variances of errors per unit					
	Coefficient	*Std. Errore*	*t*	*p-value*	
const	4,91258e-05	2,95855e-05	1,660	0,0989	*
A1	−0,0350610	0,0193772	−1,809	0,0724	*
A2	−0,0350607	0,0193774	−1,809	0,0724	*
A3	−0,0350624	0,0193773	−1,809	0,0724	*
A4	−0,0350609	0,0193774	−1,809	0,0724	*
A5	−0,0350607	0,0193773	−1,809	0,0724	*
A11	−0,0350594	0,0193773	−1,809	0,0724	*
A12	6,70135e-05	2,79084e-05	2,401	0,0176	**
A13	−1,73654e-05	7,14952e-06	−2,429	0,0163	**
A14	−1,79244e-05	7,23964e-06	−2,476	0,0144	**
A17	−0,140178	0,0775118	−1,808	0,0725	*
A19	0,140242	0,0775094	1,809	0,0724	*
A20	5,53843e-06	2,61885e-06	2,115	0,0361	**

A42	2,11491e-06	8,08161e-07	2,617	0,0098	***
A44	0,482482	0,00968896	49,80	<0,0001	***
A45	0,482472	0,00968827	49,80	<0,0001	***
A46	0,482440	0,00968844	49,80	<0,0001	***
A47	−1,15391e-05	5,13738e-06	−2,246	0,0262	**

Statistics based on weighted data:

Residual Sum of Squares	164,5208	Standard Error of Regression	1,047285
R-Squared	1,000000	Adjusted R-Squared	1,000000
F(17, 150)	3,84e+11	P-value(F)	0,000000
Log-Likelihood	−236,6238	Akaike Criterion	509,2476
Schwarz Criterion	565,4789	Hannan-Quinn	532,0690

Statistics based on original data:

Mean Dependent Variabile	26,80277	Standard Error Dependent Variable	6,213459
Residual Sum of Squares	1,73e-07	Standard Error of Regression	0,000034

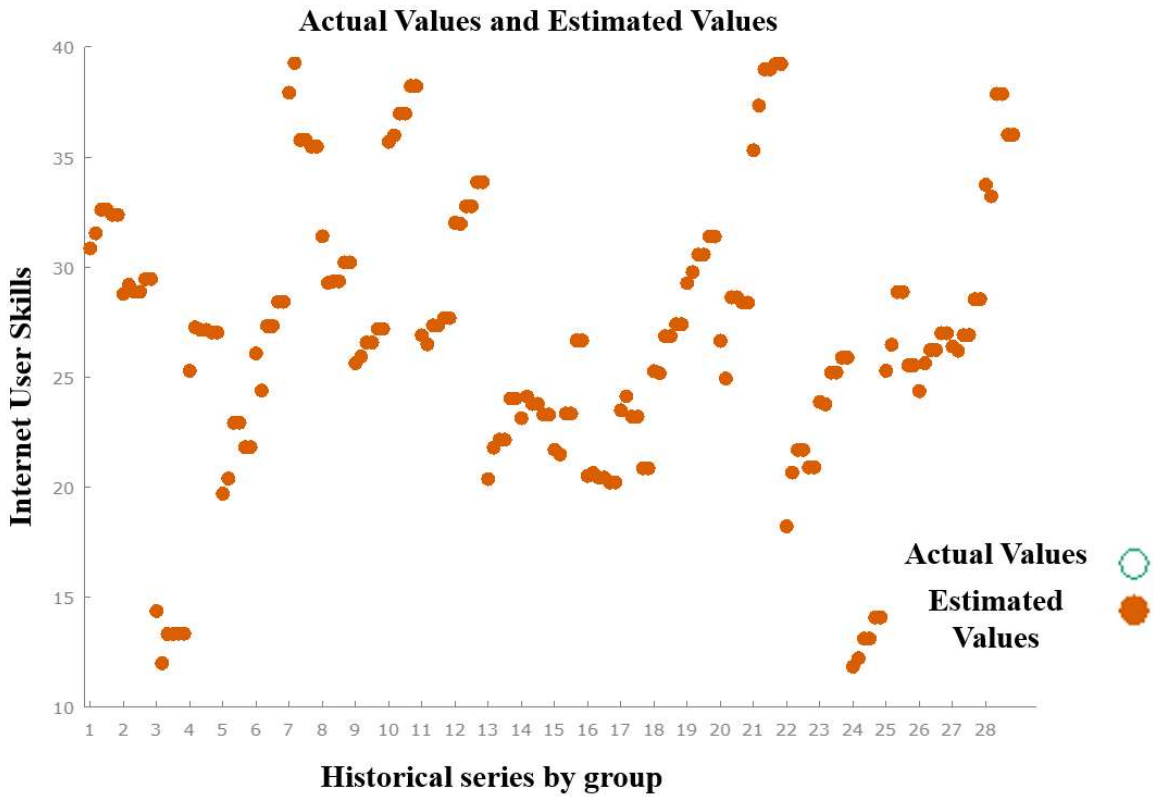

Figure 7. Actual Values and Estimated Values with WLS.

WLS corrected for heteroskedasticity, using 168 observations					
Dependent variable: A10					
	Coefficient	*Std. Error*	*t*	*p-value*	
const	2,49760e-05	2,78312e-05	0,8974	0,3709	
A1	−0,0496907	0,0174156	−2,853	0,0049	***
A2	−0,0496932	0,0174160	−2,853	0,0049	***
A3	−0,0496945	0,0174158	−2,853	0,0049	***
A4	−0,0496902	0,0174158	−2,853	0,0049	***
A5	−0,0496924	0,0174158	−2,853	0,0049	***
A11	−0,0496911	0,0174158	−2,853	0,0049	***
A12	5,05376e-05	1,92534e-05	2,625	0,0096	***
A13	−1,09711e-05	5,66562e-06	−1,936	0,0547	*
A14	−1,00977e-05	5,71674e-06	−1,766	0,0794	*
A17	−0,198733	0,0696634	−2,853	0,0049	***
A19	0,198770	0,0696634	2,853	0,0049	***
A20	5,80097e-06	2,33651e-06	2,483	0,0141	**
A42	2,25761e-06	6,92436e-07	3,260	0,0014	***
A44	0,475160	0,00870836	54,56	<0,0001	***

A45	0,475158	0,00870740	54,57	<0,0001	***
A46	0,475134	0,00870750	54,57	<0,0001	***
A47	−8,41944e-06	3,60787e-06	−2,334	0,0209	**

Statistics based on weighted data:			
Residual Sum of Squares	689,1122	Standard Error of the Regression	2,143381
R-squared	1,000000	Adjusted R-Squared	1,000000
F(17, 150)	4,58e+11	P-value(F)	0,000000
Log-Likelihood	−356,9426	Akaike Criterion	749,8853
Schwarz Criterion	806,1166	Hannan-Quinn	772,7067

Statistics based on original data:			
Mean Dependent Variabile	26,80277	Standard Deviation of Dependent Variable	6,213459
Residual Sum of Squares	1,75e-07	Standard Error of the Regression	0,000034

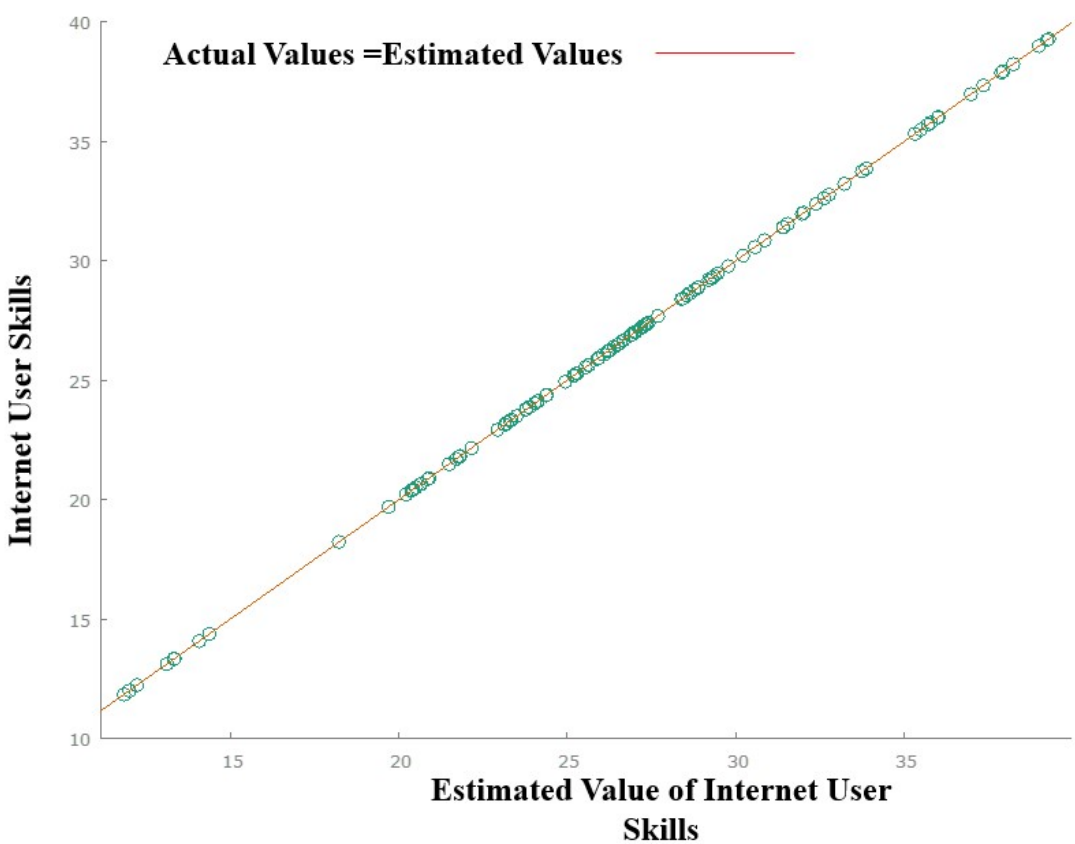

Figure 8. Estimated Value and Actual Value of WLS corrected for heteroskedasticity.

10.2 Clusterization

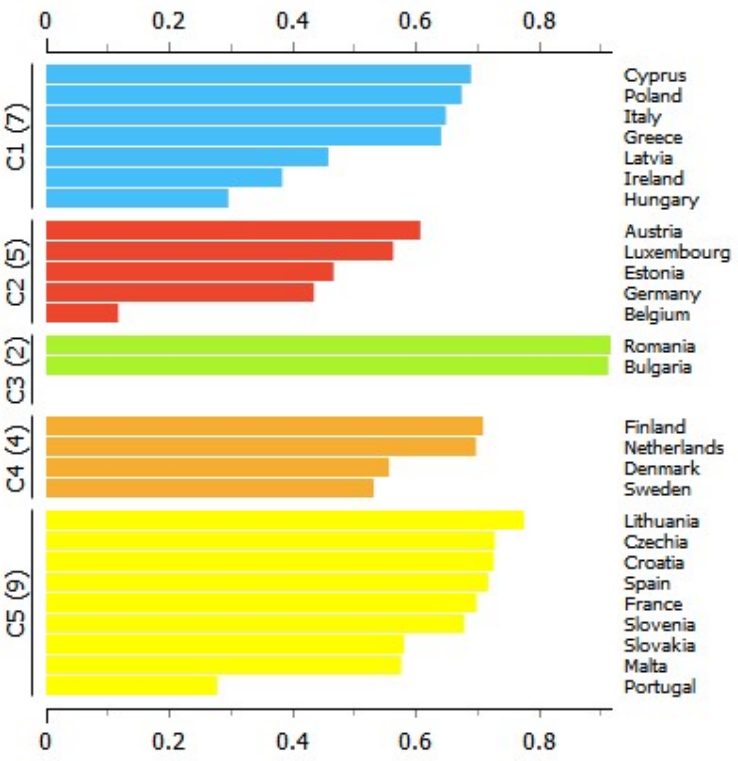

Figure 9. Silhouette Coefficient. K=5. k-Means Algorithm.

	2021	Countries	Cluster	Silhouette
5	21.8221	Cyprus	C1	0.678113
12	24.0346	Greece	C1	0.672033
13	23.3018	Hungary	C1	0.582173
14	26.6679	Ireland	C1	0.584887
15	20.2178	Italy	C1	0.67423
16	20.8585	Latvia	C1	0.621457
21	20.9058	Poland	C1	0.677444
1	32.3794	Austria	C2	0.67108
2	29.4622	Belgium	C2	0.542163
8	30.2154	Estonia	C2	0.636291
11	33.8654	Germany	C2	0.632168
18	31.3978	Luxembourg	C2	0.660519
3	13.3506	Bulgaria	C3	0.726987
23	14.0844	Romania	C3	0.727906
7	35.4817	Denmark	C4	0.657052
9	38.2305	Finland	C4	0.691501
20	39.2354	Netherlands	C4	0.689851
27	36.0145	Sweden	C4	0.647121
4	27.0347	Croatia	C5	0.69165
6	28.4300	Czechia	C5	0.694297
10	27.6888	France	C5	0.686119
17	27.4068	Lithuania	C5	0.705503
19	28.3866	Malta	C5	0.664874
22	25.8958	Portugal	C5	0.588751
24	25.5389	Slovakia	C5	0.665358
25	26.9987	Slovenia	C5	0.68649
26	28.5461	Spain	C5	0.690724

Figure 10. Silhouette Results of Clusterization.

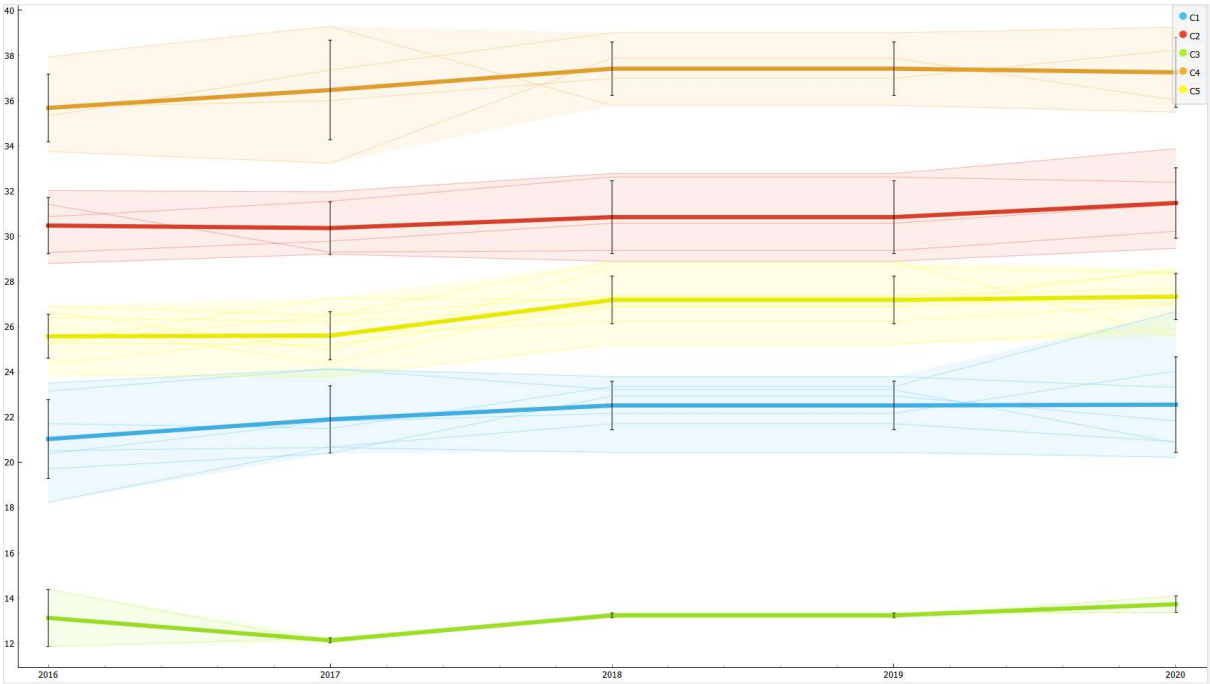

Figure 11. Line Plot of Clusters.

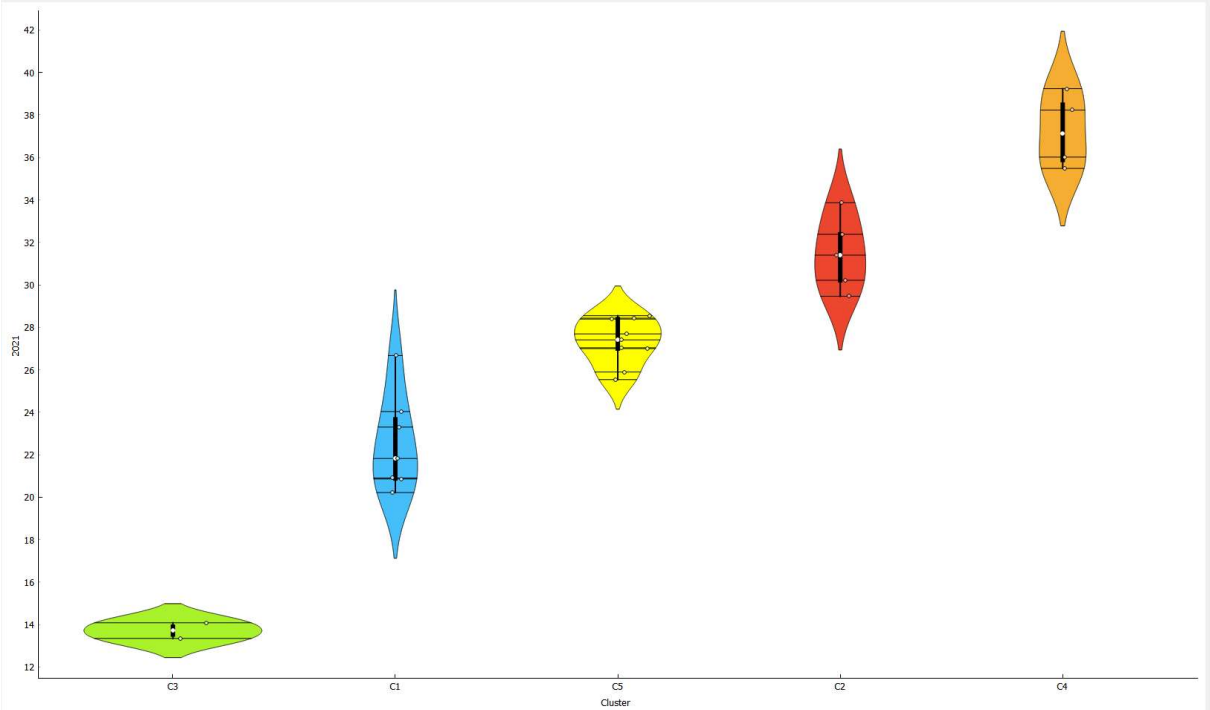

Figure 12. Violin Plot of Clusters.

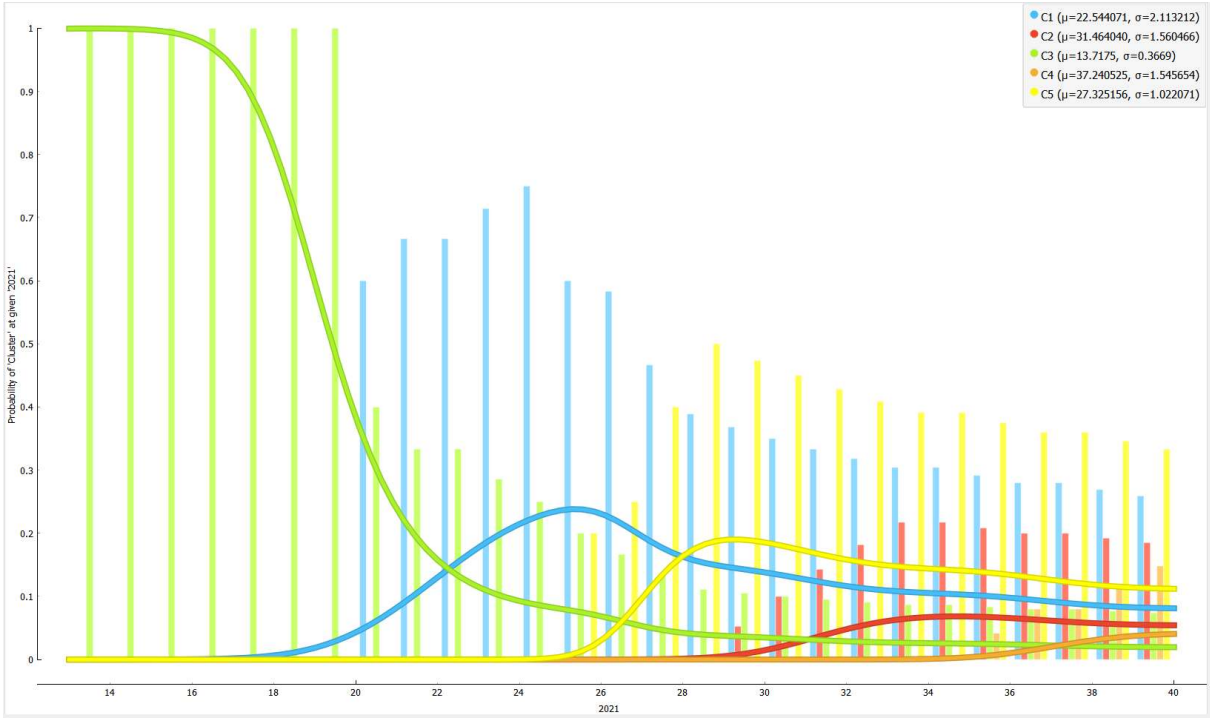

Figure 13. Distributions of clusters.

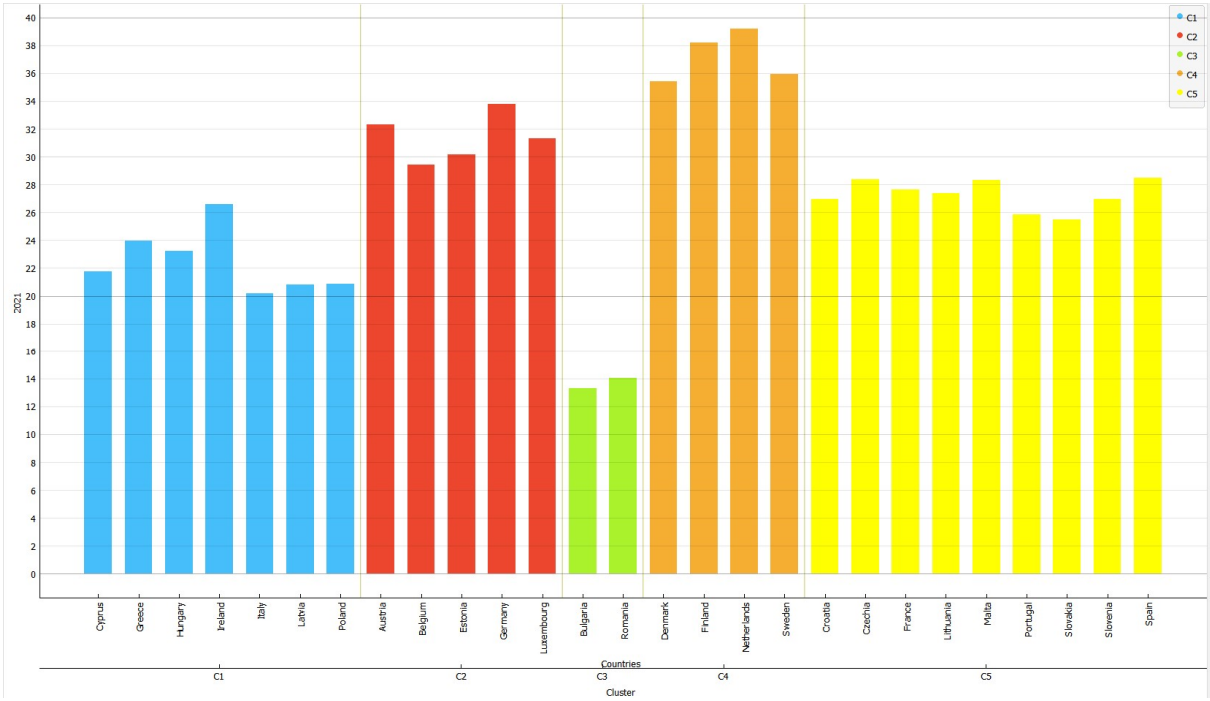

Figure 14. Cluster bar plot.

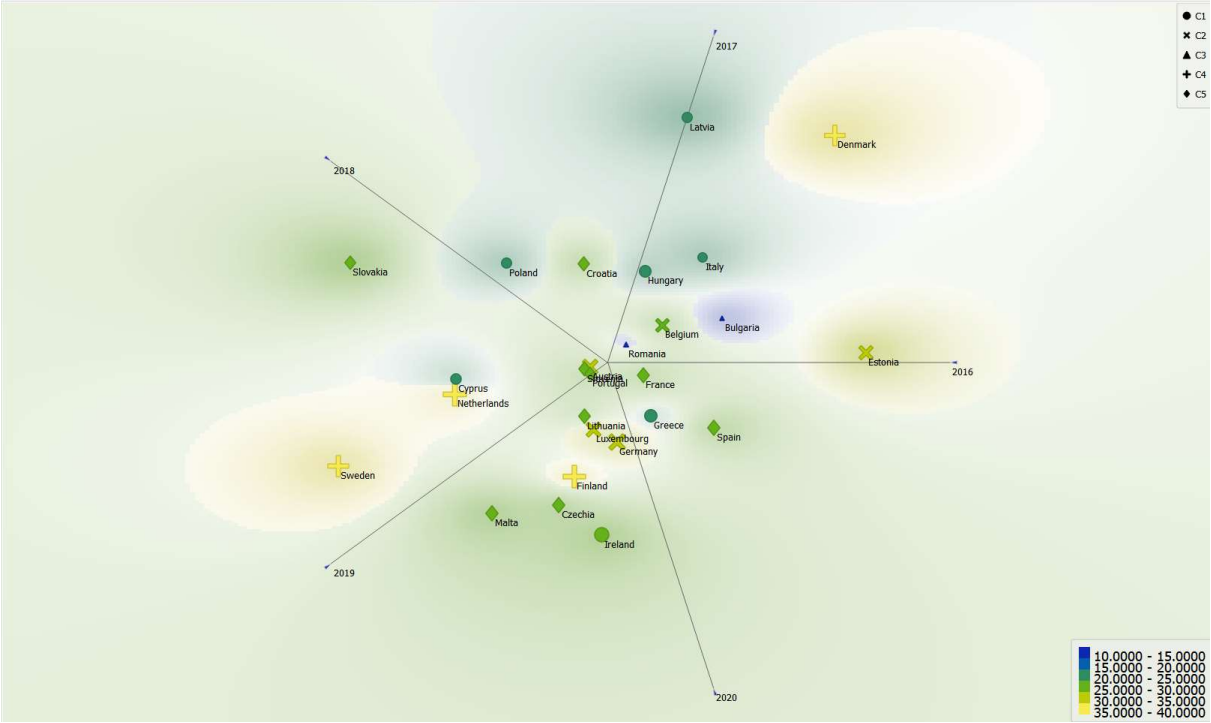

Figure 15. Free Viz of Clusters.

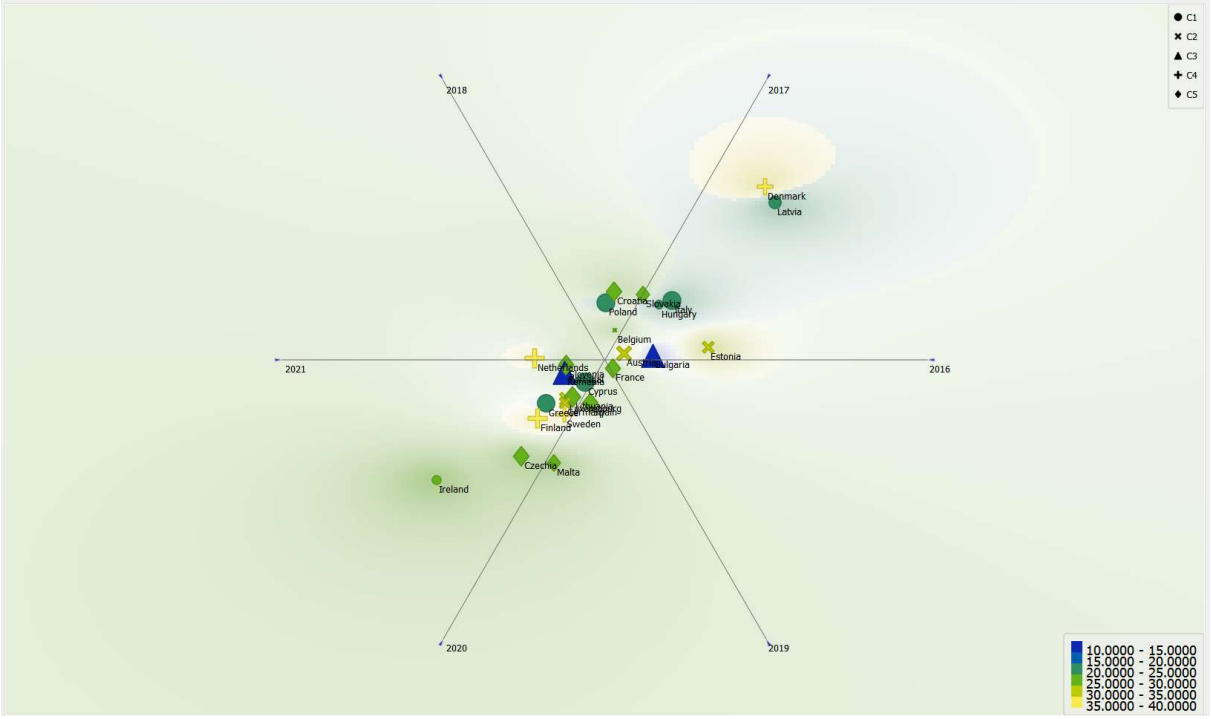

Figure 16. Linear Projection of clusters.

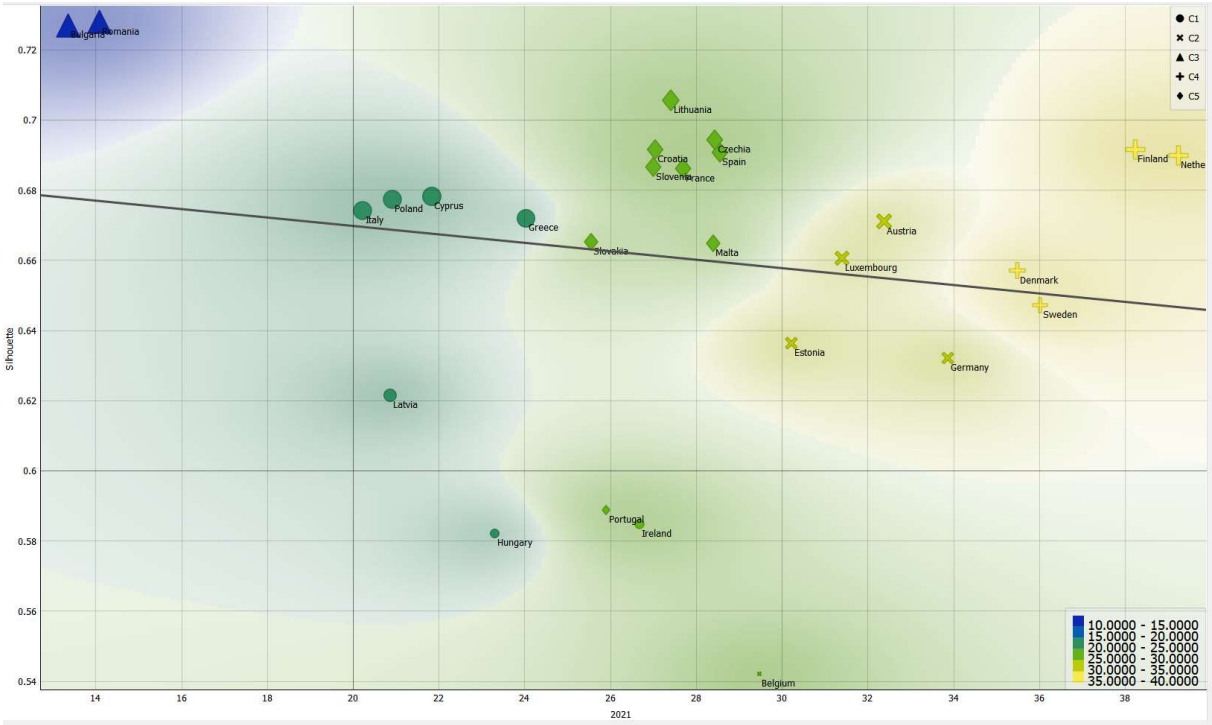

Figure 17. Scatter Plot of Clusters.

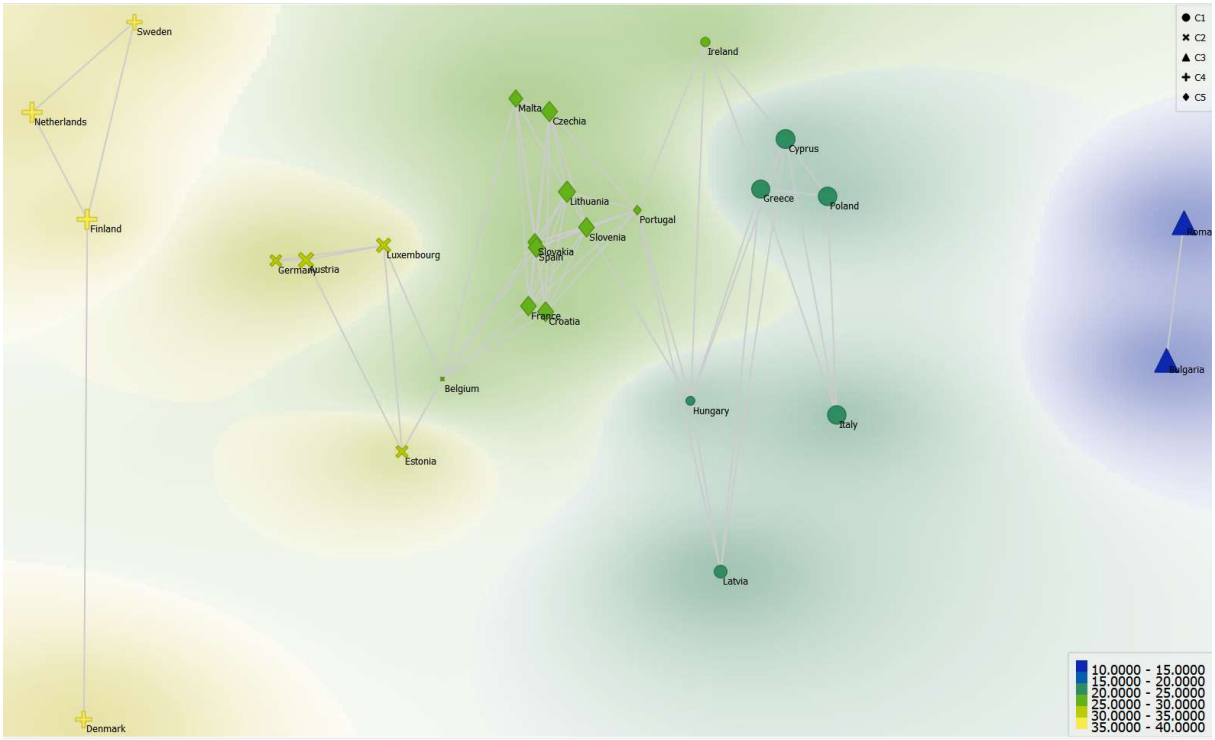

Figure 18. Multidimensional Scaling Dimension of Clusters.

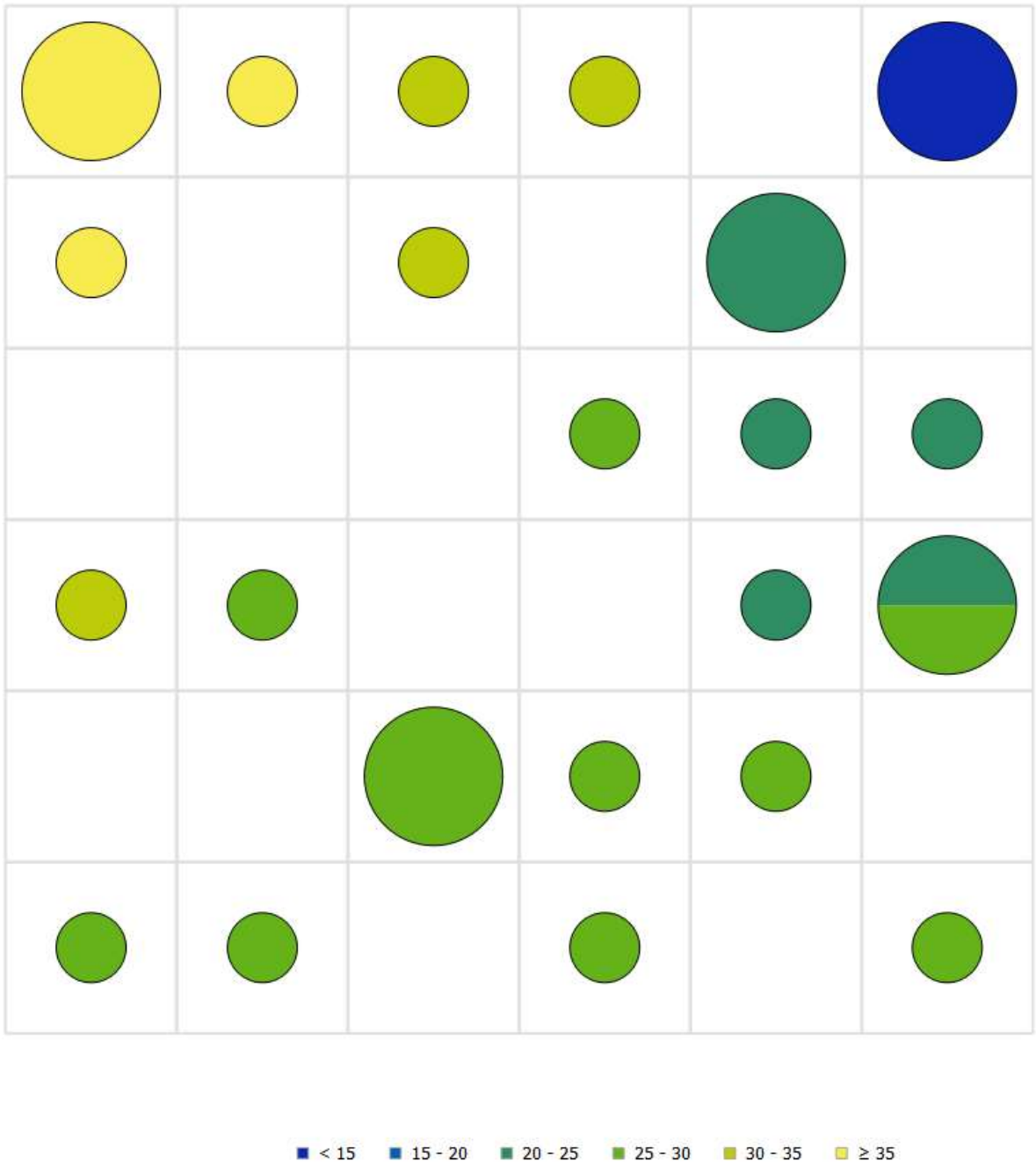

Figure 19. DBSCAN of Clusters.

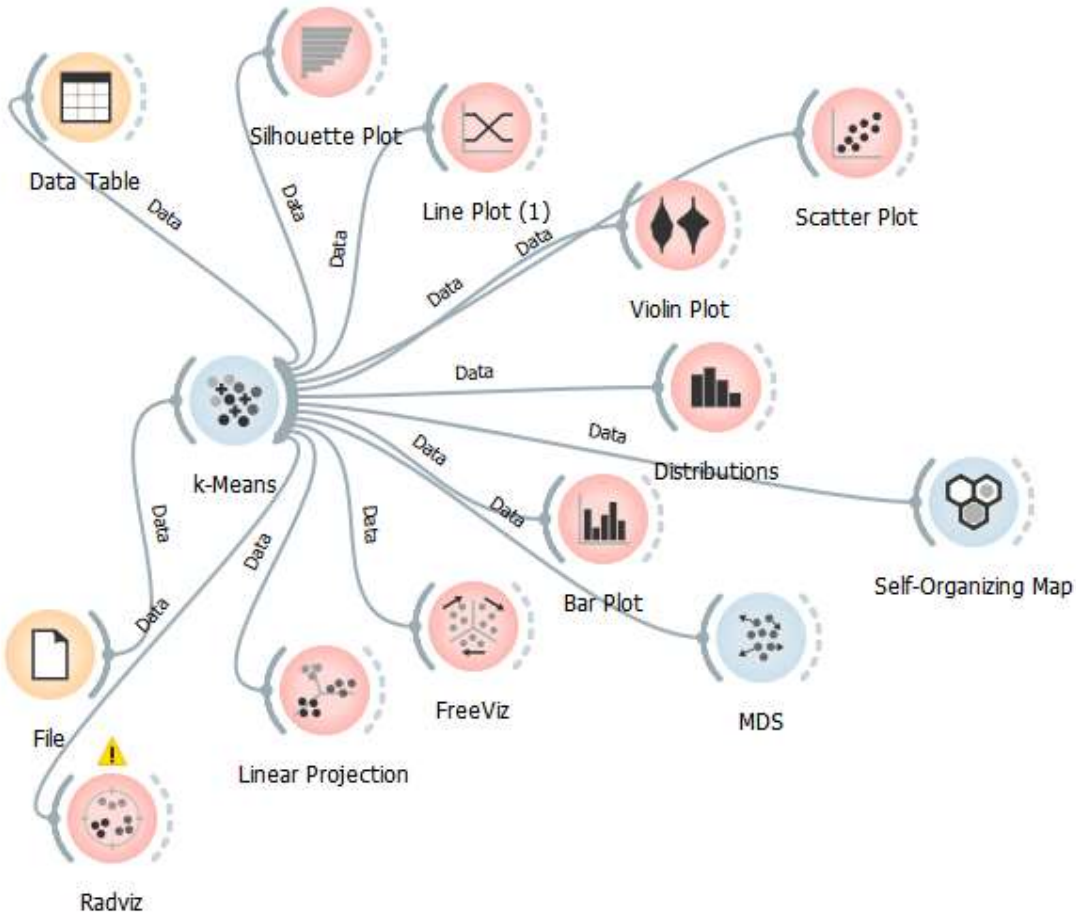

Figure 20. Orange workflow for clusterization with the k-Means algorithm.

10.3 Elbow Method

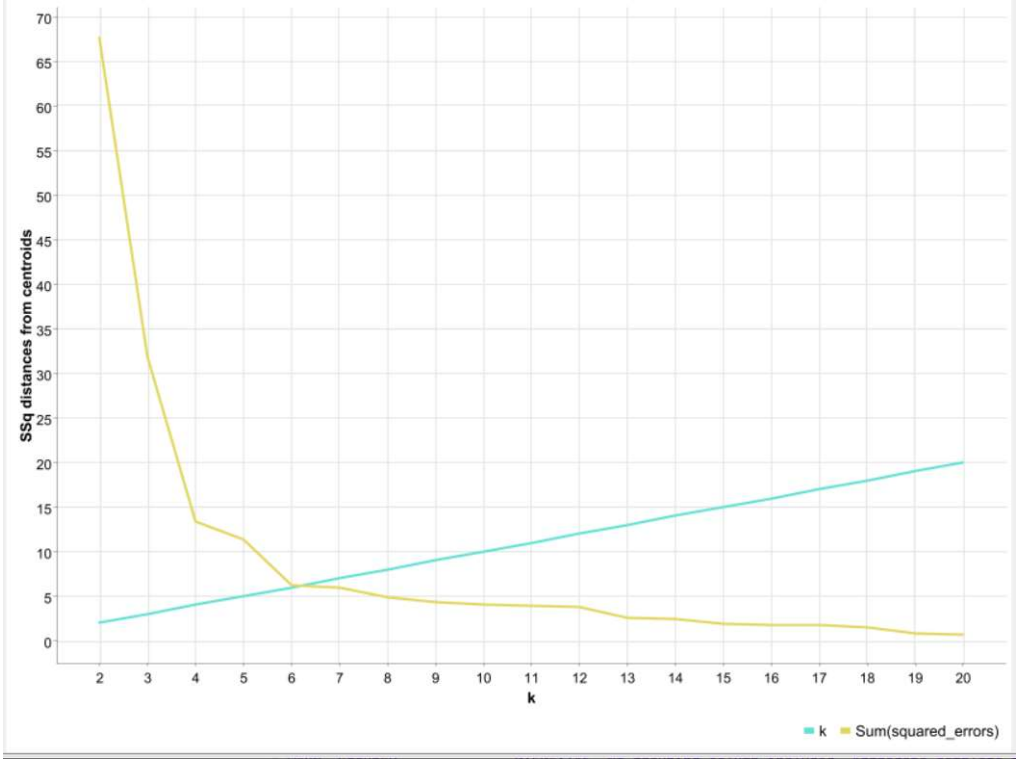

Figure 21. Grahipical Result of the Elbow Method.

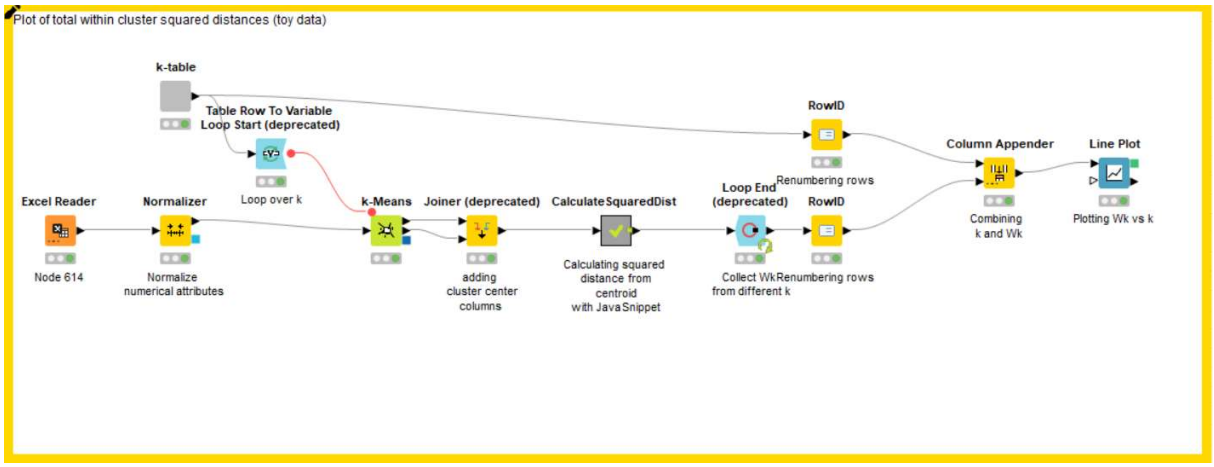

Figure 22. Workflow of Elbow Method with KNIME.

10.4 Network Analysis

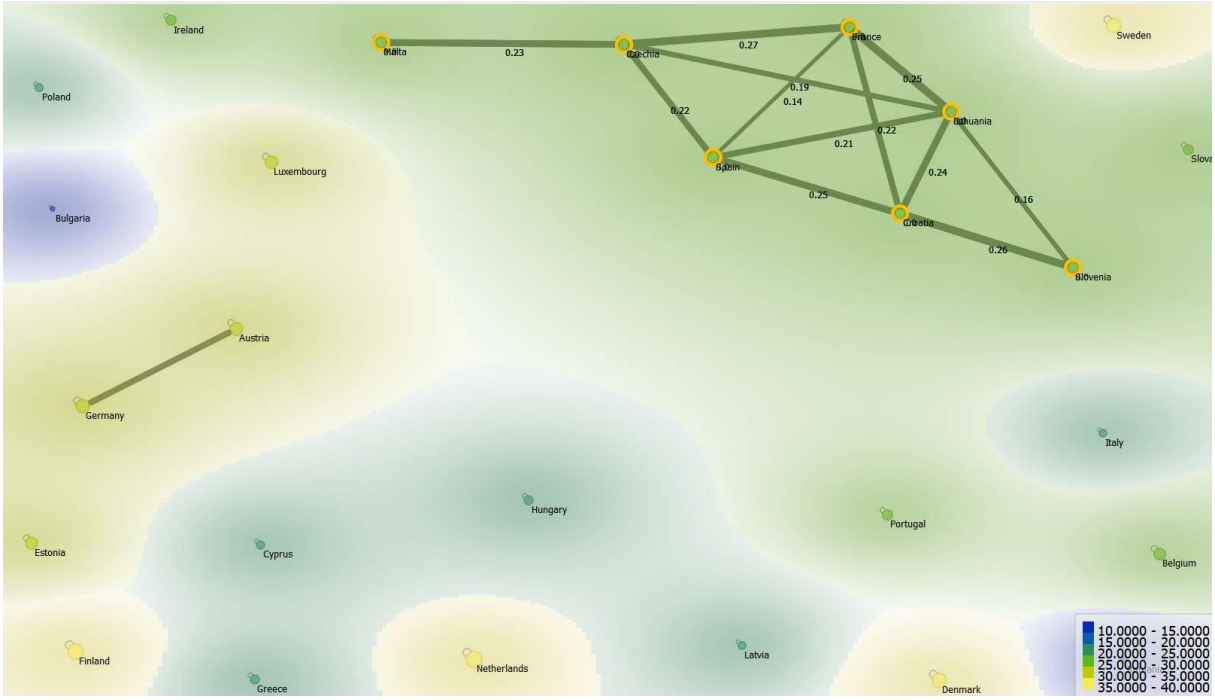

Figure 23. Graphical Representation of Network Analysis with Manhattan Distance.

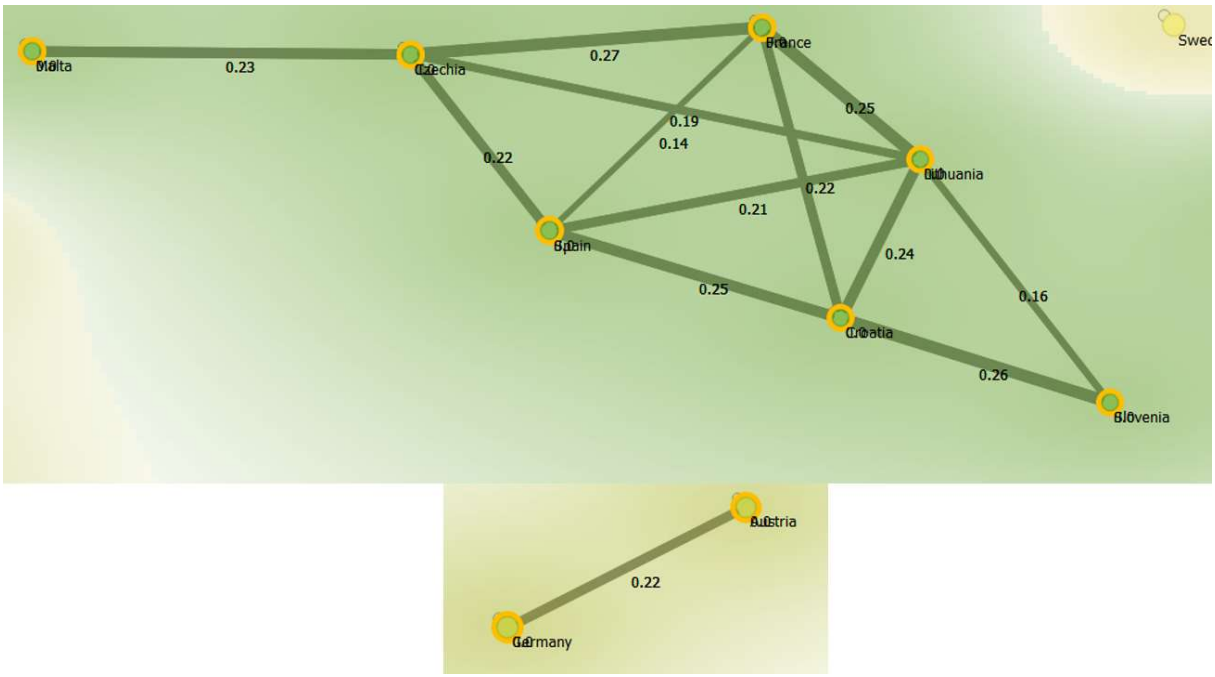

Figure 24. Focus of the Graphical Representation of the Main Structure of the Network Analysis.

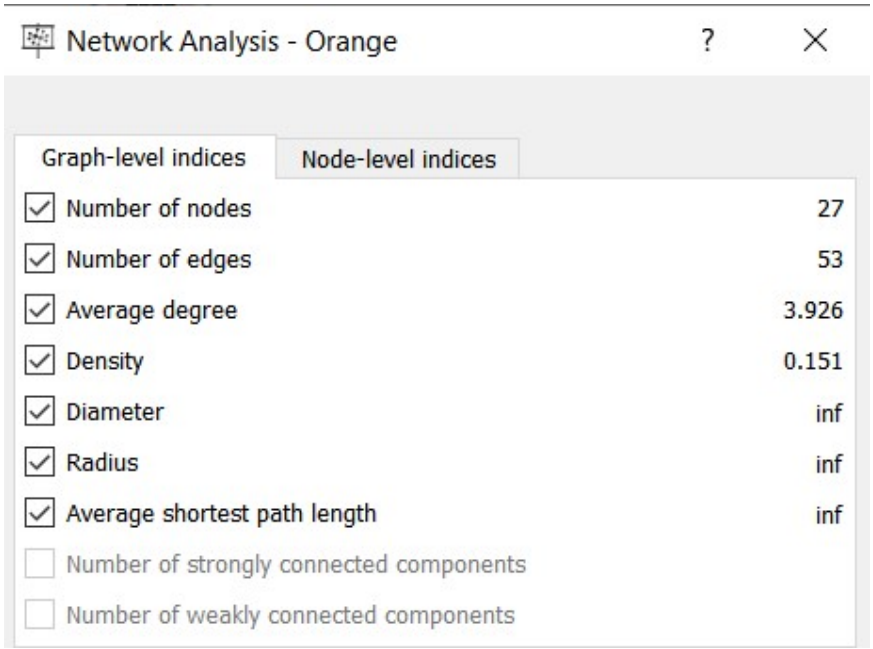

Figure 25. Metrics of Network Analysis.

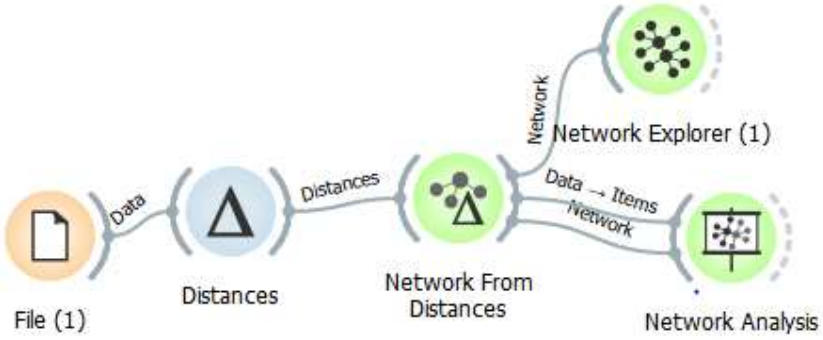

Figure 26. Workflow of the Network Analysis with Orange.

10.5 Machine Learning and Predictions with Original Data

Results of the Application of Machine Learning Algorithms for the Prediction of Internet User Skills.			
Statistical Measures	**ANN**	**PNN**	**Simple Regression Tree**
R^2	0,96388563	0,96323112	0,94205371
Mean absolute error	0,02935639	0,04174329	0,04368651
Mean squared error	0,00224388	0,00228454	0,00360035
Root mean squared error	0,04736959	0,04779691	0,06000293
Mean signed difference	0,02434017	0,02125379	0,03218298
Mean absolute percentage error	0,26268163	0,22806099	0,22812635
Adjusted R^2	0,96388563	0,96323112	0,94205371
Statistical Measures	**Gradient Boosted Tree Regression**	**Random Forest Regression**	**Tree ensemble Regression**
R^2	0,96497873	0,93988712	0,95442322
Mean absolute error	0,03765729	0,05100479	0,04263166
Mean squared error	0,00217596	0,00373497	0,00283180
Root mean squared error	0,04664720	0,06111438	0,05321468
Mean signed difference	0,01513549	0,02828109	0,02149758
Mean absolute percentage error	0,20030048	0,64628321	0,48480555
Adjusted R^2	0,96497873	0,93988712	0,95442322

Figure 27. Results of the Machine Learning Analysis for the Prediction of Internet User Skills.

Ranking of Algorithms for the Prediction with Original Data of the Value of Internet User Skills							
Algorithms	**R^2**	**Mean absolute error**	**Mean squared error**	**Root mean squared error**	**Mean absolute percentage error**	**Adjusted R^2**	**Sum**
Gradient Boosted Tree Regression	1	2	1	1	1	1	7
ANN	2	1	2	2	4	2	13
PNN	3	3	3	3	2	3	17
Tree ensemble Regression	4	4	4	4	5	4	25
Simple Regression Tree	5	5	5	5	3	5	28
Random Forest Regression	6	6	6	6	6	6	36

Figure 28. Ranking of Alghorithms for the Prediction with Original Data of the Value of Internet User Skills.

Results of the Prediction Using Gradient Boosted Tree Regression									
Country	**2021**	**Prediction**	**Absolute Variation**	**Percentage Variation**	**Country**	**2021**	**Prediction**	**Absolute Variation**	**Percentage Variation**
Austria	32,38	31,40	-0,98	-3,03	**Italy**	20,22	21,82	1,60	7,93
Belgium	27,69	28,46	0,77	2,80	**Latvia**	20,86	25,90	5,04	24,15
Bulgaria	13,35	14,09	0,73	5,50	**Lithuania**	27,41	26,89	-0,51	-1,88
Croatia	27,04	27,41	0,37	1,38	**Luxembourg**	31,39	29,46	-1,93	-6,14
Cyprus	21,82	20,22	-1,60	-7,35	**Malta**	28,84	28,43	-0,41	-1,41
Czechia	28,43	28,39	-0,04	-0,15	**Netherlands**	39,24	38,23	-1,01	-2,56
Denmark	35,48	38,23	2,75	7,74	**Poland**	20,91	21,77	0,87	4,15
Estonia	30,22	31,70	1,48	4,91	**Portugal**	25,90	25,54	-0,36	-1,38
Finland	38,23	39,25	1,02	2,68	**Romania**	14,08	13,35	-0,73	-5,20
France	27,69	28,04	0,35	1,26	**Slovakia**	25,54	26,51	0,97	3,78
Germany	33,87	36,01	2,15	6,35	**Slovenia**	28,55	28,43	-0,12	-0,41
Greece	24,04	21,82	-2,21	-9,21	**Spain**	28,55	28,37	-0,18	-0,63
Hungary	23,30	25,90	2,59	11,13	**Sweden**	36,01	38,50	2,49	6,91
Ireland	26,67	26,47	-0,20	-0,74	**Mean**	27,32	27,80	0,48	1,75

Figure 29. Results of the Predictions Using Gradient Boosted Tree Regression of the Value of Internet User Skills.

10.6 Machine Learning and Predictions with Increased Data

Results of the Machine Learning Training with Augmented Data				
Algorithms	**R^2**	**Mean absolute error**	**Mean squared error**	**Root mean squared error**
ANN	0,934993557	0,058670032	0,005242638	0,072406064
PNN	0,985315343	0,030300106	0,001184288	0,034413484
Simple Regression Tree	0,987788504	0,026053329	0,001000000	0,031382041
Gradient Boosted Tree Regression	0,762000000	2,855000000	17,792000000	4,218000000
Random Forest Regression	0,935400707	0,059069570	0,005209802	0,072178960
Tree Ensemble Regression	0,947468747	0,056234010	0,004236539	0,065088703

Figure 30. Results of the Machine Learning Training with Increased Data.

Predictions with Simple Regression Tree Trained with Augmented Data				
Country	**Prediction**	**Prediction(prediction)**	**Absolute Variation**	**Percentage Variation**
Austria	0,697	0,708	0,011	1,578
Croatia	0,543	0,582	0,039	7,182
Denmark	0,961	0,961	0,000	0,000
Finland	1,000	0,961	-0,039	-3,900
Germany	0,875	0,708	-0,167	-19,086
Hungary	0,484	0,484	0,000	0,000
Ireland	0,507	0,327	-0,180	-35,503
Luxembourg	0,622	0,583	-0,039	-6,270
Romania	0,000	0,028	0,028	0,000
Mean	0,632	0,594	-0,039	-6,099

Figure 31. Prediction with Simple Regression Tree Trained with Increased Data.

Ranking of Algorithms in Terms of Prediction Performance with Augmented Data					
Algorithm	**R^2**	**Mean absolute error**	**Mean squared error**	**Root mean squared error**	**Sum**
Simple Regression Tree	1	1	1	1	4
PNN	2	2	2	2	8
Tree Ensemble Regression	3	3	3	3	12
Random Forest Regression	4	4	4	4	16
ANN	5	5	5	5	20
Gradient Boosted Tree Regression	6	6	6	6	24

Figure 32. Ranking of Algorithms in Terms of Prediction Performance with Increased Data.

Prediction with Original Data Vs Prediciton with Augmented Data				
Statistics	**Gradient Boosted Tree Regression**	**Simple Regression Tree**	**Absolute Variation**	**Percentage Variation**
R^2	0,96497873	0,987788504	0,022809772	2,363758992
Mean absolute error	0,03765729	0,026053329	-0,011603961	-30,81464606
Mean squared error	0,00217596	0,001000000	-0,001175961	-54,04329252
Root mean squared erro	0,04664720	0,031382041	-0,015265156	-32,7247017
Mean of Errors	0,02882682	0,01947846	-0,009348359	-32,42938526

Figure 33. Prediction with Original Data vs Prediction with Augmented Data.

Figure 34. Workflow of Machine Learning Algorithms with KNIME.

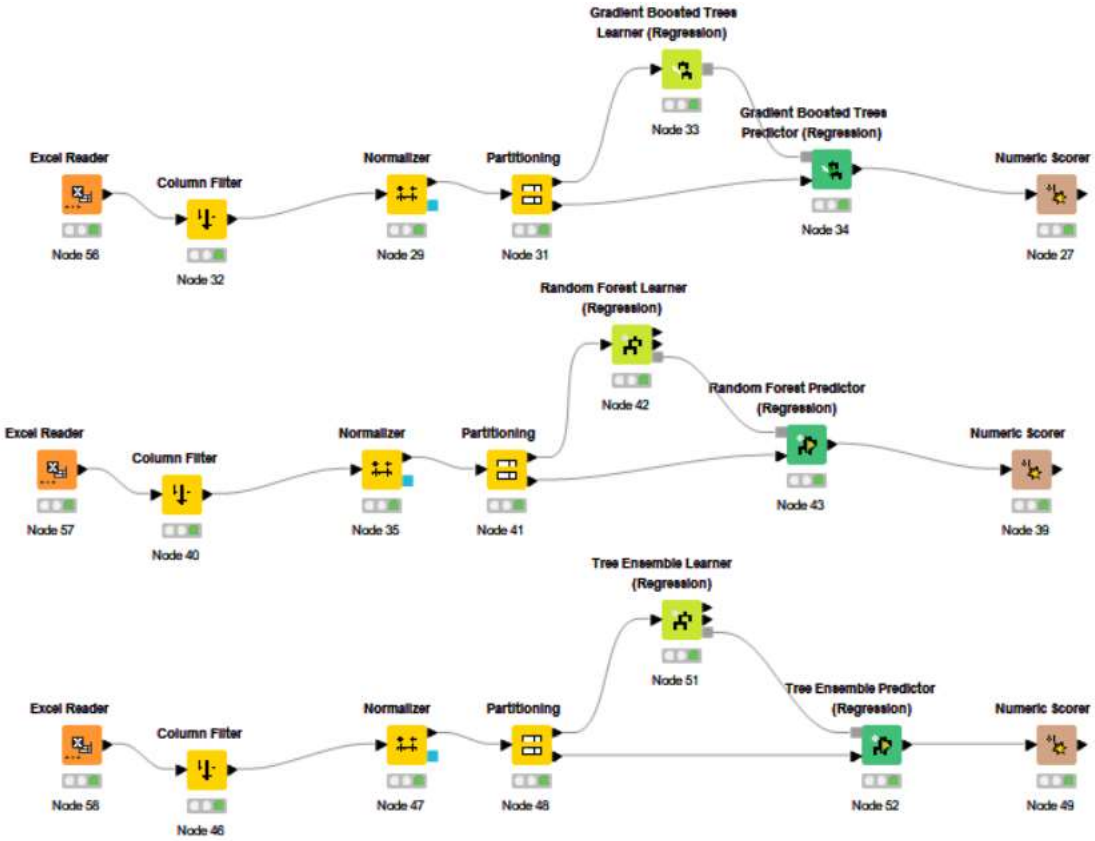

Figure 35. Workflow of the Machine Learning Algorithms with KNIME.