Original Scientific Article Journal of Decision Analytics and Intelligent Computing
Vol. 5 issue 1, (2025) 87-98 https://doi.org/10.31181/jdaic10014062025p

PCA - Enhanced regression approach for predicting internet use based on formal education

Ivana Petkovski[1,*] and Petar Vranić[1]

[1] Computer sciences, Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

* Correspondence: ivana993@turing.mi.sanu.ac.rs

Received 13 March 2025
Accepted for publication 5 June 2025
Published 14 June 2025

Abstract

The regular use of digital devices and Internet applications in modern society is significantly influenced by the user's educational attainment. In educational institutions, particularly at the primary and secondary levels, digital technologies (DT) are recognized as essential to the teaching process. Current educational programs involve active use of DT in class and thus improve digital skills necessary for active engagement in the digital world. A hybrid machine learning model was built to analyze the impact of varying educational degrees on Internet usage and innovation investments across the European Union (EU) population. Principal component analysis (PCA) was employed for detecting the primary indicators of education and innovation, resulting in the reduction of the initial variables to two factors, one including four indicators and the other containing two indicators. Linear regression (LR) was applied on the PCA factor loadings derived from the primary factor, which exhibited a statistically significant relationship with the percentage of Internet usage (r = 0.767). The results demonstrate that formal education, supported by investments in innovations within the education system, are essential preconditions for the continued development of the digital society.

Keywords: principal component analysis (PCA), correlation, linear regression, formal education, internet users.

1. Introduction

Because of the fast nomination of the Internet and the accelerated process of shifting to the digital-society phase, modern human life has been affected. This major change has occurred through different fields, affecting the routines of usual life; hence, some necessary skills need to be imparted now for potential navigation through digital technology. Educational institutions assist in empowering more people to get on the Internet and acquire digital skills (Petkovski and Vranić, 2024). Digital literacy means to be able to search, retrieve, store, manage, share, understand, analyze, evaluate and create information through the use of digital tools. This capacity can help a person to transform information into knowledge; it is an approach to teaching people about ICT principles and how to use digital technologies effectively (International Telecommunication Union, 2010; UNESCO Institute for Statistics, 2018). This proficiency not only aids individuals in securing good jobs and starting their own businesses but also facilitates their integration into society.

The OECD publication on "Skills Outlook 2019: Thriving in a Digital World" indicates the increasing need for digital skills across industries as AI, automation and data analytics change their setup (OECD, 2019). The report mentioned that digital advances dictate new skill requirements while also influencing educational systems. The World Economic Forum report, "The Future of Jobs Report 2020," similarly projects that by 2025, around 85 million jobs will be rendered obsolete by machines (World Economic Forum, 2020). Simultaneously, these digital transitions will foster the creation of 97 million new jobs. The digital business transformations are driving economic and social growth. They reduce transaction costs and influence workers' knowledge and skills, inflation, wages and employment markets, thus output. All these tend to speed up the structural and technical changes germinating in the world economy. McKinsey and Company findings imply that by increasing digital skills, there is a potential to add as much as USD 2 trillion to global GDP by the year 2030, with countries demonstrating higher levels of digital literacy expected to experience relatively greater growth (Manyika et al., 2017). According to RedSeer Consulting, the online higher education and lifelong learning market in India was poised to grow to US$5 billion by 2025, owing to government initiatives for developing online programs, improving digital infrastructure across the nation and an increased need for student upskilling (Kumar, 2024).

In spite of the availability of Internet access and digital technologies, the greatest digital divide happens among regions and nations differing from one another. The term digital divide, often also known as the digital gap, involves the ability to actually access and use information and communication technology within a country or across nations (Jamil, 2021). The existence of a digital divide within a country prevents individuals from fully engaging in societal processes and acquiring vital information in an information-led era (Van Dijk, 2013). The digital divide, characterized by inequalities in technology and internet access, poses a major challenge when it comes to educational equity, calling for targeted interventions such as technology hubs in schools and public-private partnerships to achieve an inclusive digital learning environment (Afzal et al., 2023). The results presented in Tislenko (2010) present changes in digital inequality from 2017 to 2022 in the European Union, along with the role of EU supranational policies and funding in spatial disparities. Using DESI indicators, the findings show that, although the digital divide is decreasing slightly, financial support from the EU was not evenly distributed and was only partially effective, with more Western Europe and Southern Europe obtaining the benefits than the Eastern member states. Another study, drawing on data from the EIBIS Digital and Skills Survey, sheds light on the growing digital divide between firms in the EU and the US. It reveals that many small, older companies are still lagging behind in digital adoption, while others are making significant strides. Firms that remain non-digital tend to be less innovative, create fewer jobs, and see lower profit margins, which only serves to widen the corporate digital gap (Rückert et al., 2020). The digital divide among elderly patients poses a barrier to the widespread adoption of telemedicine, as their limited digital skills may hinder system use. However, leveraging the internet for chronic disease monitoring could help bridge this gap and promote greater social and technological inclusion for this age group. The digital divide among older patients creates a significant hurdle for the widespread use of telemedicine, as their limited digital skills can make it tough for them to navigate these systems. However, by utilizing the internet for monitoring chronic diseases, we could help close this gap and encourage more social and technological inclusion for this age group (Romano et al., 2015).

Education plays a crucial role in closing this divide by providing equitable access to digital resources and information, enabling everyone to take advantage of the digital shift. A study by the International Telecommunication Union indicates that different digital literacy programs implemented in rural areas have resulted in a 20% increase in income for those involved, highlighting the impact of such educational efforts (International Telecommunication Union, 2021). Similarly, several studies emphasize the role of digital literacy in improving income growth, employment opportunities and labor migration patterns, but also social belonging and reduction in income disparities (Wang et al., 2025).

The constantly changing digital landscape, with new tools and online platforms showing up one after another, is a centerpiece in the development of needing to constantly educate oneself and acquire new skills. Formal schooling is not the end of it; embracing lifelong learning programs just makes anyone employable in this age of digitalism. Evidence suggests that people who pursue continuous learning opportunities tend to lose their jobs less often and, hence, have better chances of career advancement (Cedefop, 2020). Extra non-government organizations and community-based forums in non-formal education have a place in developing digital literacy, in particular for those social groups that have a handicap in the digital economy, such as low-income families, rural communities and the elderly. The research report "The Role of Community-Based Programs in Closing the Digital Divide" reveals that the community-based digital literacy programs not only strengthen the tech skills of the participants but also build their confidence to use technology, which further encourages them to go online for additional learning (Everyone on, 2022).

The quality of education is an undoubtable key factor for the expansion of the Internet and digitalization. However, among the plethora of indicators that describe the education system it is important to understand which ones have a dominant influence when it comes to the use of the Internet. Thus, this paper's main goal is to model this relation in order to offer a better understanding of which segments of education have a significant influence on the use of the Internet.

2. Literature review

In the broader sense, some studies have tried to model the relations between education, students, Internet use, digitalization and their impact on employability. For example, in an empirical study, Van Deursen and Van Dijk (2016) modelled relationships between traditional literacy (reading, writing, and understanding text), medium-related Internet skills (operational and formal skills), content-related Internet use (information and strategic skills), and types of Internet use (information, career, or entertainment use). The study used structural equation modeling to test the developed conceptual model and principal component analysis (PCA) with varimax rotation to identify two usage clusters. The findings show that traditional literacy is a precondition for the employment of Internet skills, i.e., traditional literacy directly affects formal and information Internet skills, while indirect effects operate on strategic Internet skills. Another study presented in Komarudin et al. (2024) intended to determine the impact of the RMS approach in teaching on the digital and mathematical literacy of students. The study employed experimental research. Data analysis of statistical distribution includes percentage, mean, standard deviation, correlation and analysis of variance (ANOVA). With the aim of exploring the correlations among variables in the study, the authors used structural equation modeling (SEM). The results show that understanding mathematical concepts can help students appreciate and put to good use digital technologies; that is, science education increases digital literacy. The authors used partial least squares structural equation modeling (PLS-SEM) to develop a value-based online learning model that predicts learners' responses to internet entrepreneurship education (Tseng et al., 2023). The results show that instructional assistance, teacher enthusiasm and teacher preparation-parameters considered in evaluating the perceived value of online learning to the acquisition of internet entrepreneurship knowledge and skills-contribute to growth-oriented bias for reuse intention. In Segbenya et al. (2023), 21st-century employability skills with respect to AI usage and its antecedents were modeled among postgraduate students in Ghana, using PLS-SEM techniques for analyzing quantitative data and thematic pattern matching for qualitative data. For checking the item loadings of those five selected variables, the authors went with CFA methods. The PLS-SEM path modeling states that both AI usage and AI challenges predict significantly how postgraduate students acquire employability skills of the 21st century. In Duong et al. (2023), an attempt was made to understand the usage of ChatGPT by higher education students. With a modification made to the technology acceptance model, the study findings, which used stratified random sampling to select 1389 higher education students from 11 universities in Vietnam, showed that effort expectancy directly affected students'

actual use of ChatGPT. It also had an indirect serial effect through performance expectancy and intention to use ChatGPT on the actual use of ChatGPT by students. There is a three-step analytical data process to evaluate the reliability and validity of the constructs being studied and to assess the proposed hypotheses. The first step was considering CFA and Cronbach's alpha for reliability and validity assessment of constructs. Next, to examine the predicted relationships in the developed model, the authors applied multiple linear regression (LR) analysis. Finally, testing mediation coefficients was performed by applying models 4 and 6 of the PROCESS macro with 5000 bias-corrected bootstrap samples.

A recent meta-analysis published in the Humanities and Social Sciences Communications examined 35 independent effect sizes to check the association between digital literacy and academic performance. The study revealed a stronger positive contribution from increased digital literacy to improved academic outcomes. Other moderators investigated in this study would be grade level, subject or gender that can modify this relationship (Li et al., 2025).

The Milken Institute Report analyzed the influence of Internet use on the educational attainment parameter. The study turns out to provide proof that the number of hours spent on the Internet per week correlates positively with educational outcome measures, thus implying that increased Internet use serves to enhance learning and academic achievement (Lee, 2017).

The rural adolescent study conducted in China investigates the causal relationship between Internet use and learning outcomes. With data taken from the China Family Panel Studies (CFPS), they report a positive effect of increased Internet access on students' academic performance (Li et al., 2021).

A study published in the Journal of Student Research investigated the relationship between broadband speed and access and student performance. Using the data for Internet access and standardized test scores in New York State, they found broadband access to have a positive effect on the test scores, thus putting in the limelight the importance of Internet infrastructure in education (Shah and Jimenez-Duran, 2023).

Another recent study in 2023 constructed an assessment model of digital literacy for vocational students with a focus on PCA. The analysis considers social responsibility and vocational competency as the base components that contribute the most to the construct of digital literacy (Peng et al., 2023).

Having an understanding of the relationship between education and the use of the Internet can help policymakers in developing strategies towards a more digitally inclusive society.

3. Materials and methods

The research framework included several independent variables cited in the European Commission (2022) dataset on education for the year 2022. The digital development as a dependent variable was regularly reported as a share of individuals using the Internet. The information about the number of Internet users was sourced from the official dataset available at the World Data Bank (2024). To ensure comparability and a precise machine learning model the scale was standardized. The empirical example was established on the EU 27 countries such as Belgium, Bulgaria, the Czech Republic, Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Croatia, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Malta, the Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland and Sweden. Table 1 describes an overview of variables along with their labels and descriptive statistics. Gross domestic expenditure R&D in higher education was expressed in euro (€) per individual for simplifying interpretability in this phase. The rest of the variables adopted a uniform reference unit expressed in percentages.

Table 1. Variables information

Indicator name	Label	Min	Max	Mean	SD
Individuals - internet use	Int_use	80.39	99.35	91.5593	4.9756
Gross domestic expenditure R&D in higher education	x_1	5.92	659.06	171.1022	151.0650
Pupils enrolled in primary education	x_2	4.00	11.00	5.6296	1.6207
Pupils enrolled in lower secondary education	x_3	3.00	6.00	3.9259	0.8286
Pupils enrolled in upper secondary education	x_4	2.18	6.22	4.1400	0.9377
Students enrolled in bachelor education	x_5	0.47	7.04	2.5852	1.1548
Students enrolled in master education	x_6	0.43	2.84	1.1933	0.4298
Students enrolled in doctoral education	x_7	0.00	0.35	0.1596	0.0759
Formal education and training of population ages 18 to 64	x_8	7.50	27.30	13.3481	4.8188
Non -formal education and training of population ages 18 to 65	x_9	15.50	63.60	42.4481	13.0459
Graduate pupils tertiary education	x_{10}	0.00	1.06	0.5285	0.2207
Graduates students bachelor education	x_{11}	0.11	0.91	0.4752	0.1633
Graduates students master education	x_{12}	0.14	1.03	0.3474	0.1717

The Internet use variable with its minimum value of 80.39% indicates a high level of digital inclusion. The presence of heterogeneity in the use of the Internet (approximately 20%) among the surveyed countries allows the analysis of influential independent variables. Investment in R&D in the field of education varies significantly across countries depending on the size of GDP and individual investment policy. This is evidenced by the high level of standard deviation (151.0650), which arises as a disparity between extreme values. A minor standard deviation value is observed throughout all levels of primary and secondary schooling. Data on tertiary education, which include bachelor's, master's and doctoral degrees, indicate a sharp drop in student enrollment at these levels of study, from the lowest to the highest level. A significantly higher value for informal forms of learning is observed when looking at the data for adults' formal and informal education. This finding is further confirmed by the fact that a smaller number of the population pursue a higher level of study after completing high school. The informal form of education is more accessible and flexible to adults than formal schooling, leading the majority to opt for this alternative. In addition, an important role is played by companies that participate in the training and education of their employees, which explains the trend of increasing informal education.

The proposed methodological framework is based on the application of principal component analysis (PCA) and linear regression (LR) to develop a predictive model of Internet use. While generally categorized as statistical methods, both techniques are also classified in the broader category of machine learning models. PCA represents an unsupervised ML learning model, while LR is a supervised ML learning model (Sarker, 2021). Prior to implementing PCA it is necessary to check the adequacy of the sample through the Kaiser-Meier-Olkin (KMO) test

and Bartlett's test of sphericity (Lakhotia et al., 2019). PCA identifies latent structures and interdependencies within data. Basically, this method is used to reduce the number of initial variables by aggregating indicators that share the highest degrees of correlation into a common factor (Abdi and Williams, 2010). The first extracted factor explains the highest degree of variance among the data, while each subsequent factor explains the remaining maximum part of the variance. The number of factors is selected based on the eigenvalue which must exceed the threshold of 1 (Greenacre et al., 2022). Several rotation methods are applied including varimax, quartimax, equamax, promax and oblimin (Wold et al., 1987). Rotation methods allow each variable to be dominantly associated with one factor, which clearly indicates the factor structure. The main goal of the PCA method is to form a smaller number of factors that explain the variance in the data without losing valuable information (Abdi and Williams, 2010).

Linear regression is a statistical procedure that although it is rooted in traditional statistics, also finds application in the field of machine learning (Sarker, 2021). LR evaluates the influence between the independent variable and the dependent variable (Weisberg, 2005). LR can be represented in its simplest form with one independent variable (Weisberg, 2005). Also, the LR model can be updated to include a larger number of independent variables. Then the LR is transformed into a form of multiple linear regression. In this instance, the structure of a LR equation with a single predictor is represented by the following formula (Bingham and Fly, 2010):

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \quad (1)$$

where,

- y is the dependent variable or the variable whose value is predicted;
- x_1 is the independent variable, or the variable whose value affects the predictive value of the dependent variable;
- β_0 is the intercept or the value of the dependent variable when the values of the independent variable are equal to zero;
- β_1 is the coefficient of the regression equation that quantifies the influence of the independent variable on the dependent variable and,
- ε is the residual or the model error that appears as the difference measured between the actual and predicted values of the dependent variable.

4. Illustrative example and results

To adequately prepare the data set for the application of PCA analysis to extract the most important latent factors, all data passed a standardization process. The observed extreme difference between the minimum and maximum values of the variable x_1 (min = 5.92; max = 659.06) demanded the application of the natural logarithm prior the standardization step. This transformation was implemented to mitigate the potentially dominant effect of the dependent variable in relation to other independent variables unlike the previous research (Petkovski and Vranić, 2024) where authors implemented only the standardization step. The stability and suitability of the data for the application of PCA were confirmed through the KMO test and the p-value of Bartlett's test of sphericity, whose reference and derived values are shown in table 2.

Table 2. Testing data adequacy for the PCA

Test name	Indicator	Reference value	Calculated value
Kaiser-Meyer-Olkin test	KMO index	> 0.600	0.716
Bartlett’s test of sphericity	p-value	< 0.050	0.000

PCA analysis extracted two factors according to which the first factor accounts for 40.656% of the total data variability, while the second factor accounts for 25.760% of the total variability. The cumulative percentage of

explaining the variable using these two factors is set to 66.416%, which indicates the retention of a significant part of the total information from the original data. Table 3 provides the values of factor loadings that surpass the threshold of 0.400 obtained across three iterations using the Varimax rotation method.

Table 3. Rotated component matrix

Extracted parameter	Factor loading	
	Factor 1	Factor 2
Formal education and training of population ages 18 to 64	0.906	
Pupils enrolled in upper secondary education	0.774	
Gross domestic expenditure R&D in higher education	0.734	
Pupils enrolled in primary education	0.673	
Students enrolled in bachelor education		0.810
Students enrolled in doctoral education		0.817

In order to determine the direction and strength of the relationship between regressions factor scores of factor one and factor two with the dependent variable, Pearson's correlation was calculated. Pearson's correlation coefficients are shown in Table 4. The results demonstrate a high positive and statistically significant correlation between the regression factor score for factor one and internet users (r = 0.767; p = 0.000). The results for factor two revealed a statistically non-significant relationship between factor 2 and the dependent variable (r = 0.000; p = 1.000). Factor two was eliminated from the LR analysis due to its insignificant contribution to the linear model for predicting the number of Internet users.

Table 4. Pearson's correlation matrix

		Int_use	$REGR_1$	$REGR_2$
Pearson Correlation	Int_use	1		
	$REGR_1$	0.767	1	
	$REGR_2$	0.000	0.000	1
Sig. (1-tailed)	Int_use	1		
	$REGR_1$	0.000	1	
	$REGR_2$	1.000	1.000	1

The presence of a single factor eliminates the trouble of multicollinearity between variables. However, due to one factor dimensionality, the coefficient of determination value equals to 76.7% comparable to Pearson's correlation coefficient. The stability and quality of the LR model were statistically confirmed with an f-statistic of 35.770 and a statistical significance of 0.000. Additional results of statistical tests are reported in Table 5.

Table 5. Model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0.767	0.589	0.572	0.6541	1.556

The LR equation described in expression (2) consists of the beta regression coefficient (0.767) and the independent variable represented by the standardized value of the regression score derived from factor one ($REGR_1$). The dependent variable is expressed through the standardized value of the number of Internet users. The constant in the equation is omitted in this instance because the value obtained is so low that it can be considered practically zero or numerical zero. Simultaneously the analysis revealed a lack of statistical significance (p = 1.000). Data standardization in the LR model dictated the structure of the LR equation as the mean values are centered

around zero and the standard deviation is equal to one. In that scenario, LR does not significantly shift the value of the dependent variable because the independent predictors already oscillate around zero, so the value of the constant is also approximately zero. In addition, the equation shows that a one standard deviation rise in factor one results in an increase in the value of the standard deviation in the dependent variable of the number of Internet users by 0.767.

$$Int\_use = 0.767 \times REGR_1 \quad (2)$$

Figure 1 uses a radar diagram to illustrate the results of applying the PCA-LR model. Both actual and predicted values are presented in their standardized form. Visual analysis, confirmed numerically by applying the formula for the absolute difference at the country level, indicates that the model succeeds in recognizing the trend of the number of Internet users using the PCA-LR model. The model is well fitted for countries like Finland, Slovenia, Hungary and France where the absolute difference ranges from 0 to 0.09. However, local differences are also evident. In Bulgaria, Portugal, Croatia and Luxembourg this difference exceeds 1, showing that the model predictions of the number of Internet users in these countries are less accurate. This is due to unequal access to financial resources that facilitate digital development, as well as disparities in national educational systems and demographic age distributions.

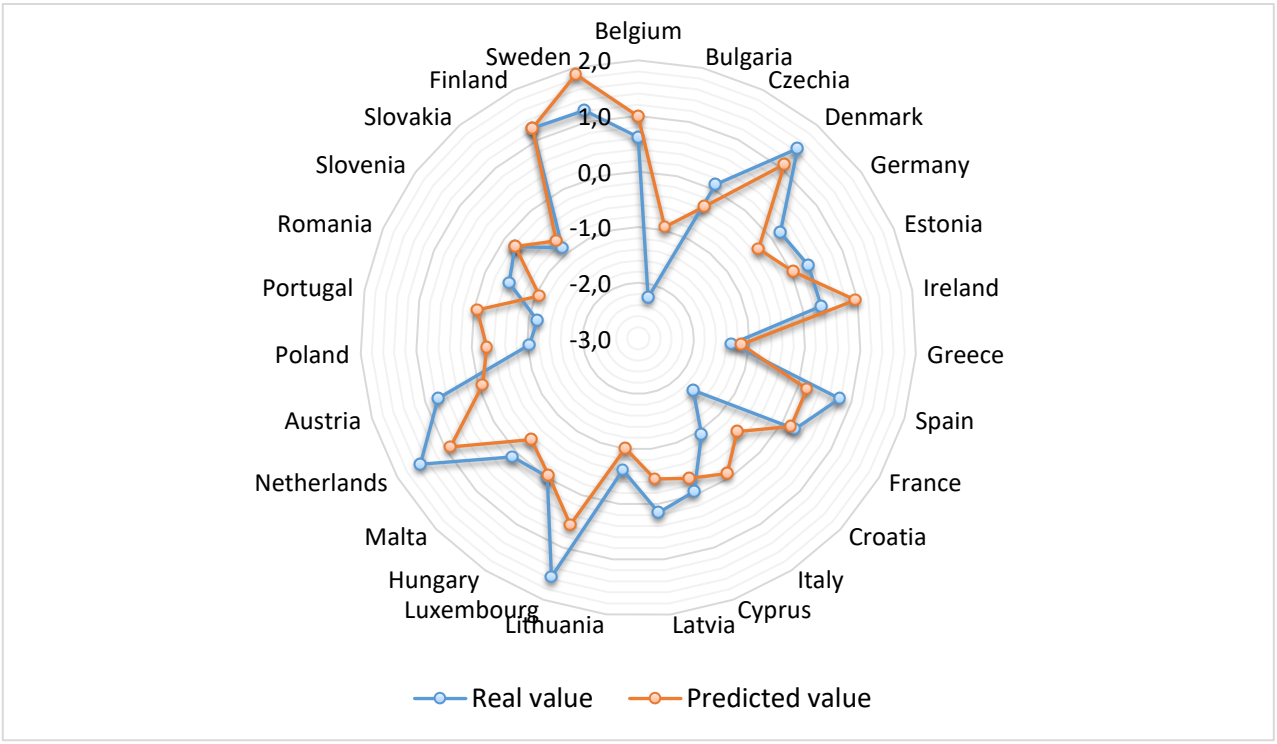

Figure 1. Real and predicted standardized values of internet users in EU27

Evaluating the data of individual countries offers analyses of trends at the regional and local levels. The maximum value of factor loadings in the PCA-ML model is the percentage of the population that participated in some form of formal education and training in 2022. According to this parameter, Scandinavian countries such as Finland and Sweden stand out, followed by Ireland, Denmark and Spain (between 17% and 30% of the adult population). Conversely, less than 10% of the population actively participates in formal education and training in Romania, Greece, Bulgaria, Italy, the Czech Republic, Slovenia, Lithuania and Slovakia. The enrollment rates of children in primary and upper secondary school are diverse and point to geographical differences. Northern and western EU countries such as Sweden, Finland, Belgium and Denmark are leading by far. An intriguing recorded

fact is that countries like Germany and Austria report a very low rate of students at the initial levels of education, which speaks of their demographic structure and the reduced share of the youngest population. The problem of depopulation is highlighted by countries such as Bulgaria and Hungary. So far, several researches have confirmed the effects of formal education on technological development and internet use (Van Deursen and Van Dijk, 2016; Komarudin et al., 2024; Tseng et al., 2023; Li et al., 2025).

Considering investment in innovations, i.e., R&D in educational institutions, Denmark and Sweden take precedence, followed by Austria, Finland, Luxembourg, Belgium and Germany. The lowest investments in R&D in education are reported by southern and central EU countries such as Bulgaria, Romania, Slovakia, Hungary, Croatia and Lithuania. Differences in investment in R&D are so large as to induce disruptions in Internet use between countries and regions. Higher investments lead to faster technological development as supported by Aghion et al. (2019), Valero and Van Reenen (2019), Barro and Lee (2013) and Bonaccorsi and Daraio (2007). From a global perspective, there is a division between EU members in the north and west of the EU, which have developed and financially supported education systems, compared to the south and east of the EU, which are characterized by smaller investments in innovation and diminished participation of the adult population in formal education and training. Specific cases from countries like Luxembourg indicate that in these countries the rate of Internet use depends on other factors that may be more dominant than the influence of education itself.

5. Managerial implication and conclusions

In the paper, a machine learning model was developed to predict the use of the Internet based on educational criteria. The hybrid PCA-LR model enables the formation of latent factors that group related educational factors. Subsequently, based on the factor loadings obtained through PCA, a linear regression equation is generated. The PCA analysis identified two dominant factors of education, which were further processed using correlation analysis. The results showed a statistically significant and strong positive influence in the ratio of factor loadings of the first factor and the number of Internet users, which was adopted as a single predictor in LR. This factor represents a synthesis of participation in formal education and training, lower education levels of education and investment in R&D in the education sector. LR indicates that over 76% of the variation in Internet use can be explained by the latent factor defined in this scope. The obtained result provides excellent performance considering that it covers only the education framework. This model offers the possibility to apply machine algorithms in the planning of educational resources, predicting the need for teaching staff or managing the process of digital exclusion. At the same time, it offers a methodological framework for measuring the success of educational strategies aimed at fostering digitalization.

The regression model indicated deviations in the prediction in local cases, indicating differences in education systems across the EU. In accordance with the research, practical implications are proposed that can achieve balance between education and digital development.

Formal education of children and adults is a basic condition for technological development of the country. Accordingly, it is necessary to ensure greater budgetary allocations for education in those countries and areas that are exposed to a lack of financial capacity for the operation of primary and secondary schools, as well as higher education institutions. Such an approach enables a larger percentage of the population to acquire the necessary digital skills and actively use the Internet. In addition, countries lagging behind in the digitalization process should develop or modify existing strategies for digitalization in education.

Investing in research and development at the level of secondary education offers opportunities for further growth in digitalization through increased use of the Internet. However, not all member states are prepared to allocate the same financial resources for these activities. Accordingly, it is recommended to establish EU funds dedicated to southern and central countries as financial support for the process of digitalization of educational systems. By applying for funding in programs such as Horizon or Erasmus, individual educational institutions could

secure funding for equipping classrooms with digital devices, improving digital infrastructure or acquiring digital skills for professional staff, which can often be an obstacle to digital progress.

It is necessary to adjust the fiscal policy to introduce tax incentives for those employers who offer the possibility of formal education to their employees and apply the concept of lifelong learning. This policy would emphasize increasing the share of adults who are involved in some form of formal education.

Socially sensitive categories require complimentary or subsidized access to education and Internet resources to equally participate in the digital progress of society. It is necessary to offer programs of financial support and incentives for the purchase of digital devices and equipment to socioeconomically disadvantaged populations. This approach mitigates the risk of digital exclusion.

A limitation of the study is reflected in the results that indicated deviations in predicting Internet use for individual countries. Future research would focus on testing regression models by introducing different socioeconomic variables. Another limitation is reflected in the size of the data set on which the regression models were formed.

Acknowledgements

This work was supported by the Serbian Ministry of Science, Technological Development and Innovation through the Mathematical Institute of the Serbian Academy of Sciences and Arts.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), 433-459.

Afzal, A., Khan, S., Daud, S., Ahmad, Z., & Butt, A. (2023). Addressing the Digital Divide: Access and Use of Technology in Education. Journal of Social Sciences Review, 3(2), 883-895.

Aghion, P., Boustan, L. P., Hoxby, C. M., & Vandenbussche, J. (2019). The contribution of higher education to innovation and productivity: Evidence from U.S. states. Journal of Economic Growth, 24(1), 35-82.

Barro, R. J., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. Journal of Development Economics, 104, 184-198.

Bingham, N. H., & Fry, J. M. (2010). Regression: Linear models in statistics. Springer Undergraduate Mathematics Series. London: Springer-Verlag.

Bonaccorsi, A., & Daraio, C. (2007). Universities and strategic knowledge creation: Specialization and performance in Europe. Cheltenham: Edward Elgar Publishing Limited.

Cedefop (2020). Empowering adults through upskilling and reskilling pathways: Vol. 1: adult population with potential for upskilling and reskilling. Luxembourg: Publications Office. Cedefop reference series, No 112.

Duong, C. D., Vu, T. N., & Ngo, T. V. N. (2023). Applying a modified technology acceptance model to explain higher education students' usage of ChatGPT: A serial multiple mediation model with knowledge sharing as a moderator. The International Journal of Management Education, 21(3), 100883.

European Commission. (2022). Indicator database. https://ec.europa.eu/eurostat/data/database, Accessed 15 July 2024.

Everyone On. (2022). The digital skill and trust. https://static1.squarespace.com/static/5aa8af1fc3c16a54bcbb0415/t/61fc71248a56247e899c2a20/1643933997111/EveryoneOn_Report_2_DigitalSkills_and_Trust.pdf, Accessed 10 May 2025.

Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A. & Tuzhilina, E. (2022). Principal component analysis. Nature Reviews Methods Primers, 2(1), 100.

International Telecommunication Union. (2010). World Telecommunication/ICT Development Report 10: Monitoring the WSIS Targets. https://uis.unesco.org/sites/default/files/documents/monitoring-the-wsis-targets-a-mid-term-review-world-telecommunication-ict-development-report-2010-en.pdf, Accessed 10 May 2025.

International Telecommunication Union. (2021). Digital skills insights. https://academy.itu.int/sites/default/files/media2/file/21-00668_Digital-Skill-Insight-210831_CSD%20Edits%206_Accessible-HD.pdf, Accessed 10 May 2025.

Jamil S. 2021. From digital divide to digital inclusion: Challenges for wide-ranging digitalization in Pakistan. Telecommunications Policy, 45(8), 102206.

Komarudin, K., Suherman, S., & Vidákovich, T. (2024). The RMS teaching model with brainstorming technique and student digital literacy as predictors of mathematical literacy. Heliyon, 10(13), e33877.

Kumar, V. (2024). Digitalization Higher Education in India. Journal of dvanced esearch in ducation, 3(1), 14-17.

Lakhotia, R., Nagesh, C. K., & Madgula, K. (2019). Identifying Missing Component in the Bechdel Test Using Principal Component Analysis Method. arXiv preprint arXiv: 1907.03702.

Lee, J. (2017). Internet Usage Effect on Educational Attainment: Evidence of Benefits (pp. 1-17). California: Milken Institute.

Li, F., Cheng, L., Wang, X., Shen, L., Ma, Y., & Islam, A. Y. M. (2025). The causal relationship between digital literacy and students' academic achievement: a meta-analysis. Humanities and Social Sciences Communications, 12(1), 1-12.

Li, L., Ma, Y., Friesen, D., Zhang, Z., Jin, S., & Rozelle, S. (2021). The impact of Internet use on adolescent learning outcomes: evidence from rural China. China Agricultural Economic Review, 13(3), 569-592.

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R., & Sanghvi, S. (2017). Jobs lost, jobs gained: Workforce transitions in a time of automation. McKinsey Global Institute. https://www.mckinsey.com/~/media/McKinsey/Industries/Public%20and%20Social%20Sector/Our%20Insights/What%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/MGI-Jobs-Lost-Jobs-Gained-Executive-summary-December-6-2017.pdf , Accessed 01 May 2025.

OECD. (2019). OECD Skills Outlook 2019 Thriving in a Digital World. Paris: OECD Publishing.

Peng, F., Guo, M., Zheng, C., Wang, S., Wang, X., & Xu, M. (2023). An Assessment Model of Digital Literacy for the Students in Vocational Education Based on Principal Component Analysis in Machine Learning. Proceedings of the 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (pp. 1382-1386). Chongqing: IEEE.

Petkovski, I., & Vranić, P. (2024). Modeling Education and Internet Usage: PCA and Linear Regression Approaches. Proceedings of the 6h Virtual International Conference Path to a Knowledge Society-Managing Risks and Innovation (PaKSoM) (pp. 71-77). Belgrade: Mathematical Institute SANU.

Romano, M. F., Sardella, M. V., Alboni, F., Russo, L., Mariotti, R., Nicastro, I., Barletta, V., & Di Bello, V. (2015). Is the digital divide an obstacle to e-health? An analysis of the situation in Europe and in Italy. Telemedicine and e-Health, 21(1), 24-35.

Rückert, D., Veugelers, R., & Weiss, C. (2020). The growing digital divide in Europe and the United States (No. 2020/07). EIB Working Papers.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), 160.

Segbenya, M., Bervell, B., Frimpong-Manso, E., Otoo, I. C., Andzie, T. A., & Achina, S. (2023). Artificial intelligence in higher education: Modelling the antecedents of artificial intelligence usage and effects on 21st century

employability skills among postgraduate students in Ghana. Computers and Education: Artificial Intelligence, 5, 100188.

Shah, A., & Jimenez-Duran, R. (2023). Broadband Access and Standardized Test Scores: A Causal Analysis. Journal of Student Research, 12(4), 1-10.

Tislenko, M. I. (2024). Digital divide in the European Union: assessing spatial disparities and neighborhood effects. Journal of the Geographical Institute "Jovan Cvijić" SASA, 74(2), 181–194.

Tseng, T. H., Wu, T. Y., Lian, Y. H., & Zhuang, B. K. (2023). Developing a value-based online learning model to predict learners' reactions to internet entrepreneurship education: The moderating role of platform type. The International Journal of Management Education, 21(3), 100867.

UNESCO Institute for Statistics. (2018). A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2. https://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf, Accessed 10 May 2025.

Valero, A., & Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. Economics of Education Review, 68, 53-67.

van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2016). Modeling traditional literacy, Internet skills and Internet usage: An empirical study. Interacting with computers, 28(1), 13-26.

van Dijk J. A. G. M. (2013). A theory of the digital divide. In M. Ragnedda, & G.W. Muschert (Eds.), The digital divide: the internet and social inequality in international perspective (pp. 29-51). (Routledge advances in sociology; Vol. 73, No. 73). Routledge.

Wang, Y., Xu, G., Cao, J., Chen, Y., & Wu, J. (2025). Does digital literacy affect farmers' adoption of agricultural social services? An empirical study based on China Land Economic Survey data. PLoS One, 20(4), e0320318.

Weisberg, S. (2005). Applied linear regression (Vol. 528). New Jersey: John Wiley & Sons.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52.

World Data Bank. (2024). Indicator database. https://data.worldbank.org/, Accessed 15 July 2024.

World Economic Forum. (2020). The Future of Jobs Report 2020. https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf, Accessed 10 May 2025.