RIGA TECHNICAL UNIVERSITY

Ieva Vītoliņa

RESEARCH OF THE E-INCLUSION PROCESSES AND TECHNOLOGICAL SOLUTIONS

Summary of the Doctoral Thesis

RTU Press
Riga 2021

RIGA TECHNICAL UNIVERSITY
Faculty of E-Learning Technologies and Humanities
Distance Education Study Centre

Ieva Vītoliņa
Doctoral Student of the Study Programme "E-Learning Technology and Management"

RESEARCH OF THE E-INCLUSION PROCESSES AND TECHNOLOGICAL SOLUTIONS

Summary of the Doctoral Thesis

Scientific supervisor
Dr. Phys.
ATIS KAPENIEKS

RTU Press
Riga 2021

Vītoliņa, I. Research of the E-Inclusion Processes and Technological Solutions. Summary of the Doctoral Thesis. – Riga RTU Press, 2021. – 59 p.

Published in accordance with the decision of the Promotion Council “RTU P-21” of 21 September 2021, Minutes No. 1.

NATIONAL DEVELOPMENT PLAN 2020

EUROPEAN UNION
European Social Fund

INVESTING IN YOUR FUTURE

https://doi.org/10.7250/9789934226984
ISBN 978-9934-22-698-4 (pdf)

DOCTORAL THESIS PROPOSED TO RIGA TECHNICAL UNIVERSITY FOR THE PROMOTION TO THE SCIENTIFIC DEGREE OF DOCTOR OF SCIENCE

To be granted the scientific degree of Doctor of Science (Ph. D.), the present Doctoral Thesis has been submitted for the defence at the open meeting of RTU Promotion Council on 16 December 2021 at 11.00 a.m. at the Faculty of E-Learning Technologies and Humanities of Riga Technical University, 1 Kronvalda Boulevard, Room 200.

OFFICIAL REVIEWERS

Dr. paed Sarma Cakula
Riga Technical University

Dr. comp. Māris Vītiņš
University of Latvia

PhD Peter Francis Chatterton
University of London, UK

DECLARATION OF ACADEMIC INTEGRITY

I hereby declare that the Doctoral Thesis submitted for the review to Riga Technical University for the promotion to the scientific degree of Doctor of Science (Ph. D.) is my own. I confirm that this Doctoral Thesis had not been submitted to any other university for the promotion to a scientific degree.

Ieva Vītoliņa ……………………………. (signature)
Date: ………………………

The Doctoral Thesis has been written in Latvian. It consists of Introduction; 4 sections; Conclusion; 52 figures; 35 tables; 10 appendices; the total number of pages is 184, not including appendices. The Bibliography contains 387 titles.

Contents

GENERAL DESCRIPTION OF THE DOCTORAL THESIS

Topicality of the Research

More and more information, employment, household, education, and public administration services are available digitally. Information and communication technologies (ICT) have become an integral part of everyday life. E-inclusion focuses on the participation of all individuals and communities in all aspects of information society. E-inclusion policies aim to reduce the disparities in the use of ICT by different individuals and to promote the effective use of ICT by all individuals for education, personal development, and professional development, thus contributing to economic growth and full participation of individuals in information society (DiMaggio & Bonikowski 2008; FreshMinds, 2008; Johansson & Tjäder, 2013).

The strategic goals of the e-inclusion policy in the European Union were set in 2006, but they have not been achieved. The European Commission's Communication of 10 March 2021 entitled "2030 Digital Compass: The European way for the Digital Decade states that there is still a gap in society between those who can and do not benefit from digital technologies, so the European Commission's vision for 2030 is a digital society that leaves no one behind (European Commission, 2006; European Commission, 2021).

Statistics show that individuals do not take full advantage of the opportunities offered by technology. Data compiled by Eurostat (European Commission, 2017) show that in 2017, although 96.7 % of EU citizens have the skills to work online, they do not use the opportunities offered by technology. For example, only 67.6 % use Internet to search for information about goods and services, 51 % use Internet banking, 16.8 % use it to search for job offers, 7.35 % use it for on-line studies (e-learning courses).

The European Union has consistently recognized the importance of basic digital skills for all citizens and included them in both the Riga e-Inclusion Declaration and the Digital Agenda for Europe (European Commission, 2006; European Commission, 2010). Digital competence has been identified as one of the eight key competences that are essential for everyone (Council of Europe, 2006; Council of Europe, 2018). However, in 2019, only 56 % of adults in the European Union had basic digital skills (European Court of Auditors, 2021).

The predictions of researchers who believe that the differences between individuals' digital skills and their use will disappear over time and that there is no need to influence the e-inclusion process were not confirmed (ITU, 2006; Samuelson, 2003). On the contrary, several researchers indicate that differences between individuals in their ability to use technology are not diminishing but increasing (Haight, Quan-Haase, & Corbett, 2014). The urgency of promoting the e-inclusion process is determined by technological developments, as new technological opportunities are constantly emerging, so individuals must continuously develop their skills to use new technologies (European Commission, 2021; Yu et al., 2018). With the development and change of technology, the problem of digital skills shortage remains. Europe suffers from a growing shortage of professional ICT skills and digital skills (European Commission, 2012; Santos, Azevedo, & Pedro, 2013). The European Commission has set a target of having at least 80 % of all adults in basic digital skills by 2030 and employing 20 million professionals in the EU's ICT sector.

To promote the e-inclusion of individuals, European Union policy documents have been developed, and various studies on e-inclusion processes have been carried out. The results of several studies characterize e-inclusion, substantiate the need for e-inclusion, and provide recommendations for the development and implementation of the e-inclusion declaration (FreshMinds, 2008). Different target groups have been studied: people with disabilities, the elderly, minorities, people living in economically underdeveloped regions (Abad, 2014; Aerschot & Rodousakis, 2008). Recommendations are provided for public administration and private sector e-services to ensure that their quality meets the needs and usability of target groups (Achituv et al., 2008; Bélanger & Carter, 2009; European Commission, 2006). Recommendations promoting digital skills are given (DLHLEG, 2008). Recommendations for the implementation of e-learning are given (Casacuberta, 2007). The research focuses on how to ensure the availability of technologies (Rapaport, 2009).

Previous research on e-inclusion has focused on identifying differences in various socio-demographic, economic, and geographical indicators between individuals who use and individuals who do not use information and communication technologies (Drabowicz, 2014; Haight, Quan-Haase & Corbett 2014; Hidalgo et al., 2020). However, some studies indicate that not only socio-demographic and economic indicators are important, but also new factors characterizing e-inclusion need to be sought (Sanz & Turlea 2012). The number of such studies is insufficient, where correlations are sought between the e-included individual and the factors that characterize the individual, his / her behavior, e-inclusion process, digital skills acquisition processes.

Several studies suggest that it is necessary to look for relationships that characterize e-included individuals based on several factors simultaneously (De Haan, 2004). It is necessary to further study the processes of e-inclusion by identifying the factors that influence the process of e-inclusion so that individuals learn new technologies and use them meaningfully (Guillen-Gamez et al., 2020; Hatlevik et al., 2015). There is currently no comprehensive method that looks at the e-inclusion process from the point of view of the meaningful use of digital skills.

The results of existing research have not helped to achieve the goals of e-inclusion, both for those who lead the process at the administrative level or participate as professionals, and for individuals who need to be included. Social and demographic parameters alone cannot explain the differences in motivation, ICT availability, digital skills, and ICT use (Zillien & Hargittai, 2009). Although research has been carried out on e-inclusion processes, there is no consensus on how to promote the meaningful use of newly acquired digital skills. Existing e-inclusion studies are descriptive, they find that there are differences in the use of ICT for different groups, compared by one or more socio-demographic, economic, or other characteristics of individuals, and the studies describe these groups.

Most of the existing research is on e-inclusion risk groups, such as the elderly, immigrants, the disabled persons. But today, new, previously unidentified risk groups are emerging, such as young people and individuals who need to change occupation or who need to be able to use ICT in their professional activities (Csordás, 2020; Drabowicz, 2014; Sanz & Turlea, 2012). Educators have a key role to play in bridging the digital divide. Digital education requires them to be both digital skills experts to teach others and to continuously develop their own professional digital skills, as set out in the European Union's Digital Education Action Plan 2021–2027. (European Commission, 2020; LR Izglītības un zinātnes ministrija, 2021). The digital skills of educators in educational institutions are assessed as insufficient (Instefjord & Munthe, 2017; Jerrim & Sims, 2019). There is a small number of studies on digital inclusion of these groups.

The possibilities of data and learning analytics in promoting learning achievements are emphasized in the European Union Digital Education Action Plan for 2018–2020. This plan emphasizes the need for research in the field of artificial intelligence and learning analytics (European Commission, 2018). Although various cases of the use of learning analytics are available in the literature, a comprehensive approach to learning analytics to promote the digital skills acquisition process that would ensure the inclusion of the individual is still missing.

The Object of the Research

The object of the research is the process of knowledge creation and transfer in the information system predicting the e-inclusion of the individual.

The Subject of the Research

The subject of the research is the development of a model predicting the e-inclusion of the individual.

The Main Aim of the Doctoral Thesis

The research aims to develop a model that predicts an individual's e-inclusion in e-learning environment.

Research Questions

The main research question is: What technological solutions, using the factors characterizing the individual, allow to predict the degree of e-inclusion of the individual in e-learning environment in the digital skills acquisition course?

To determine the extent to which the linear regression model, using the predefined e-inclusion factors that characterize the knowledge flow between the instructor and the individual, can predict the degree of e-inclusion for VET (vocational education and training) teachers' learning digital skills under the guidance of an instructor in an e-learning environment, the following research questions have been defined:

1. What is the degree of correlation between the student's assessment of the instructor's willingness to share knowledge and the student's degree of e-inclusion?
2. What is the degree of correlation between the student's level of satisfaction with e-learning materials and the student's e-inclusion level?
3. What is the degree of correlation between the student's level of satisfaction with the e-learning environment and the student's e-inclusion level?
4. What is the degree of correlation between the student's desire to learn and the student's degree of e-inclusion?
5. What is the degree of correlation between the student's learning abilities and the student's degree of e-inclusion?

6. To what extent is it possible to predict the degree of student's e-inclusion based on the student's level of interest and ability to learn, the student's level of satisfaction with e-learning materials and e-learning environment and the instructor's willingness to share knowledge?

To find out (1) individual's e-inclusion factors that characterize the flow of knowledge between the instructor and the individual, (2) the differences between VET teachers who acquire digital skills under the guidance of an instructor in the e-learning environment and have different degrees of e-inclusion, the following research questions have been defined:

1. How does the students' assessment of the instructor's willingness to share knowledge differ for students who use the newly acquired digital skills after the end of the course and for students who do not use the newly acquired skills after the end of the course?
2. How does the level of student satisfaction with e-learning materials differ for students who use the newly acquired digital skills after the end of the course and for those students who do not use the newly acquired skills after the end of the course?
3. How does the level of student satisfaction with the e-learning environment differ for students who use the newly acquired digital skills after the end of the course and for those students who do not use the newly acquired skills after the end of the course?
4. How does students' willingness to learn differ for students who use the newly acquired digital skills after the end of the course and for those students who do not use the newly acquired skills after the end of the course?
5. How do students' learning abilities differ between students who use the newly acquired digital skills after the end of the course and those who do not use the newly acquired skills after the end of the course?

In order to determine to what extent it is possible to predict e-inclusion for VET teachers with classification-based methods and pre-determined factors, the following research questions have been defined:

1. To find out which classifiers generate e-inclusion predictive models with higher performance indicators.
2. To find out if there is a classifier with which the generated e-inclusion predictive models have the highest performance indicators for all training courses.
3. To find out how the performance indicators of the models differ for different types of data sets.

To determine to what extent, by combining linear regression, cluster analysis and classification methods, it is possible to train a model predicting e-inclusion with higher performance indicators compared to individual model indicators, the following research questions have been defined:

1. For which combinations of linear regression, cluster analysis and classification models, the performance indicators of the e-inclusion predictive model are higher?
2. What percentage of all VET teachers at-risk of digital exclusion who acquire digital skills can the model predict as belonging to the group at risk?
3. What percentage of the predicted VET teachers at-risk of digital exclusion actually belong to the group at risk?

To determine the performance of the model predicting e-inclusion and whether the model meets its objectives, the following research questions have been defined:

1. What percentage of all VET teachers at-risk of digital exclusion who acquire digital skills can the model predict as belonging to the group at risk?

2. What percentage of predicted VET teachers at-risk of digital exclusion actually belong to the risk group?

3. To what extent is it possible to predict, using model (without significantly lowering the performance indicators of the model), e-inclusion for individuals who acquire digital skills in courses that are different from the courses with which data the predictive model is trained?

To evaluate the drift of the model predicting e-inclusion, the following research questions have been defined:

1. By what percentage does the recall of the model and base models change?
2. By what percentage does the precision of the model and base models change?
3. By what percentage does the F measure of the model and the base models change?

The Tasks of the Doctoral Thesis

To achieve the goal of the Doctoral Thesis, the following tasks are set:

1. To develop an algorithmic model predicting e-inclusion of individuals.

A. To analyze the scientific literature and other sources related to the e-inclusion processes.

B. To analyze the scientific literature and other sources related to the predictive technologies and methods.

C. To create an algorithmic model predicting e-inclusion of individuals.

2. To create a technological model (prototype) predicting e-inclusion of individuals.

3. To evaluate the technological model predicting the e-inclusion of the teachers of the vocational education institutions.

Thesis Statements

1. The e-inclusion of an individual can be predicted using linear regression, cluster analysis, classifiers, and artificial intelligence methods.

2. The degree of e-inclusion of an individual can be predicted technologically by the following factors: the level of individual's satisfaction with the e-learning environment and e-learning materials used by the individual to acquire new digital skills; the individual's ability and interest in learning of the new digital skills; and the instructor's willingness to share knowledge. The intensity of the knowledge flow allows to predict technologically the degree of an individual’s e-inclusion.

3. The e-inclusion prediction model can be used for prediction of the VET teachers’ degree of e-inclusion during the acquisition of digital skills.

Methods of Research

The following theoretical methods have been used in the theoretical part of the Doctoral Thesis:

1. Research and analysis of scientific literature and European Union policy planning documents in the field of e-inclusion processes. Research and analysis of scientific literature on the factors characterizing students related to knowledge management. Research and analysis of scientific literature on the possibilities, tendencies, and technological solutions of

the analysis of factors characterizing students for the development of a model predicting the degree of e-inclusion.

2. 4EM (For Enterprise Modeling) method for structuring knowledge about e-inclusion, its goals, business rules, concepts, resources, participants, processes.

The following methods have been used to develop the practical part.

1. Data acquisition methods:

A. Surveys in e-learning environment Moodle for the determination of e-inclusion factors for the predictive model.

B. Surveys by e-mail and telephone surveys to determine the real e-inclusion level of students.

2. Data analysis methods:

A. Central tendency analysis to characterize the main features of the sample set.

B. Pearson correlation analysis to determine the relationship between e-inclusion degree and student characteristics.

C. Multivariate linear regression methods to determine the effect of e-inclusion risk factors on student e-inclusion degree.

D. Data mining methods. Cluster analysis – Expectation Maximization and kMeans algorithm to identify student clusters. Classification methods – LMT (Logistic Model Tree), LWL (Locally Weighted Learning), Naïve Bayes classifier, Simple Logistic regression, OneR classifier for predicting student e-inclusion.

3. The following methods have been used for the development and evaluation of the algorithmic and technological model predicting e-inclusion:

A. Cross-industry standard process for CRISP-DM data mining and CRISP-ML (Q) machine learning model development and quality assessment.

B. Data mining methods.

C. 10-fold cross-validation for model evaluation.

D. Confusion matrix, recall, precision, accuracy, and F-measure values to evaluate model performance.

Stages of Research

The study has been conducted in several stages.

2008–2010: The problem area of e-inclusion was studied, literature analysis was performed using the EKD (4EM) modeling approach. The theoretical design and methodology of the research have been developed.

2010–2013: E-inclusion prediction technologies and methods were studied; knowledge management theory was studied, characteristics of the vocational school teachers who acquired digital skills have been studied. Data analysis and linear regression modelling were performed characterizing the degree of e-inclusion of the VET teachers.

2013–2014: Development of technological model.

2014: The description of the problem area of e-inclusion was updated and a cluster analysis was performed for vocational school teachers who acquired digital skills.

2014–2016: Additional data on teachers who acquired digital skills were collected.

2019–2021: The problem area of e-inclusion was updated; used e-inclusion prediction technologies and methods were updated, the e-inclusion predictive algorithmic and technological model was developed and evaluated.

The Basis of the Research

The basis of the research is data on 767 vocational school teachers who acquired digital skills in the e-learning system Moodle environment from 2011 to 2012 and 160 vocational school teachers who did the same from 2014 to 2016.

Scientific Novelty, Practical Significance and Theoretical Significance of the Doctoral Thesis

The novelty of the Doctoral Thesis research

A model for predicting the inclusion of an individual and the risk factors influencing it has been developed. The model contains a new technology (algorithm) based on linear regressions, cluster analysis, and classification methods to determine the individual's e-inclusion risk and the factors influencing it.

The theoretical significance of the Doctoral Thesis research

1. The developed model of the e-inclusion prediction consists of new technology (algorithm) for determining an individual's e-inclusion degree in the context of digital skills acquisition.
2. The developed software prototype for predicting the e-inclusion of an individual provides an opportunity to analyze and evaluate the individual's risk factors for further research in this problem area.
3. Theoretical aspects of the e-inclusion prediction provide a theoretical basis for further research in this problem area.

The practical significance of the Doctoral Thesis research

1. The technological model of e-inclusion predicts the risk of an individual and determines the individual's risk factors in the context of acquiring digital skills.
2. The developed prototype of e-inclusion prediction software can be used in digital skills acquisition courses to identify factors hindering the meaningful use of newly acquired skills.
3. The developed prototype of e-inclusion prediction software can be used to evaluate and eliminate the identified risk factors by preparing new digital skills acquisition courses.
4. The developed prototype of e-inclusion prediction software can be used in digital skills courses as a support tool for the instructor to decide on the most appropriate learning approach for the individual.
5. The developed prototype of e-inclusion prediction software can be used in the development of new e-learning systems, a student behavior analysis tool, or learning analytics tools.

Approbation of Results of Research

Research results presented in conferences:

1. Vītoliņa I. Knowledge management model to facilitate e-inclusion. In 7th International JTEFS/BBCC Conference “Sustainable Development. Culture. Education”, May 5–8, 2009, Daugavpils University, Daugavpils, Latvia.

2. Vītoliņa I. Assessment of learning outcomes and collaborative efforts in computer supported environments. In RTU 50th International Scientific Conference, October 16, 2009, Riga, Latvia.

3. Vītoliņa I. E-inclusion modeling to improve digital skills of society. Liepājas Universitātes 13. starptautiskā zinātniskā konference “Sabiedrība un kultūra: haoss un harmonija” Liepāja, 2010. gada 28.–29. aprīlis.

4. Vītoliņa I. Wiki approach and student engagement in the learning process. 8th International JTEFS/BBCC Conference “Sustainable Development. Culture. Education”, May 17–19, 2010, Paris, France.

5. Vītoliņa I. E-inclusion process and digital skill development of society. 9th International JTEFS/BBCC Conference “Sustainable Development. Culture. Education”, May 18–21, 2011, Siauliai University, Lithuania.

6. Vitoliņa I., Kapenieks A. 2012. A Study of the e-inclusion process in a real-life e-course delivery context. 10th International JTEFS/BBCC Conference “Sustainable Development. Culture. Education”, May 22–25, 2012, Savonlinna, Finland.

7. Vitoliņa I., Kapenieks A. E-inclusion and knowledge flows in e-course delivery. Proceedings of the 5th International Conference on Computer Supported Education CSEDU 2013, Aachen, Germany, 6–8 May, 2013, pp. 417–422. Aachen: SCITEPRESS, 2013. ISBN 9789898565532

8. Vītoliņa I. Zināšanu plūsmu analīze e-iekļaušanas procesā. Rīgas Tehniskā universitātes 54. starptautiskā konference, Latvija, Rīga 14.10.2013.

9. Vītoliņa I. A. User analysis for e-inclusion in a blended learning course delivery context. Rēzeknes Augstskolas starptautiskā konference “Sabiedrība, integrācija, izglītība”, 2014. gada 23. maijs.

10. Vitolina I., Kapenieks A. E-inclusion prediction modelling in blended learning courses. 23rd International Conference on Interactive Collaborative Learning (ICL2020), September 23–25, 2020.

11. Vītoliņa I., Kapenieks A. (2021). Comparison of e-inclusion prediction models in blended learning courses. 19th International Conference on e-Society (ES 2021), March 3–5, 2021.

12. Vītoliņa I., Kapenieks A. Modeling the e-inclusion prediction system. 13th International Conference on Computer Supported Education CSEDU 2021, April 23–25, 2021.

Publications:

1. Vītoliņa I. (2009). A knowledge management model to promote e-inclusion. *Proceedings*

of the 7th International JTEFS/BBCC Conference "Sustainable Development. Culture. Education": Research and Implementation of Education for Sustainable Development, Latvia, Daugavpils, 5–8 May 2009, pp. 354–374.

2. Vītoliņa I. (2011). E-inclusion modeling to improve digital skills of society. *Liepājas Universitātes 13. starptautiskās zinātniskās konferences "Sabiedrība un kultūra" rakstu krājums*, 869–878. lpp.

3. Vītoliņa I., Kapenieks A. (2013). E-inclusion measurement by e-learning course delivery. *Procedia Computer Science*, *26*, pp. 101–112.

4. Vītoliņa I., Kapenieks A. (2013). E-inclusion and knowledge flows in e-course delivery. *Proceedings of the 5th International Conference on Computer Supported Education CSEDU 2013*, pp. 417–422.

5. Vītoliņa I., Kapenieks A. (2013). Factors predicting e-inclusion in a blended learning course delivery context. *Reorientation of teacher education towards sustainability through theory and practice.*, pp. 199–212.

6. Vītoliņa I., Kapenieks A. (2014). User analysis for e-inclusion in a blended learning course delivery context. *Proceedings of the International Scientific Conference,* Vol. 2, pp. 367–378.

7. Vitolina, I. (2015). E-inclusion process and societal digital skill development. *Discourse and Communication for Sustainable Education*, *6*(1), pp. 86–94.

8. Vitolina, I. (2015). E-inclusion modeling for blended e-learning course. *Procedia Computer Science*, *65*, pp. 744–753.

9. Vitolina I., Kapenieks A. (2021). E-inclusion prediction modelling in blended learning courses. *Educating Engineers for Future Industrial Revolutions: Proceedings of the 23rd International Conference on Interactive Collaborative Learning (ICL2020),* Volume 1, 23, pp. 327–337.

10. Vītoliņa I., Kapenieks A. (2021). Comparison of e-inclusion prediction models in blended learning courses. *Proceedings of the 19th International Conference on e-Society (ES 2021),* pp. 101–108.

11. Vītoliņa I., Kapenieks A., Grada I. (2021). Modeling the e-inclusion prediction system. *Proceedings of the 13th International Conference on Computer Supported Education CSEDU 2021,* Vol. 2, pp. 258–265.

Structure of the Doctoral Thesis

The Doctoral Thesis consists of an introduction, 4 sections, conclusions, references and 6 appendices.

Section 1 gives a description of e-inclusion, identifies the problem area of e-inclusion. Section 2 presents technologies and methods for predicting e-inclusion. Section 3 describes the development of the model predicting the e-inclusion of the individual and the algorithm used in the model. Section 4 describes the technological model (prototype) predicting e-inclusion and evaluates the technological model predicting the degree of e-inclusion of teachers in vocational education institutions. The conclusions summarize the results and conclusions of the Thesis, as well as the directions and possibilities of further research.

1. ANALYSIS OF E-INCLUSION AND ITS PROCESSES

The section aims to analyze the issues of e-inclusion and its processes in order to promote the achievement of the European Union's e-inclusion policy goals and to clarify the scope of e-inclusion, which will be studied in the Doctoral Thesis. The following tasks have been performed in this section:

1. The model describing e-inclusion has been developed with the enterprise modeling method 4EM.
2. The business goals of the model predicting the individual’s e-inclusion have been determined.

The initial version of the descriptive model of e-inclusion is given in the author's publication [1] (in Section “General Description of the Doctoral Thesis”).

1.1. Method for Creating a Model Describing E-inclusion and Its Processes

According to the theory of J. Bubenko (2007), enterprise modeling is a process that results in the creation of a model that represents a company or an object from different aspects. Evaluating the convenience of the availability of method documentation and the existing experience in modeling, the 4EM (previously known as EKD) method was chosen for modeling e-inclusion in the Doctoral Thesis (Stirna & Persson, 2018).

The 4EM method forms a unified model consisting of several interrelated sub-models: goal model, business rules model, concept model, actors and resources model, business process model, as well as technical components and requirements model. Each of the sub-models solves a certain level of problems and uses certain components (Stirna & Persson, 2018).

1.2. Model of E-inclusion and Its Processes

When modeling e-inclusion processes, the author deviates from the 4EM guidelines, as the modeling seminar is not organized, instead the author uses documents – EU or scientific research, EU policy planning documents and progress reports, as well as statistical data as sources of knowledge. As a result of the work, a 4EM model describing e-inclusion has been obtained, which consists of the e-inclusion goals model, rules model, resource and actor model, concept model, business process model (full versions of models are available in the Doctoral Thesis).

E-inclusion Goals Model

E-inclusion has been one of the European Union's digital policy goals for more than a decade, which has not been achieved. The European Union's goal of e-inclusion was defined in 2006 when the European Commission published the Riga e-Inclusion Declaration, aiming to overcome the exclusion, and improve economic performance, employment opportunities, quality of life, and social participation of all individuals and society as a whole. The

Declaration aims to achieve the goal of e-inclusion by promoting the use of ICT and reducing disparities in the use of ICT (European Commission, 2006).

The author uses the goals of the e-Inclusion Declaration as the main ones for the e-inclusion goals model. Their achievement is also relevant in the European Commission's Communication "2030 Digital Compass: The European way for the Digital Decade", which sets digital goals to be achieved by 2030.

The e-Inclusion Declaration identified 6 strategic goals:

- Address the needs of older workers and elderly people
- Reduce geographical digital divides
- Enhance-accessibility and usability
- Improve digital competencies
- Promote cultural diversity in relation to inclusion
- Promote inclusive e-government

However, the goals of the e-Inclusion Declaration were not achieved in line with the results set. In 2010 the European Commission launched the Digital Agenda for Europe, which set the European Union's digital goals for 2020 (European Commission, 2010). The Digital Agenda for Europe updates the objectives set out in the 2006 e-Inclusion Declaration. In 2019, the European Commission published a digital strategy for the next ten years, and on 10 March 2021, the European Commission proposed the Digital Compass (European Commission, 2020; European Commission, 2021). The Digital Compass continues to pursue the goals of e-inclusion by emphasizing that the digital space must provide the same opportunities and rights for individuals as are available to individuals offline. To ensure this, people must have access to a secure and high-quality internet, as well as the opportunity to acquire digital skills and use non-discriminatory digital services.

Barriers to achieving the strategic goals of e-inclusion. Analyzing the problems that hinder the achievement of the goals, it can be concluded that the following repeatedly appear as obstacles:

1. <u>ICT is not available</u>. For example, no internet connection, no access to a computer or smart device for the elderly or immigrants, no services for the elderly, and not provided e-accessibility and usability (Driessen et al., 2011; European Commission, 2006; European Commission, 2010; European Commission, 2020c; European Commission, 2021; Rana et al., 2013; Soja et al., 2019),

2. <u>The lack of digital skills</u>. For example, older people and people in economically disadvantaged regions (Chena, Liu, 2013; European Commission, 2006; European Commission, 2010; European Commission, 2021; Khalil Moghaddam & Khatoon-Abadi, 2013), people with disabilities (Benda et al., 2011; Farbeh-Tabrizi, 2012), immigrants and ethnic minorities (Lupiañez et al., 2015; Yu et al., 2018) have no digital skills.

3. <u>Socio-demographic and economic factors.</u> For example, older people cannot afford to pay for Internet access (Ala-Mutka et al., 2008; Amy, 2011; Townsend et al., 2013).

Analyzing what influences the achievement of e-inclusion goals, it can be seen that **improving digital skills** is a precondition for achieving any strategic goal of e-inclusion. Digital competence has been identified as one of the eight key competences that are essential for everyone (Council of Europe, 2006; Council of Europe, 2018). The European Commission has set a target of having at least 80 % of adults with basic digital skills by 2030.

Digital education ecosystem – for digital literacy. Skills acquisition needs to be ensured on a continuous and regular basis to develop skills in line with technological developments (European Commission, 2010). It is essential to develop a high-performance digital education ecosystem to foster the development of digital skills (European Commission, 2020b; European Commission, 2021). Digital education encompasses two distinct but complementary aspects: firstly, the development of digital competences for both learners and education providers (teachers) and secondly, the use of digital technologies in pedagogical work, education and training systems (European Commission, 2018). Acquisition of digital skills in blended learning courses is welcomed (Gudmundsdottir & Vasbø, 2017; Guillen-Gamez et al., Martínez-Alcalá et al., 2018; Patmanthara et al., 2018).

Data analysis and prediction to improve digital skills. The Digital Education Action Plan 2018–2020 emphasizes the importance of data in digital education processes. One of the priorities is to ensure better data analysis and forecasting (European Commission, 2018). The Digital Education Action Plan states that pilot projects in the field of artificial intelligence and learning analytics should be carried out to facilitate the provision of digital education. A study on the use of learning analytics by the European Joint Research Center indicates that the potential of learning analytics has not been exploited. There are differences between the possibilities indicated in the research literature and the learning analytics introduced in practice. In the European Union, learning analytics is in its early stages of research (Ferguson et al., 2016; Maennel, 2020; Viberg et al., 2018). Therefore, further research is needed.

Model of the E-inclusion Processes and Its Relation to the Target Groups

Since the author concludes that the improvement of digital skills is necessary to achieve any e-inclusion strategic goal and that digital skills are among the most important prerequisites for individual e-inclusion, the individual e-inclusion process model is based on the acquisition of digital skills.

Today the emphasis has shifted from the availability of technology to the meaningful use of ICT (Yu et al., 2017). The author defines that **an individual is e-included if the individual uses digital skills for professional or private needs,** not if the individual has acquired digital skills but does not use them.

In order for an individual to use ICT meaningfully, the individual must go through several stages of the e-inclusion process (Fig. 1.1) (Scheerder et al., 2017; van Dijk, 2006; Yu et al., 2018):

Stage 1: the individual is motivated to use the new technology.

Stage 2: the new technology is physically available to the individual (1st generation digital divide).

Stage 3: the individual has digital skills to use technology (2nd generation digital divide).

Stage 4: The individual uses the new technology in a meaningful way, that is, to address a professional or private need (3rd generation digital divide, digital inequality).

It should be emphasized that the acquisition of new digital skills and their meaningful use are two different steps in the e-inclusion process. The existence of digital skills alone does not yet ensure an individual's e-inclusion (Lerchner et al., 2007; Ono & Zavodny, 2008). Nowadays, there is the concept of "3rd generation digital divide". This divide has developed between those who use technology meaningfully for the benefit of professional needs or personal development and those who use technology but do not benefit from it (Robles & Torres-Albero, 2012). In such situations, **a gap in the use of ICT is formed**, or **a gap**

between an individual's knowledge of ICT and the use of ICT meaningfully (van Deursen & van Dijk, 2015).

The process of e-inclusion is repeated for the individual with the emergence of new ICT (van Dijk, 2006; Yu et al., 2018). In the context of e-inclusion, it is necessary to indicate that the individual's e-inclusion is situational. It refers to the use of specific technologies for a specific purpose, in a specific situation. An individual's total e-inclusion is formed by the set of his situational e-inclusion.

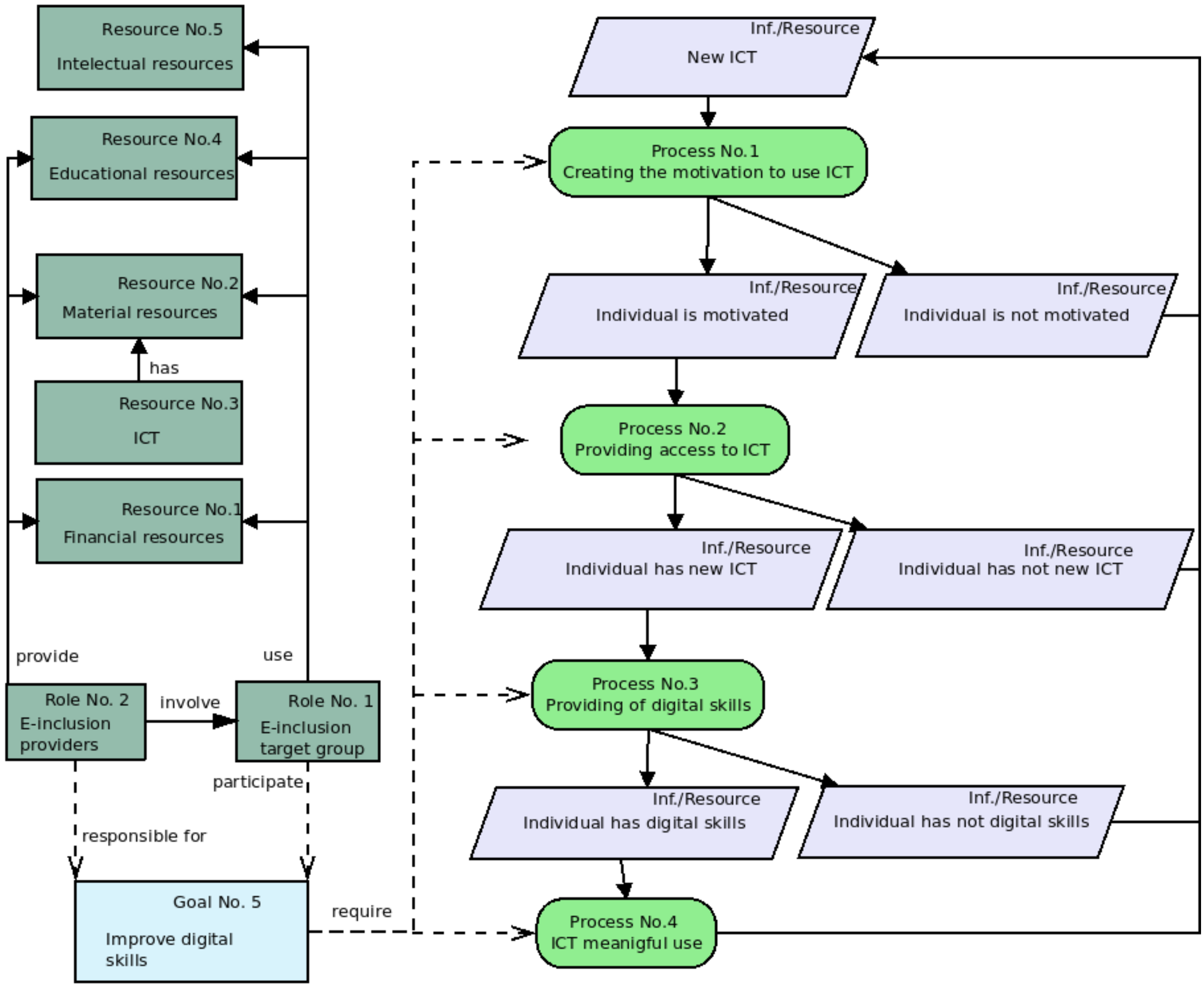

Fig. 1.1. The process model of the e-inclusion of an individual.

E-inclusion Target Groups

E-inclusion refers practically to everyone. The target group of e-inclusion consists of: elderly people, people with lower incomes and people with low level education, people from less developed regions, people with disabilities, women, the unemployed, immigrants, ethnic minorities (Basili, 2013; de Hoyos et al., 2013).

Contrary to the view that a generation that has become familiar with the Internet in childhood has no problems using it, it is pointed out that this generation lacks digital literacy (information and strategic skills) (Santos et al., 2013; van Deursen & van Dijk, 2009).

Digital skills need to be improved for prospective employees and today's employees. One of the target groups for e-inclusion is educators. Their professional activities increasingly require the meaningful use of ICT (Altun, 2019; Rintamäki & Lehto, 2018; Záhorec et al., 2019). In 2014, the "Guidelines for the Development of Education for 2014–2020" supported by the Cabinet of Ministers indicated that it was necessary to continue increasing the professional capacity in the field of ICT for the teaching staff of vocational education (Izglītības un zinātnes ministrija, 2014). A study by the Organization for Economic Cooperation and Development (OECD) in 2018 (Jerrim & Sims, 2019) showed that less than 40 % of educators felt ready to use digital technologies in teaching and that there are large differences within the EU. Educator training is needed both to improve educators' digital skills and to make educators later be part of a high-performance digital education ecosystem.

1.3. Summary and Conclusions

Improving digital skills can be seen as a key goal in achieving the strategic goals of e-inclusion. Improving individuals' digital skills also contributes to other e-inclusion policy goals. The most important step in the e-inclusion process for any individual is the meaningful use of digital skills, which is closely related to both the availability of technology and the acquisition of digital skills.

1. It is necessary to achieve an effective transition of individuals from the stage in which they acquire digital skills to the next stage in which the individual uses acquired digital skills meaningfully. In the Doctoral Thesis, the author has studied how to ensure that an individual moves from the stage of skills acquisition to the stage of meaningful use of skills.

2. It is necessary to identify the factors that facilitate the individual's transition to the last stage of the e-inclusion process, in which the individual uses the acquired digital skills. Current research mainly characterizes individuals using socio-demographic and economic factors and concludes these factors about the age, gender and other characteristics of people. However, socio-demographic and economic factors are difficult to change. It is necessary to look for individual factors in the context of knowledge creation and transfer processes that can be influenced during learning.

3. It is necessary to conduct digital skills training for e-inclusion of individuals as blended learning. Blended learning and online learning are in line with the EU Digital Education Action Plan (2021–2027), which calls for education and training systems to adapt to the digital age.

4. Target group – teachers. Studies and EU policy planning documents indicate that in the context of e-inclusion, attention should be paid not only to "traditional" risk groups, but to everyone. The rationale for choosing teachers is based on the fact that most teachers do not have sufficient digital skills. The choice of teachers as the target group for e-inclusion is also in line with the priorities of the EU Action Plan for Digital Education (2021–2027), which envisages the provision of educational staff and teachers with stable digital competences.

5. The EU Digital Education Action Plan points to the need to use learning analytics to promote digital education.

The author sets the following goals for the predictive model of e-inclusion:

1) to identify an individual who is at risk of not using newly acquired digital skills (risk of not being e-included) by learning analytics;

2) to determine for an individual in the context of knowledge creation and transfer processes risk factors that influence the individual's learning outcome – the e-inclusion of the individual.

2. ANALYSIS OF E-INCLUSION PREDICTION TECHNOLOGIES AND METHODS

To determine the requirements for predictive technologies and methods for the e-inclusion prediction model, the following tasks have been performed in this section:

1. The scientific literature on the technologies and methods of creating predictive models, learning analytics, and its application in the development of predictive models has been studied.
2. Conclusions have been made on the methods and technologies for the development, evaluation, and performance characterization of the e-inclusion prediction model.
3. The stages of the process of creating an individual's e-inclusion prediction model have been determined.

2.1. Predictions, Learning Analytics and Related Concepts

It is assumed that **prediction** means foresight based on facts, observations or assumptions, the formation of a future image with a certain degree of probability (Skujiņa, 2000). Prediction methods can be divided into qualitative and quantitative (Anderson et al., 2012; Tilde, 2014).

Predictive analytics involve a variety of data mining, predictive modeling, machine learning, and statistical techniques to predict future events based on an analysis of current and historical facts (Nyce & Cpcu, 2007).

Predictive modeling is a model development process that generates the most accurate forecast possible (Kuhn & Johnson, 2013).

Prediction is related to almost every discipline of science, and prediction using learning data is a topical issue in educational data mining and learning analytics research (Rudin, 2014).

Learning analytics is the measurement, collection, analysis and review of learner-related data to understand and optimize the learning process and the learning environment (LAK, 2011). Three main user groups are involved in learning analytics: learners, educators/ instructors, institutions (Miteva & Stefanova, 2020; Ortiz-Rojas et al., 2019).

Predictive learning analytics is the analysis of historical and current data obtained from learners and the learning process to create models that allow predicting what improves the learning process and the environment in which it takes place (ECAR-Analytics Working Group, 2015).

Learning analytics uses statistical, data mining, and machine learning methods and technologies to develop prediction models (Lenar et al., 2019).

The growing amount and availability of data related to the learning process make it possible to use new approaches to the analysis of student data, which have already been used in such fields as artificial intelligence and related machine learning (Machine Learning and Learning Analytics Workshop, 2014). **Machine learning** involves teaching computers (systems) to predict the future or in some way unknown events, using statistical and data mining techniques. The goal of the machine learning process is to provide training in a computer program based on experience (Cios et al., 2002).

2.2. Methods and Technologies for Creating Predictive Models

Methods and algorithms based on statistical, data mining and machine learning technology can be used to create predictive models. Several strategies are used in machine learning (Schuh et al, 2020): unsupervised learning, supervised learning, semi-supervised learning, reinforcement learning. An example of unsupervised learning is clustering tasks. The supervised learning addresses classification and regression tasks.

Although the primary task of clustering is to group objects into clusters, several studies are available in the literature where the application of cluster analysis is aimed at prediction (Sorour et al, 2014). One of the algorithms used in cluster analysis for prediction purposes is the *kMeans* algorithm (Sorour et al., 2014; Tamada et al., 2019).

In machine learning, the task of classification is to use a training data set to learn to identify a class that corresponds to a previously unseen object. Classification algorithms can be divided according to the way they predict: (1) predict the class to which the object belongs; (2) predict the probability that shows the extent to which the object belongs to the class.

Another type of classification algorithms is based on how they form a classification model and classify unseen data (Galván et al., 2011). The first type is called lazy learning because the classification model uses the training data set only when it is necessary to determine how to classify a particular unseen example. The second type of learning is eager learning. In this case, a classification model is first created based on the training data, which is then used to classify unprecedented data. Most of the classifiers are eager learning algorithms, such as algorithms based on rules, functions, trees, as well as Bayesian algorithms. An example of a classifier of lazy learning is the kNN (k-nearest neighbor) algorithm. Other lazy learning algorithms are based on the locally weighted learning approach.

In machine learning, the task of regression is to learn how to predict future events with a numeric score using previously known example values. Linear regression is the most widely used supervised predictive modeling for systems based on machine learning approaches (Caraciolo, 2011).

2.3. Predictive Model Evaluation Methods and Performance Metrics

Evaluation of Predictive Models

The main methods for model evaluation are the collection of new data to test the accuracy of predictions, or the division of existing data into several sets to obtain independent measurements of model accuracy (Snee, 1977). Dividing existing data into several subsets to assess the accuracy of the model can be done in several ways. The two methods most commonly used to evaluate the developed model are the holdout method and the cross-validation method (Arlot & Celisse, 2010). In situations where a limited amount of data is available, one of the cross-validation methods is used to objectively evaluate the model (Yadav & Shukla, 2016). One of the cross-validation methods is the k-fold method. The ten-fold cross-validation method is considered to be the best in theory. The cross-validation method has been used to evaluate student achievement classifier models (Buraimoh et al., 2021; Pereira et al. 2019).

Performance Metrics of Predictive Models

Performance metrics of classification model. There is no consensus in the research literature on which performance metrics is best used to evaluate the model (Seliya et al., 2009). The selection of the most appropriate performance metrics should depend on the problem to be solved and the characteristics of the available data set (Novaković et al., 2017).

When it is important to distinguish between types of errors or there are no similar number of examples in the classes, the confusion matrix methodology is used to assess the performance of classification models (Table 2.1).

Table 2.1

Confusion Matrix in a Binary Classifier

	Predicted class Positive	**Predicted class Negative**
Actual Positive class	True positive (tp)	False negative – type II error (fn)
Actual Negative class	False positive – type I error (fp)	True negative (tn)

The confusion matrix consists of 4 different combinations of predicted and actual values: true positive (tp); true negative (tn); false positive (fp); false negative (fn). A true positive is an outcome where the model correctly predicts the positive class. A true negative is an outcome where the model correctly predicts the negative class. A false positive is an outcome where the model incorrectly predicts the positive class. And a false negative is an outcome where the model incorrectly predicts the negative class.

In the binary classification model, its performance is evaluated using precision (2.1), sensitivity (recall) (2.2), accuracy (2.3), balanced accuracy (2.4), F1 score (2.5), and F2 score (2.6):

$$precision=\frac{tp}{tp+fp} \tag{2.1}$$

$$recall=\frac{tp}{tp+fn} \tag{2.2}$$

$$accuracy=\frac{tp+tn}{tp+tn+fp+fn} \tag{2.3}$$

$$balanced\ accuracy=\left(\frac{tp}{tp+fn}+\frac{tn}{tn+fp}\right)/2 \tag{2.4}$$

$$F1\ measure=\frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{2.5}$$

$$F2\ measure=\frac{4 * \text{precision} * \text{recall}}{5 * \ \text{precision} + \text{recall}} \tag{2.6}$$

Precision indicates what percentage of the predictions on the positive class is correct. **Recall** shows the proportion of the positive class (not e-included) instances that are correctly classified. **Accuracy** is defined as the ratio of correctly classified examples in both classes to the total number of examples. **Balanced accuracy** is especially useful when each class has a different number of examples (Brodersen et al., 2010). Balanced accuracy is defined as the average ratio of correctly classified examples in each class.

Complementarity is one of the most important features of evaluating classification models (Novaković et al., 2017). The **F1 measure** is the harmonic mean between precision and recall. The **F2 measure** combines precision and recall, double emphasizing the importance of recall.

Since the Doctoral Thesis research aims to identify students at risk of digital exclusion *(not e-included*), the author considers that a positive class is *not e-included*, but an *e-included* class is a negative class. The true positive (tp) is the number of examples *not e-included* that are correctly predicted. The true negative (tn) is the number of examples of the *e-included* in the class that are predicted correctly. A false positive (fp) is the number of *e-included* class instances that are predicted as *not e-included*. A false negative (fn) is the number of instances of a class *not e-included* that is incorrectly predicted (as *e-included*).

The F measure is used as the primary indicator of model performance in the dissertation. An individual's e-inclusion prediction model is important if its prediction covers as many students at risk as possible. Therefore, not only the F1 measure but also the F2 measure has been used in the dissertation study to emphasize the importance of coverage.

The F measure is used as the primary indicator of model performance in the Doctoral Thesis. An e-inclusion prediction model makes sense if it covers as many students at risk as possible. Therefore, we will also use the F2 measure in our study to emphasize the importance of recall.

Performance Metrics of Regression Models

The most common metrics for estimating regression problems are Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), coefficient of determination, and corrected coefficient of determination.

Performance Metrics of Cluster Analysis Models

As the cluster analysis will be used for prediction and the real values of the dependent variable will be known, it will be possible to use the performance indicators for the classification tasks.

Techniques for Improving the Performance of Predictive Models

Combining predictive models to improve the performance of a student achievement prediction model. To improve the performance of the model (accuracy, precision, recall, etc.), one possibility is to base the final prediction on several models, defining a scheme for combining the individual prediction result of each model (Hung et al., 2019). Another approach to improving the performance of models for predicting student achievement is to form classifier ensembles (Atallah & Al-Mousa, 2019; Mulyani et al., 2019).

Addressing the problem of unbalanced classes in classification. Methods for solving the problem of unbalanced classes are divided into two groups (Haixiang et al., 2017): (1) the level of algorithms, where the goal is to improve the classification algorithms; (2) data level, where the goal is to balance the amount of data in the classes. The most commonly used methods include data-level methods. As the research of the Doctoral Thesis uses relatively small data sets, the author considers methods to increase the number of examples in the smallest class. To avoid duplication of examples, the SMOTE (Synthetic Minority Over-Sampling Technique) method is proposed (Chawla et al., 2002). The SMOTE method synthesizes new examples from the smallest class examples. The SMOTE method has been used to develop predictive models for at-risk students (Mulyani et al., 2019).

2.4. Features Used in Predicting Student Achievement

The author concludes that the features of students presented in the scientific literature are diverse and can be divided according to the change of their values during the learning process (Romero & Ventura, 2020):

1. Constant features during training – demographic and socio-economic data (gender, age, region of residence, economic situation); administrative data (information about the educational institution and teachers, instructors) (Niet et al., 2016; Moncada, 2018).
2. Time-varying features – student activity and interaction with the teacher and learning environment (navigation data, test, task and exercise data, forum reports, etc.); students' psychological and cognitive abilities (motivation, emotional states); students' self-assessments of the learning environment and process-related activities (Cobos & Olmos, 2018; Herrera et al., 2019; Maennel, 2020).
3. Combinations of time-varying and time-constant features. (Mahboob et al., 2016).

Another important aspect in predictive modeling is how to choose the features to be used for prediction:

1. The choice of features is based on a theory that substantiates the relationship between predictors and outcome data. To benefit from learning analytics in the pedagogical process, it is important to understand how individuals learn when selecting the data to be analyzed (Ferguson et al., 2016; Verhoeven et al., 2020). Previous research does not strongly indicate the most appropriate theoretical framework for learning analytics (Nistor, 2015).
2. The features are technically determined with the help of modeling tools. Using the capabilities of machine learning tools, the features with which the model has the highest performance are determined (Márquez-Vera et al., 2013). In such cases, the disadvantage is that the features can be determined by risk factors that cannot be changed during the learning process.

An analysis of the scientific literature shows that when developing models for predicting individual achievement, the goals are mostly limited to predicting whether the student will complete the course or the student will have a successful final grade, but do not identify risk factors that can be influenced to improve student achievement (Cobos & Olmos, 2018; Mahboob, 2016; Sorour et al., 2014).

The author concludes that there is no single set of features that can be taken over and used in a model for predicting e-inclusion. To benefit from learning analytics in the pedagogical process, when choosing the data to be analyzed, it is important to understand how individuals learn and to make the choice of data (attributes) in relation to the individual learning process. When choosing the attributes of the predictive model, it is important to choose the attributes that characterize students, which reflect preventable risk factors, characterize the student learning process, and are based on knowledge management theory.

2.5. Analysis of Models Predicting Students' Achievements

Some research provides information on predictive models that determine an individual's level of ICT use. These models are mainly based on demographic factors or an individual's personality traits and prior digital skills; these factors cannot be changed during the learning process to prevent risk (Azcona et al., 2019; Akhtar et al., 2017; Guillén-Gámez et al., 2020a;

Hidalgo et al., 2020; Verhoeven et al., 2020).

Data sets of different sizes have been used to develop predictive models, for example, models for e-inclusion predictions have both 52 participants (Berkowsky et al., 2017) and 17 000 participants (Hidalgo et al., 2020). As individuals often acquire digital skills in professional development and non-formal education courses, where the number of participants is relatively small, the author in the literature analysis focused on research where prediction technologies have been used for small groups of students. The scientific literature shows that relatively small data sets were used for models that predict student achievement for model training; 76 student data sets (Suresh et al., 2016); 149 student data sets (Azcona et al., 2019); or 273 student data sets (Baksa-Haskó & Baranyai, 2018). A study using a data set of students containing less than 100 examples concludes that a combination of machine learning methods and predictive models can reach as many as 97–100 % correctly predicted at-risk students (Lykourentzou et al., 2009).

Prediction methods in the studies available in the literature are diverse, using cluster analyses (Luo et al., 2020), linear regressions (Akhtar et al., 2017; Berkowsky et al., 2017; Guillén-Gámez et al., 2020a; Xu et al., 2020); and classification methods (Baksa-Haskó & Baranyai, 2018; Buraimoh et al., 2021; Cobos & Olmos, 2018; Hidalgo et al., 2020; Pereira et al., 2019; Suresh et al., 2016).

Studies reviewed in the literature have often used the accuracy of the prediction model as a metric of performance (Alamri et al., 2019; Buraimoh et al., 2021; Mahboob et al., 2016). The precision, recall, F measure are also indicated as metrics of model performance (Márquez-Vera et al., 2013). Coefficients of determination have also been used as performance metrics depending on the chosen method (Berkowsky et al., 2017). The values of the performance metrics differ for the algorithms used. Performance metrics of models created with the same classifier algorithm with different data sets and features are different. This indicates that for each prediction task, an appropriate classifier algorithm must be found, with which the model presents the best performance metrics. Research reveals that in order to improve performance, models use multiple classifiers, compare their performance metrics, and then select those classifiers with higher performance metrics or classification ensembles for further use (Azcona et al., 2019). The analysis of the literature shows that the performance metrics of models predicting student achievement range from 50 % to 100 %, depending on the features used to characterize students and the algorithms used.

The author concludes that it is not possible to directly use the models available in the literature to predict an individual's e-inclusion process because the available studies do not show a model, a set of attributes that would be transferable and reusable. The development of models is influenced by both the purpose of the model and the context in which the learning takes place. The content of the study, the environment, and the characteristics of the students are important. The author's statement that it is not possible to adopt directly created models is in line with the results of a study on the use of predictive models in 17 courses, where it is concluded that it is not possible to reuse existing predictive models in other similar courses or years (Conijn et al., 2017). One of the challenges of learning analytics in developing predictive models is to ensure that models are not only for single use or per course, but can be generalized for a wider range of courses (Hung et al., 2019; Romero & Ventura, 2019). The author concludes that it is necessary to use different data mining methods and to create several models to compare their performance metrics in order to choose models with higher performance metrics.

2.6. Predictive Information System Processes in the Context of Machine Learning Technologies

An information system developed using predictive models based on machine learning technologies is characterized by two main functions: model training and prediction (Berral et al., 2010). In the prediction phase, the knowledge learned for the system is used to perform the tasks assigned to the system (Ning et al., 2011). The prediction stage also includes the following activities – analysis of the obtained information and interpretation of the results (Halkidi et al., 2001). The final stage of the prediction process is the action of the system according to the obtained results (Ribeiro De Carvalho Martinho et al., 2013). For information systems, the development of which is based on the use of machine learning technologies, the quality maintenance function of the predictive model is also important (Studer et al., 2021).

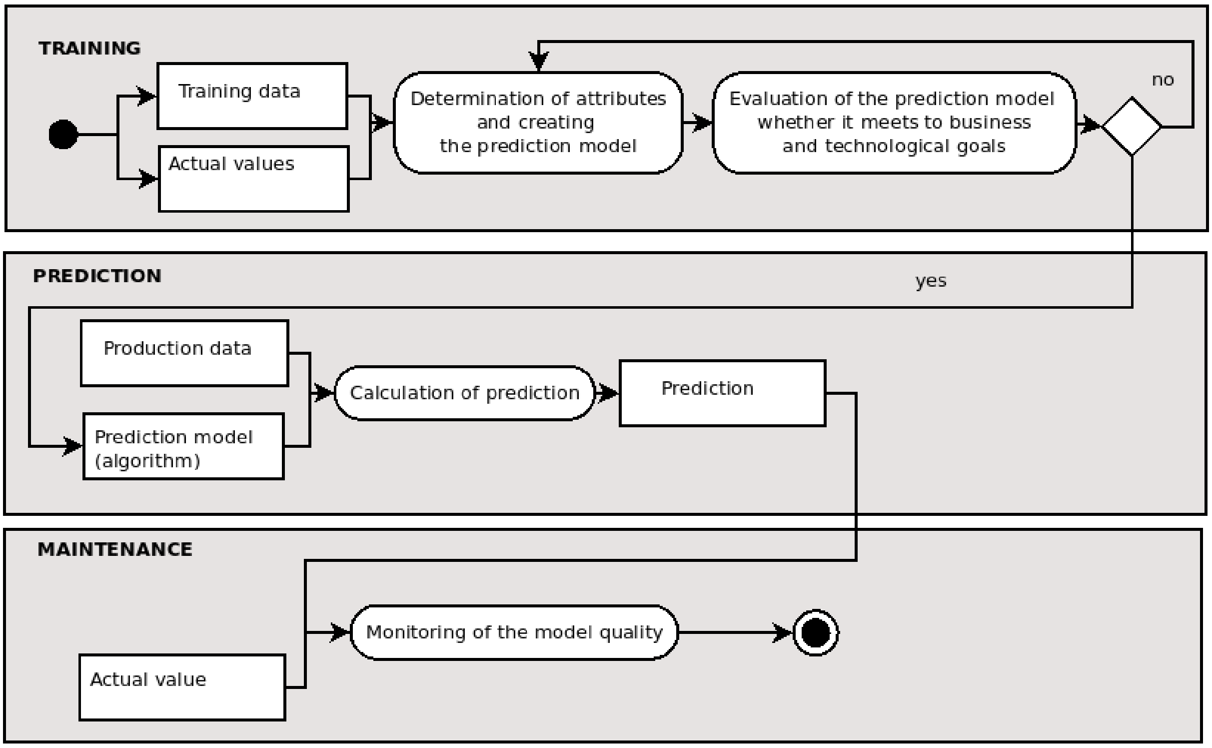

Fig. 2.1. The main functions of predictive information systems: model training, prediction, maintenance.

Conclusions on the e-inclusion predictive information system. The author concludes that it must be ensured that the predictive e-inclusion system (Fig. 2.1):

- is able to learn using data characterizing students, finds an algorithm to identify students at risk;
- is able to predict students at risk.

An optional function of the system for predicting e-inclusion is to monitor and maintain the developed model to ensure the quality of the prediction (Maskey et al., 2019).

2.7. The Process of Creating a Predictive Model and Its Application in the Development of a Technological Model Predicting E-Inclusion

Learning analytics includes the following key processes: acquire and store the data; analyze data; provide information on the result of the analysis (Sclater, 2017). The learning analytics process can be divided in 6 stages according to the CRISP-DM (CRoss-Industry Standard Process for Data Mining) standard, if the predictive model is developed using data mining methods (Baksa-Haskó & Baranyai, 2018). The main steps in creating a data mining model are (1) to understand the problem area (business); (2) to understand the data; (3) to prepare data; (4) to create a model; (5) to evaluate the model; (6) to deploy (implement) the model. Although CRISP-DM is a data mining standard, it is widely used in the development of machine learning-based software (Ekubo, 2020; Shearer, 2000). Based on the CRISP-DM standard, the CRISP-ML (Q) (CRoss-Industry Standard Process Model for the Development of Machine Learning applications with Quality assurance methodology) method has been developed for the development of machine learning models and their quality assurance (Studer et al., 2021). The CRISP-ML method complements the CRISP-DM with a final stage, which provides monitoring and maintaining the quality of the model.

2.8. Steps for the Development of a Model Predicting E-Inclusion

Based on Section 2.7, the author determines the stages of the development process of an e-inclusion prediction model.

The author determines the business understanding stage as the first stage for the development of an individual e-inclusion prediction machine learning-based system. Section 1 of the Doctoral Thesis includes the analysis of e-inclusion processes and the business objectives of the e-inclusion prediction model. The author determines that the second stage is the research of predictive technologies to implement the business goals set for the e-inclusion prediction model. The model must be able to identify risk factors that could affect student achievement. Section 2 of the Doctoral Thesis defines the goals of data mining and machine learning technologies and their application according to the business goals of the e-inclusion process.

The author determines that the third stage is the choice of factors influencing e-inclusion and the development of the prediction model. The development of a model also includes the evaluation of the model using the cross-validation method and the comparison of the model performance metrics. Section 3 of the Doctoral Thesis includes the stages of data understanding and preparation and model development.

During the fourth stage, the author performs model evaluation with test data, which includes prototype development, its operation with test data and evaluation. Section 4 of the Doctoral Thesis contains the evaluation stage of the model predicting e-inclusion.

The fifth and sixth stages, which are not considered in the framework of the Doctoral Thesis, are the full deployment and operation of the model in the production environment and the monitoring and maintenance of the model.

3. STUDY OF FACTORS AFFECTING INDIVIDUAL E-INCLUSION AND PREDICTION MODELS IN THE CONTEXT OF THE PROCESS OF KNOWLEDGE CREATION AND TRANSFER

The section aims to create a model for predicting e-inclusion by performing the following tasks:

1) using knowledge management theory, to determine possible factors,

2) using potential factors, to create a model with which it is possible to predict students at risk of exclusion and to determine which factors should be used in the prediction model.

3) to determine the most appropriate model for predicting student e-inclusion by comparing model performance metrics.

Research on the development of a model (algorithm) predicting the e-inclusion has been performed in accordance with the following methodology: (1) three different data mining methods were used sequentially (linear regression, cluster analysis, classification), and with each of them, a prediction model was created, performance indicators for the prediction model were evaluated, as well as whether the model could be generalized to digital skills courses; (2) combinations of prediction models have been developed, it has been assessed for which combinations of models the performance indicators are improving, as well as it has been assessed whether the model can be generalized to digital skills acquisition courses.

The studies described in this section are detailed in the author's publications [2]–[7], [9], [10].

3.1. Factors Based on Knowledge Management Theory and Influencing E-Inclusion

The gap between knowledge and the practical application of knowledge

The gap between knowledge and the practical application of knowledge is one of the reasons why employees in organizations, when returning from training courses, do not fully use newly acquired knowledge to achieve the goals of the organization (Pfeffer & Sutton, 1999). The gap between the existence of digital skills and their practical application is considered to be the gap between knowledge and the practical application of knowledge in the context of digital skills. We assume that the ideal situation is when the number of actual users of the technology converges with the number of potential users (Becker et al., 2008).

Knowledge Flows in the Context of E-Inclusion

According to Nissen (2006), the knowing-doing gap can stem from problems with knowledge flows.

A knowledge flow has three crucial attributes: direction (sender and receiver), carrier (medium), and content (shareable) (Zhuge, 2004). In the context of the e-inclusion process, the knowledge sender is the instructor or the expert of digital skills; the receiver is the student whose digital skills are improved by these means.

The development of ICT has enhanced the importance of technology within the learning process. Nowadays traditional forms of teaching and learning are often substituted by e-learning to achieve better learning outcomes (Mason, & Rennie, 2008). The carrier can be the e-learning environment and the Internet. Oye and his colleagues emphasize the role of e-learning environments in knowledge transfer; the e-learning environment not only helps students make sense of content, but it also enables on-going communication between students and instructors (Oye et al., 2011).

Nissen stated that for knowledge to flow at the individual level, the instructor or expert must be willing and able to share; the student must be willing and able to learn; and the organization must be willing and able to help him/her do so (Nissen, 2006, p.11).

The e-inclusion of an individual is characterized by the fact that he has digital skills and they are used meaningfully. According to Nisen's approach, in order for the knowledge about digital skills to reach an individual, the instructor must be willing and able to share knowledge about digital skills, the individual must be willing and able to learn and acquire digital skills; the organization must be willing and able to assist the student (for example, by ensuring the availability of technology) and the instructor in transferring and receiving knowledge.

The author proposes to use an approach where the probability that the newly acquired digital skills will be meaningfully used (the individual will be e-included) is determined based on the following factors: Factor 1 – the degree to which the instructor is willing and able to share knowledge; Factor 2 – the degree to which the student is willing to learn; and the learning capacity of the students; Factor 3 – the degree to which the organization supports learning development; the degree to which the organization promotes learning.

Factor 1: Instructor's willingness and ability to share knowledge

The instructor's willingness to share knowledge is understood as the support given to students to facilitate learners' needs. If students use an e-learning environment, then the role of the instructor in sharing knowledge decreases. Knowledge sharing depends on the quality of the content, i.e., learning materials, and the usability of the e-learning environment for convenient use of content and communication with the instructor. In our model, we proposed that the instructor's ability to share knowledge determines the quality of e-course materials and e-learning environment.

Several studies indicate that the instructor influences how students use ICT and how well they are able to learn ICT (Quintana & Zambrano, 2014; Sundqvist et al., 2020). The impact of the learning environment on student achievement is indicated in a study on the development of predictive models (Maennel, 2020).

Factor 2: Students' willingness and ability to learn

Research identifies students' interest as an important motivating element that influences learning outcomes (Subramaniam, 2009). Studies have identified student motivation as a prerequisite for acquiring digital skills (Hatlevik et al., 2015). According to John Dewey's theory (1913), learning outcomes depend on the student's interests.

The author describes the student's ability to learn with the student's previous experience, which is reflected in the level of knowledge. According to the theory of constructivism, each student constructs new knowledge based on his or her existing experience (Powell et al., 2009; Vedins, 2011).

Factor 3: Organizational support and promotion of learning development

In terms of this study, we assumed that the organization is non-biased against all students. All students have the opportunity to complete an e-course for digital skills improvement. The organization actively supports all students.

3.2. Acquisition and Preparation of Data for Research on the Development of Predictive Models and Evaluation of Possible Factors Influencing E-Inclusion

767 teachers of vocational education institutions who studied the module "Increasing Information Technology Skills Competencies" in the period from November 2011 to May 2012, are involved in the research of this Doctoral Thesis. Eleven topics of the module were used for Doctoral Thesis. For some studies, sample data were supplemented with data obtained from 160 teachers of vocational education institutions, who in the period from 2014 to 2016 completed three courses of the study program "Education of Modern Interests" – Mobile Technologies, Robotics, and Video Technologies and Design.

The author used 25 surveys as the data acquisition method. Surveys are an appropriate form of feedback to improve student achievement (Prokofyeva et al., 2019). 24 questionnaires were placed in the Moodle system during the training course. After completing the course, a telephone survey was conducted or an e-mail survey was sent to students to find out the actual use of digital skills for professional or private purposes.

Description of Variables and Attributes Used in the Research

Independent variables. Based on the theory discussed in Subsection 3.1, the author uses two factors and the corresponding independent variables to characterize the e-inclusion process (Table 3.1).

Table 3.1

Factors Characterizing the E-Inclusion Process and the Corresponding Independent Variables

Factor	**Independent variable**	**Data acquisition**
I. Instructor's willingness and ability to share knowledge	Instructor's willingness to share knowledge, IWS Value: from 1 (very low) to 5 (very high)	Survey after the course
	Student's evaluation of e-learning materials, ELM Value: from 1 (completely dissatisfied) to 5 (completely satisfied)	Survey after completion of each topic
	Student's evaluation of e-learning environment, ELE Value: from 1 (completely dissatisfied) to 5 (completely satisfied)	Survey after completion of each topic
II. Student's willingness and ability to learn	Student's willingness to learn, SWL Value: from 1 (no desire at all) to 5 (expressed desire)	Survey before starting the course
	Student's ability to learn, SAL This variable is interpreted as the percentage increase in the student's level of knowledge Variable value: 0 % (no ability) to 100 % (excellent ability) or scaled from 1 to 5	Survey before learning each topic Survey after completion of each topic

In addition to the above-mentioned independent variables, the author uses the following variable:

- Student's general digital skills, DS. This variable is derived from the student's self-assessment of his/her digital and Internet skills. Variable value: from 0 (no skills) to 1 (excellent skills). The value of the variable was scaled from 1 to 5.

Dependent variables. To determine the degree of e-inclusion of an individual in the e-inclusion process, the author uses several characteristics that indicate the e-inclusion of the individual: student's prognosis about the use of newly acquired skills, actual use of observed skills, possible use of skills. Information about students' self-predictions in the use of newly acquired skills is considered important by the organizers of various online courses (Future Learn, 2020), as well as predictions about the use of digital technologies are used in research to determine individuals' digital competencies (Kreijns et al., 2014). The variables that characterize the degree of e-inclusion of the individual are given in Table 3.2.

Table 3.2

Characteristics of Dependent Variables

Dependent variable	**Data acquisition**
Student's self-prediction on whether the student will use the acquired skills for professional or private needs after completing the course, PU. Value: from 1 (completely dissatisfied) to 5 (completely satisfied)	Survey after the course
Observed skills usage, OU. This variable is derived from the student's self-assessment of the use of newly acquired skills after the completion of the course. The variable has three possible values: 0 – No, I have not used any skills related to this topic at all 1 – No, but I use skills at the same level as before the course 2 – Yes, I use the newly acquired skills	Survey on the use of newly acquired digital skills, depending on the type of course, 4 to 8 weeks after the end of the course or up to 6 months after the end of the course.
Potential skills usage, PU&OU. The value of this variable is obtained by combining the values of the variables Student's self-prediction and Observed skills usage. Variable value: from 1 (no probability) to 7 (there is a strong probability). The variable value for some studies was scaled from 1 to 5	Survey after the course. Survey on the use of newly acquired digital skills, depending on the type of course, 4 to 8 weeks after the end of the course or up to 6 months after the end of the course.

3.3. Development of a Model Predicting E-Inclusion Based on Linear Regression Approach and Evaluation of Factors Influencing the Prediction

This study aims to test the extent to which the linear regression model, using predefined possible e-inclusion factors that characterize the knowledge flow between the instructor and the individual, can predict the degree of e-inclusion for VET teachers learning digital skills under the guidance of an instructor in the e-environment.

Research questions:

1. What is the degree of correlation between the student's assessment of the instructor's willingness to share knowledge and the student's degree of e-inclusion?
2. What is the degree of correlation between the student's level of satisfaction with e-learning materials and the student's e-inclusion level?
3. What is the degree of correlation between the student's level of satisfaction with the

e-learning environment and the student's e-inclusion level?

4. What is the degree of correlation between the student's desire to learn and the student's degree of e-inclusion?

5. What is the degree of correlation between the student's learning abilities and the student's degree of e-inclusion?

6. To what extent is it possible to predict the degree of student's e-inclusion based on the student's level of interest and ability to learn, the student's level of satisfaction with e-learning materials and e-learning environment and the instructor's willingness to share knowledge?

Data analysis method. To find out the relationship between e-inclusion factors and e-inclusion degrees, the author uses a Pearson correlation coefficient. In order to create a model predicting e-inclusion and to find out the influence of e-inclusion factors on the degree of e-inclusion, the author uses multifactor linear regression analysis. The compliance of the variables with the normal distribution was checked.

Results: Correlations between e-inclusion factors and e-inclusion degree

According to Pearson's correlation calculations, the correlation coefficients between the variables characterizing the e-inclusion factors and the variables characterizing the e-inclusion degree range from 0.23 to 0.64.

Results: Linear regression-based model predicting e-inclusion

Correlation calculations show that linear regression models can be created to predict e-inclusion using as outcome variable: (1) the student's self-prediction that after completing the course the student will use the acquired skills for professional or private needs; (2) potential use of skills after completion of the course. Correlation studies show that the relationship between e-inclusion factors and an individual's e-inclusion level varies across the course topics.

According to the research questions, it is concluded that it is possible to predict the degree of student's e-inclusion with factors, which can be expressed by the following variables – student's interest level and ability to learn, student's satisfaction level with the e-learning materials and e-environment. The degree of student's e-inclusion in the linear regression model is measured in two ways: (1) as a student's self-prediction that the student will use the acquired skills for professional or private tasks after completing the course; (2) as potential use of newly acquired skills after completion of the course.

The results of linear regression modeling show that the models used for prediction and the variables of e-inclusion factors are different for different courses. The results do not show a single model predicting e-inclusion.

Predicted usage and predictors. The most commonly used predictors for student's self-predicted usage are the following independent variables of Factor 1: Student'*s evaluation of e-learning* materials (43.75 %) and *Student's evaluation of e-learning environment* (37.5 %).

Factor 2 variable *Student's willingness to learn* affects the predicted usage relatively less, it is included in the models as the second factor 12.5 %. Variable *Student's ability to learn* determined the predicted usage only in 6 % of models.

Potential usage and predictors. The most commonly used predictors for potential usage are the following independent variables of Factor 1: *Student's evaluation of e-learning environment* (37.5 %), *Student's evaluation of e-learning materials* (25 %), and *Student's* *ability to learn* (25 %).

Factor 2 variable *Student's willingness to learn* affects the potential usage relatively less, it is included in the models as the second factor 12.5 %.

The obtained results show that linear regression models can be used to predict the degree of e-inclusion of an individual. With the obtained linear regression e-inclusion prediction models it is possible to explain the degree of individual e-inclusion from 13.1 % to 46.2 % of the total number of variations. Since the linear regression method explains a relatively small percentage, the author continues the research of e-inclusion prediction models using cluster analysis methods in Subsection 3.4.

3.4. Development of Clusters Characterizing an Individual's E-Inclusion and Evaluation of Differences in the Factors Influencing E-Inclusion in the Clusters

The study aims to find out (1) the possible e-inclusion factors, which characterize the flow of knowledge between the instructor and the individual; (2) the differences between VET teachers who acquire digital skills under the guidance of an instructor in the e-environment and have different degrees of e-inclusion.

Research questions:

1. How does the students' assessment of the instructor's willingness to share knowledge differ for students who use the newly acquired digital skills after the end of the course, and for students who do not use the newly acquired skills after the end of the course?

2. How does the level of student satisfaction with e-learning materials differ for students who use the newly acquired digital skills after the end of the course, and for those students who do not use the newly acquired skills after the end of the course?

3. How does the level of student satisfaction with the e-learning environment differ for students who use the newly acquired digital skills after the end of the course, and for those students who do not use the newly acquired skills after the end of the course?

4. How does the students' willingness to learn differ for students who use the newly acquired digital skills after the end of the course, and for those students who do not use the newly acquired skills after the end of the course?

5. How does the students' learning abilities differ between students who use the newly acquired digital skills after the end of the course and those who do not use the newly acquired skills after the end of the course?

Data analysis method. To group VET teachers based on their characteristics and to compare these groups according to several parameters, the author used an iterative distance-based clustering approach known as *kMeans* algorithm, where k is number of clusters. It is difficult to determine the optimal number of clusters (Hamerly & Elkan, 2003).

To determine the number of clusters for the *kMeans* algorithm, the author used the *EM* (Expectation-Maximization) algorithm, as well as empirically determined the value of k (Osamor et al., 2012). The author performed the cluster analysis with the open source software WEKA. Students were divided into three clusters depending on their answer about the usage of newly acquired skills: (1) I have not used any skills related to this topic at all; (2) I use the skills at the same level as before the course; (3) I use the newly acquired skills.

Results according to the research questions

Instructor's willingness to share knowledge. Students' assessment of the instructor's willingness to share knowledge does not differ for students who continue to use the acquired digital skills after the end of the course and for those students who do not use the newly acquired skills after the end of the course.

E-learning materials. The level of student satisfaction with e-learning materials differs for students who continue to use the acquired digital skills after the end of the course and for those students who do not use the newly acquired skills after the end of the course.

A higher evaluation of e-learning materials indicates that students use newly acquired digital skills. To achieve the goals of the e-inclusion process, it is necessary to ensure in the e-learning environment that the study materials are adapted to the needs of the student.

E-learning environment. The level of student satisfaction with the e-learning environment differs for students who continue to use the acquired digital skills after the end of the course and for those students who do not use the newly acquired skills after the end of the course. The highest score of 4.5, which is close to the maximum possible of 5.0, indicates that the student belongs to a cluster of students who will continue to use the newly acquired skills, while the lowest scores are close to the ratings of other clusters of students that do not use them. The author concludes that attention should be paid to how the student feels in the e-learning environment.

Student's willingness to learn. Students' willingness to learn does not differ for students who continue to use the acquired digital skills after the end of the course and for those students who do not use the newly acquired skills after the end of the course.

Student's ability to learn. Students' learning abilities differ for students who continue to use the acquired digital skills after the end of the course and for those students who do not use the newly acquired skills after the end of the course.

The obtained results show that the ability of some students who use the newly acquired skills to learn is higher than the values of other student clusters. But for some students who also continue to use their newly acquired digital skills, their ability to learn is the lowest. The low ability of students to learn can be explained by the fact that the in the beginning the student's self-assessment of his knowledge of the topic has been relatively high or even the highest. Therefore, the student has not had the opportunity to show that his knowledge of the topic has increased.

The results of the cluster analysis indicate that students can be divided into clusters based on the way how they use the newly acquired digital skills after completing the courses. The centroids of the variables characterizing students are different for different clusters, however, it can be observed that those clusters of students who use newly acquired skills have higher values of centroids compared to those clusters where students do not use newly acquired skills. The author concludes that cluster analysis can be used to predict student e-inclusion using individual e-inclusion factors.

3.5. Modeling an Individual's E-Inclusion with Classification Algorithms and E-Inclusion Factors

The study aims to find out to what extent, using classification-based methods and pre-defined factors characterizing an individual's e-inclusion, it is possible to predict e-inclusion for teachers of vocational education institutions who acquire digital skills under the guidance of an instructor.

Research questions:

1. To find out which classifiers generate e-inclusion predictive models with higher performance indicators.

2. To find out if there is a classifier with which the generated e-inclusion predictive models have the highest performance indicators for all training courses.

3. To find out how the performance indicators of the models differ for different types of data sets.

The author prepared 12 data sets for the research. Data sets contain the records of all students or contain only the records of students who have the technology to use the newly acquired skills. The data sets were also balanced. The datasets were created for each of the courses: Mobile Technologies, Robotics, and Video Technologies and Design.

Method of creating a predictive model. In order to classify VET teachers into *e-included* and *not e-included*, 5 classification models were developed using the following classifier generation algorithms: NaiveBayes, SimpleLogistic, LWL, OneR, and LMT in the Waikato Environment for Knowledge Analysis (WEKA) platform.

To evaluate the performance of the developed classification models, the author uses the methodology of the confusion matrix and determines F1 measures. 10-fold cross-validation is used to evaluate classification models predicting e-inclusion (Yadav & Shukla, 2016).

Results: Modeling of individual's e-inclusion with classifiers and e-inclusion factors

According to the research question, it has been found that different types of data sets have different performance indicators for the F1 measure. The F1 measure obtained in the study confirms that balancing the data set before the classification model generation procedure improves the performance results for the classification models, especially the models generated by the lazy.LWL algorithm.

The study compares five different classification models and finds that (1) models that use balanced data sets and (2) students who have access to technology have the best performance for the F1 measure.

According to the research question, it has been found that there is no single classifier generation algorithm with which the classification models show the best performance indicators in all three training courses. The classification model created with the LMT algorithm has the highest F1 measure of 0.842 in the Robotics course. In the course of Mobile Technology, the highest F1 measure is 0.818 for the classification model created with the lazy.LWL algorithm, but in the Video Technology and Design course, the highest F1 measure is 0.804 for the classification model created with the LMT algorithm.

According to the research question, it has been found that out of all five classifier generation algorithms used two can be highlighted: lazy.LWL and LMT – the models created with these algorithms showed the highest performance metrics.

In addition, the models created with lazy.LWL and LMT algorithms for balanced data sets showed the highest average F1 measure of all study courses: the average F1 measure of lazy.LWL classification models is 0.768, the average F1 measure of LMT classification models is 0.770. Thus, it can be considered that lazy.LWL and LMT classification models can be used for various digital skills development courses.

The performed research allows concluding that the factors of individual e-inclusion with classification methods allow predicting the individual e-inclusion and that it is possible to train the model predicting the individual e-inclusion using classifier development algorithms. The results of the study did not confirm that any of the classifier algorithms used to generate the classification models show the highest performance in all three courses.

Based on these conclusions, the author conducts research in the next subsection, creating a model for predicting an individual's e-inclusion, training it with data sets that combine data from several digital skills courses.

3.6. Development of an Algorithmic Model Predicting the E-Inclusion of an Individual

The study aims to combine linear regression, cluster analysis and classification methods to create a model that predicts individual's e-inclusion with the highest possible performance indicators while recognizing as many students at risk of digital exclusion as possible.

Research questions:

1. For which combinations of linear regression, cluster analysis and classification models, the performance indicators of the e-inclusion predictive model are higher?

2. What percentage of all VET teachers at-risk of digital exclusion who acquire digital skills can the model predict as belonging to the group at risk?

3. What percentage of the predicted VET teachers at-risk of digital exclusion belong to the at-risk group?

Method of creating a predictive model. In order to create a model predicting an individual's e-inclusion, three models were developed by training them with a data set containing data from various digital skills development courses: (1) model based on classifier ensemble, (2) model based on cluster analysis, (3) model based on linear regression.

Then combinations of these three models were developed, looking for combinations with the highest performance metrics but taking into account the condition that the model should be able to recognize as many students at risk of digital exclusion as possible.

The author used the WEKA platform to train models (Frank et al., 2009). Model performance was assessed using recall, precision, F1 measure, and F2 measure. 10-fold cross-validation is used to evaluate the model predicting e-inclusion.

Prediction Model M1: a Model Based on Classification Ensemble

Based on the previous results that the models generated by lazy.LWL and LMT algorithms have the highest performance and their combination with the models created using algorithms: NaiveBayes, Simple Logistic and OneR, the author determined the model combinations with the highest performance metrics (F1 and F2 measures). Model M1 is an ensemble classifier that combines:

1) 4 classification models developed with lazy.LWL (with Random Forest), LMT, OneR, and Simple Logistic algorithms, using the majority voting approach, if the F1 measure is considered as the model performance criterion;

2) 3 classification models developed with lazy.LWL (with Random Forest), LMT, and Simple Logistic algorithms using the majority voting approach, if the F2 measure is considered as the model performance criterion.

Prediction Model M2: a Model Based on Cluster Analysis

The prediction model M2 is developed using a cluster analysis approach. The *kMeans* cluster analysis method is used to group students into 2 clusters (*e-included* and *not e-included*).

Prediction Model M3: a Model Based on Linear Regression

The linear regression model predicts the degree of e-inclusion as a percentage of the maximum possible degree. To determine the individual's e-inclusion level threshold, the predicted e-inclusion level was compared with the actual observation of whether the student corresponds to the class *e-included* or *not e-included*. The author concluded that the highest value of F1 measure is when the degree of e-inclusion is 75%, the highest value of F2 measure is when the degree of e-inclusion is 85 % (Fig. 3.1).

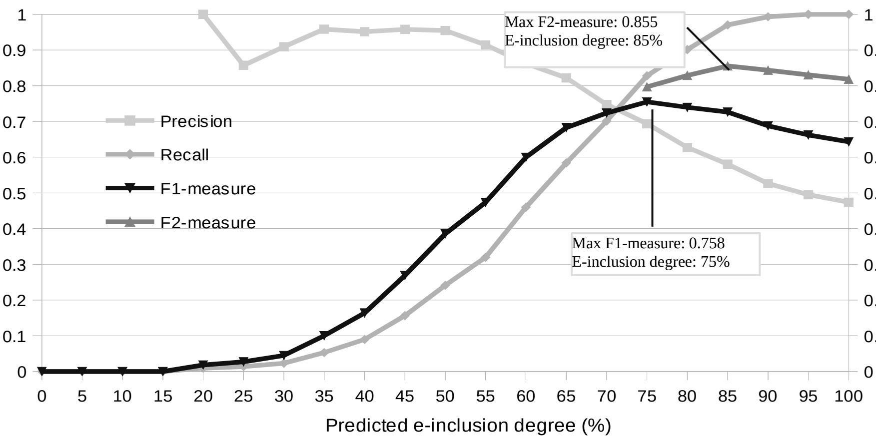

Fig. 3.1. Changes in performance metrics depending on the predicted degree of e-inclusion. Determining the e-inclusion threshold.

The prediction Model M3 is a multifactor linear regression model that assumes that the learner is at risk of digital exclusion:

1) if the predicted degree of e-inclusion is less than 75 % and the F1 measure is the determining metric;

2) if the predicted degree of e-inclusion is less than 85 % and the F2 measure is the determining metric.

Combinations of E-inclusion Models

The author studied how to improve recall and to reduce the number of students the model predicts as e-included, but in reality, students are not e-included. The author created four additional models: Models M4, M5, M6, M7. These models are combinations of Model M1, M2 and M3 which are based on changing the prediction value from *e-included* to *not e-included* if the prediction value of the second (added) model is *not e-included*. Thus, as a result of prediction, the recall increases, and the number of students who are at risk of digital exclusion and who are recognized by the model as students at risk increases.

Results: F1 Measure as a Key Performance Indicator

The obtained results (Fig. 3.2) show that t<u>he F1 measure is the highest for the model that combines a classifier ensemble and model based on clustering</u>.

The highest F1 value of 0.812 is for Model M4, which is obtained by combining Model M1 and Model M2. Model M1 uses classification models based on algorithms lazy.LWL, LMT, OneR and Simple Logistic, combining them with the ensemble and majority voting method. Model M2 is a *kMeans* cluster analysis model where students are grouped into *e-included* and *not e-included* clusters. The recall of the M4 model is 0.828, the precision is 0.796.

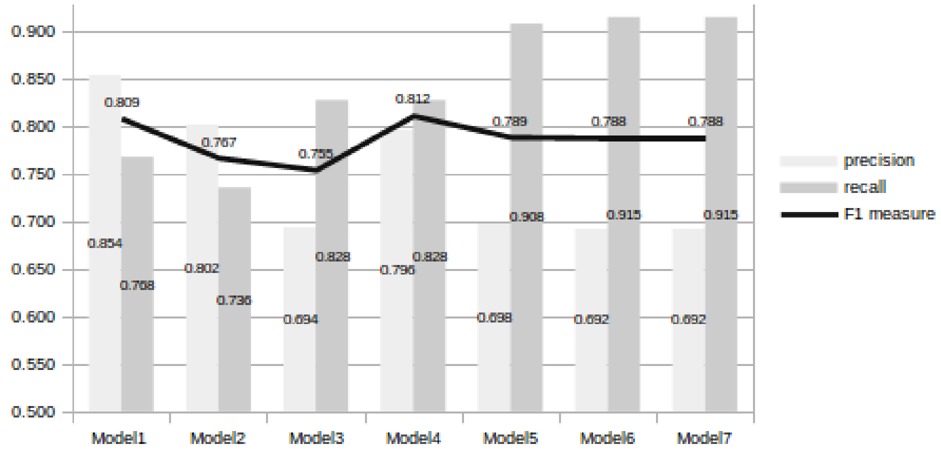

Fig. 3.2. F1 measure, precision, recall values for e-inclusion prediction models considering recall and precision with the same importance.

Results: F2 Measure as a Key Performance Indicator

The obtained results (Fig. 3.3) show that the performance indicator <u>F2 measure is the highest for models obtained by combining a classifier ensemble with a linear regression model or a classifier ensemble with a linear regression model and a cluster analysis model.</u>

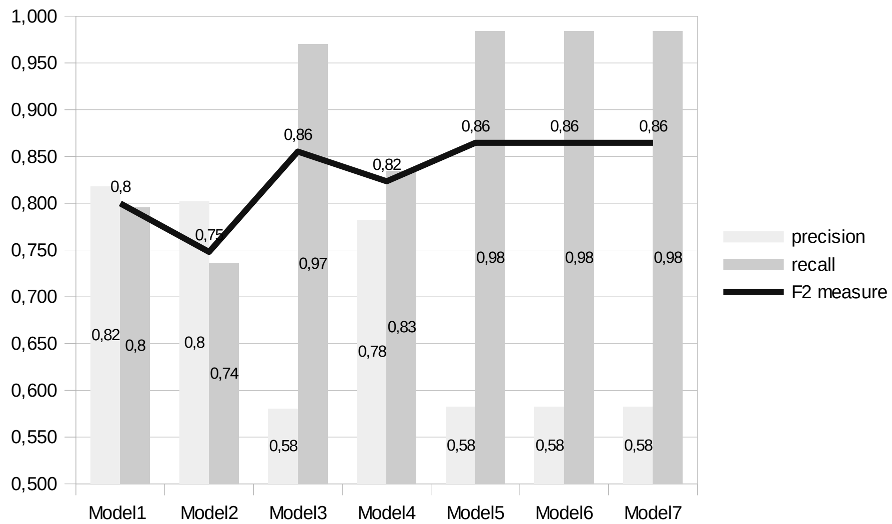

Fig. 3.3. F2 measure, precision, recall values for e-inclusion prediction models considering recall more important than precision.

The highest F2 measure value 0.865 is for Models M5, M6 and M7. Model M5 is obtained by combining Models M1 and M3. Model M1 uses classification models based on the lazy.LWL, LMT and Simple Logistic algorithms, combining them with the ensemble method and using the majority voting method. Model M3 is a linear regression model where the learner is considered to be e-included if his e-inclusion degree is higher than 85 %. Models M6 and M7 are the result of a combination of Models M1, M2, and M3. A combination of classification, cluster and linear regression methods has been used for these models. For ModelsM5, M6, and M7, the recall is 0.984 and the precision is 0.582.

The author concludes that if the goal is to cover as many students at risk as possible, then the classifier ensemble method combined with a linear regression model or with cluster analysis and linear regression model is appropriate. Such a model is able to recognize **98.40 %** of those who are digitally excluded, but at the same time, only **58.20 %** of those who are predicted to be digitally excluded are true students at risk.

If precision and recall are equally important, then the highest performance indicators can be obtained by combining the classifier ensemble model and the cluster analysis model. This prediction model recognizes **82.80 %** of digitally excluded students and is able to correctly predict **79.60 %** of those students who are predicted as at-risk students.

Determining Individual's E-Inclusion Risk Factors

In order to determine the risk factors of a particular student, the author uses the values of cluster centroid as values that must be reached for the attributes of a particular student for the student to correspond to the e-included group (Fig. 3.4).

Final cluster centroids:

		Cluster#	
Attribute	Full Data	0	1
	(928.0)	(529.0)	(399.0)
SWL	3.7884	4.0457	3.4471
DS	3.7183	3.8601	3.5303
SAL	3.1094	3.7719	2.231
ELM	4.0477	4.4805	3.4739
IWS	4.6056	4.8507	4.2807
ELE	4.1525	4.5388	3.6404
PU	3.986	4.569	3.213

Fig. 3.4. Centroids for *e-included* and *not e-included* student clusters in the M2 model. Cluster "0" is for the *e-included* class, cluster "1" is for the *not e-included* class. Attributes are students' self-assessments. SWL – student's willingness to learn, DS – digital skills level, SAL – student's ability to learn, ELM – student's evaluation of e-learning materials, IWS – instructor's willingness to share knowledge, ELE – student's evaluation of e-learning environment, PU – student's self-prediction that the student will use the newly acquired skills.

A second view of student risk factors is provided using the linear regression Model M3. The obtained coefficients of the linear regression model are given in Fig. 3.5.

```
PUOU =

      0.538  * SWL +
      0.2807 * SAL +
      0.6354 * ELM +
      0.3163 * ELE +
     -1.7483
```

Fig. 3.5. Values of linear regression model coefficients of Model M3. Attributes are students' self-assessments. SWL – student's willingness to learn, SAL – students' ability to learn, ELM – student's evaluation of e-learning materials, ELE – student's evaluation of e-learning environment. PUOU – potential value of e-inclusion degree.

Linear correlation coefficients indicate that student characteristics have different effects on prediction. E-learning materials and student motivation have a greater impact, the e-learning environment and the student's ability to learn have a smaller impact. Depending on the attributes of a particular student, it is possible to determine with a linear regression model which are the determining risk factors for the student.

Predicting the e-inclusion of VET teachers, the author believes that it is more important to recognize as many representatives of the risk group as possible (greater recall) than to ensure greater precision of the model but not to notice individuals at risk. Comparing the performance indicators of the models, the author concludes that the highest recall value, while maintaining the highest overall performance of the model, is in the case when the determining performance indicator is F2 measure and prediction is based on Models M5, M6, M7, obtained by combining a classifier ensemble with a linear regression model or a classifier ensemble with a linear regression model and a cluster analysis model.

4. TECHNOLOGICAL MODEL AND ITS EVALUATION FOR PREDICTING INDIVIDUAL E-INCLUSION

The section aims to evaluate the algorithmic model predicting e-inclusion by deploying it in the prototype of the e-inclusion prediction system.

To achieve the goal, the following tasks have been performed:

1) the software requirements of the e-inclusion prediction system are defined;

2) a prototype (technological model) of the system predicting e-inclusion has been developed following the specified requirements, deploying an algorithmic model predicting e-inclusion in it;

3) the algorithmic model predicting e-inclusion and the developed prototype are evaluated.

The author's publications on the research described in this section are available [8], [11].

4.1. Functional Requirements and Main Principles of the System Predicting E-Inclusion

Goals of the E-Inclusion Prediction System

The main goal of the e-inclusion prediction system is to identify students at risk of e-inclusion. The first sub-goal is to build a knowledge base consisting of a database of training examples and a trained prediction model. The second sub-goal is to predict students at risk of digital exclusion based on the pre-defined e-inclusion threshold and calculations obtained using a prediction model. If the student is at risk of being excluded, then the task is to determine what factors affect student performance. The third sub-goal is to monitor the performance quality of the prediction model.

Functional Requirements for the E-Inclusion Prediction System

According to the goals of the system, the main functional requirements for the e-inclusion prediction system are:

1. Predict students at risk of digital exclusion
2. Ensure the quality of results – maintain the performance of prediction model
3. The interface must be simple and easy to use

E-Inclusion Prediction System Basic Processes and Data Flows Between Them

Figure 4.1 presents a context-level data flow diagram for the e-inclusion prediction system. The main user of the e-inclusion prediction system is an instructor who teaches students in the blended e-learning courses. The instructor sets values of the e-inclusion degree threshold level and receives information on risk students and risk factors. The e-inclusion prediction system receives student data and the topic from the learning management system (LMS). To get feedback from students on the usage of the learned skills, the system sends SMS messages to students' smartphones. The decision to use the SMS approach for communication with students is based on previous experience delivering blended learning courses by the multi-screen approach. Then the database is supplemented with data on the actual use of newly acquired skills.

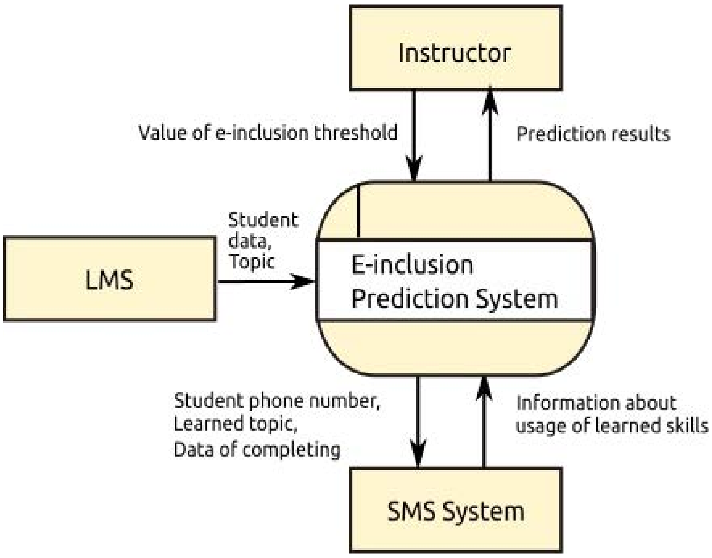

Fig. 4.1. Context-level data flow diagram showing the relationship between the e-inclusion prediction system, instructor, LMS, and SMS system.

Algorithm for predicting student e-inclusion

The e-inclusion predictive algorithm consists of algorithm of e-inclusion prediction model training process, a prediction process algorithm, and a performance monitoring algorithm.

Algorithm of e-inclusion prediction model training process. Figure 4.2 shows the training process of the e-inclusion prediction model. To obtain the e-inclusion prediction model PROGN, the first three different prediction Models M1, M2 and M3 are trained, then the optimal combination of these models is determined and the final prediction function PROGN = f (M1, M2, M2, e-inclusion threshold) is calculated.

Model M1 is a prediction model that combines classification models created with the following algorithms in the classification ensemble – lazy.LWL with Random Forest, LMT and Simple Logistic algorithms, using the majority voting approach. The prediction Model M2 uses the *kMeans* clustering algorithm; it divides students into two clusters, where each of the clusters corresponds to *e-included* or *not e-included* students. The prediction Model M3 is a multifactor linear regression model that predicts the learner's e-inclusion level according to a predefined threshold.

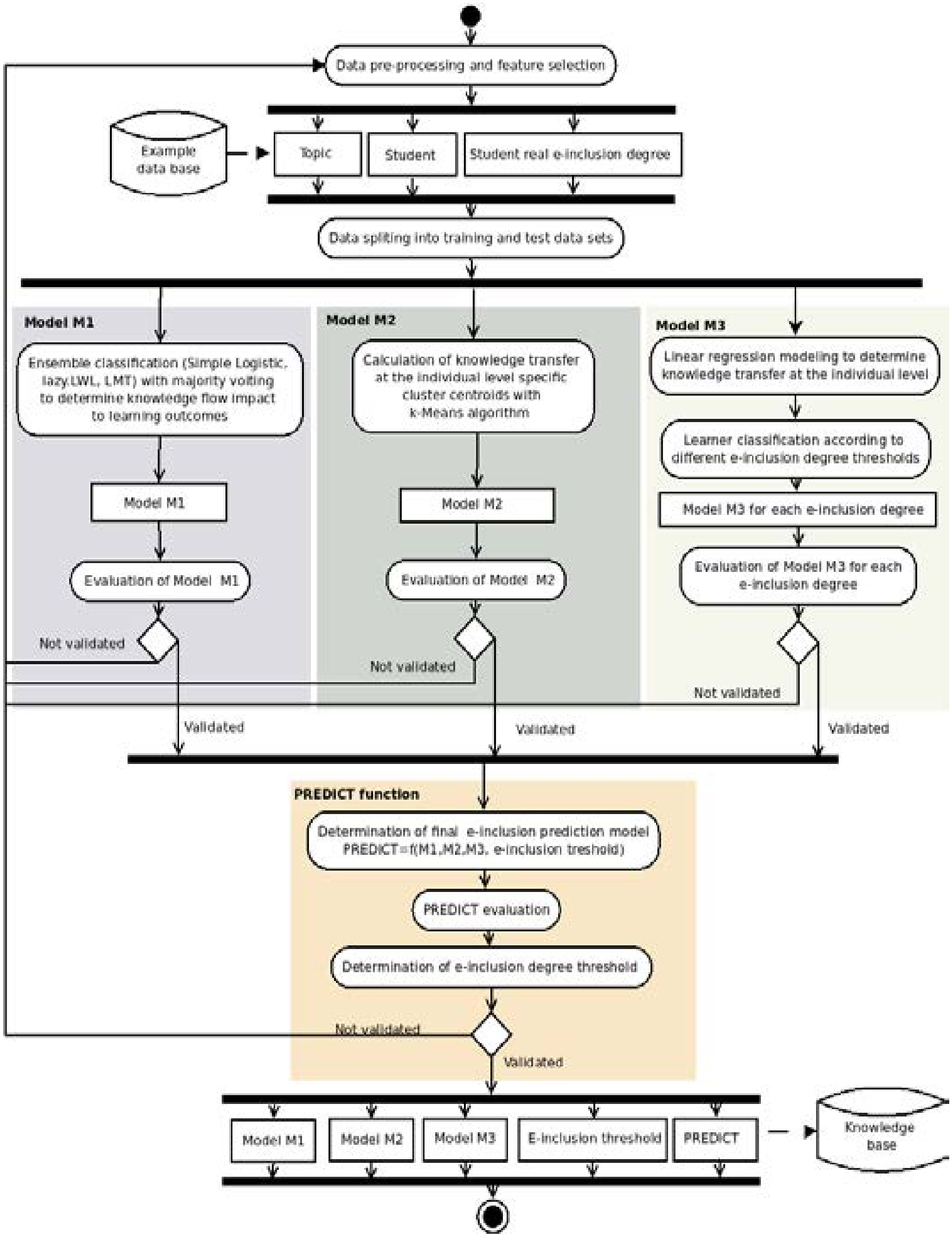

Fig. 4.2. The training process of an individual's e-inclusion prediction model (algorithm) using three model training and PROGN function calculation.

The PROGN function finds the optimal combination of three models' predictions (Fig. 4.3). If Model M1 predicts that the student will not be digitally included, then the final result will be that the student is at risk. If Model M1 predicts that the student will be digitally included, then the next step is to test the prediction of Model M2. If Model M2 predicts that the student is not digitally included, then the final result is again that the student is at risk. Model M3 is similarly tested. This approach is chosen to cover as many students as possible who are potentially at risk. In only one case the system predicts that the student is not at risk, i.e., if all three models predict that the student is digitally included.

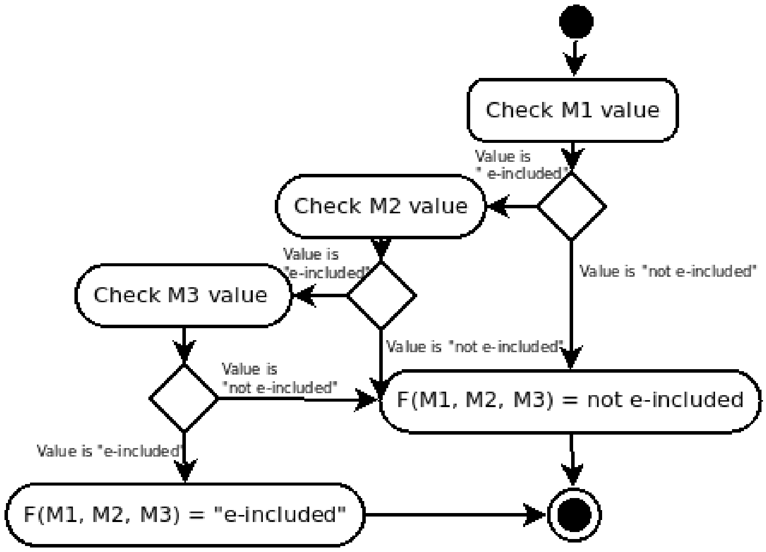

Fig. 4.3. Process of determining the final prediction PROGN based on predictions of Models M1, M2, M3.

Algorithm of e-inclusion prediction process. The algorithm of the prediction process for the e-inclusion of a specific student is given in Fig. 4.4. To make a prediction, it is necessary to know the topic, student data, the pre-defined e-inclusion threshold and to have a trained model with which to make the prediction. If the model predicts that there is a risk of digital exclusion for the student, the risk factors for the particular student are identified, such as the e-learning environment or e-learning materials, or the instructor's ability to share knowledge. The value of the prediction and the risk factors are shown to the instructor so that he can decide on further actions, for example, to change something in the communication with the student, or to offer other study materials, etc.

The level of precision of the prediction is determined as follows. If the model makes a prediction based on Model M1, then according to Subsection 3.6. the level of prediction precision is high. If the model uses a combination of Models M1 and M2 for the prediction, then the level of precision is average. If the prediction is based on a combination of Models M1, M2, M3, and if the student is predicted to be at risk, then the level of precision is low, but if the student is predicted to be not at risk, the level of precision is high.

Algorithm of the process of monitoring the performance of the prediction model. The process of monitoring the performance of the prediction model to determine whether re-training of the model is required is as follows. The database of examples is updated so that it can be used to evaluate the performance of the model and to retrain the model. At a certain time (e.g. 8–10 weeks) after processing the student's data, the system sends the student a question about the real use of digital skills. The answer is recorded in the example database. At predetermined time intervals, the performance of the model is evaluated, if the quality of the model decreases, the predictive model is retrained by system.

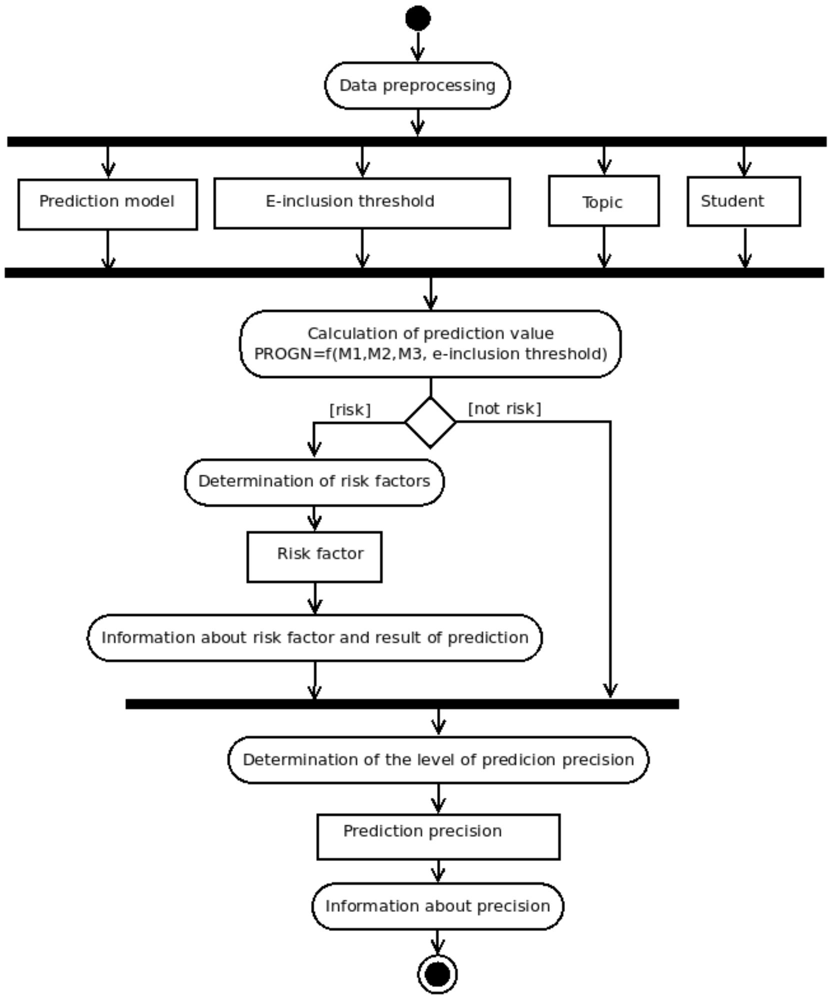

Fig. 4.4. Algorithm of student e-inclusion prediction process.

4.2. Prototype of E-Inclusion Prediction System

The prototype is web-based software using the JAVA programming language and open source software WEKA libraries. The prototype is based on a 3-tier application architecture that consists of a presentation tier, an application tier and a data tier.

The prototype is an early version of the e-inclusion system and consists of the base functionality. The main task of the e-inclusion prediction prototype is to provide functionalities that inform instructors about at-risk students and evaluate the performance of the basic functionality of the e-inclusion prediction system.

The main functionality of the e-inclusion prediction system for the instructor: to set an e-inclusion degree threshold, to search for students, to display prediction results for students (e-included or not e-included); to display factors impacting the prediction result (for example, student motivation, student self-evaluation of learning materials or e-learning environment; download prediction results (Fig.4.5).

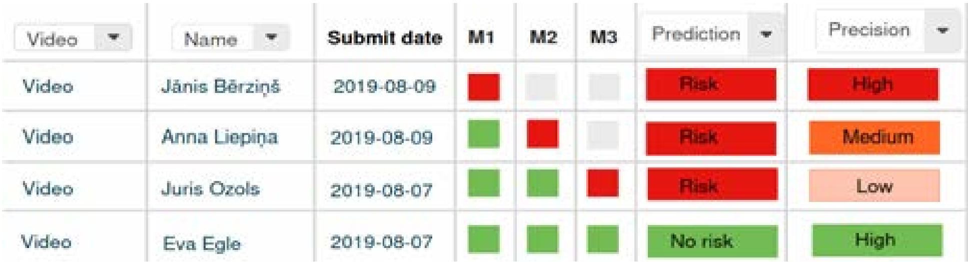

Fig. 4.5. Prototype view of different types of the results predicting the risk to be digitally excluded for learners and presenting the level of precision for the prediction.

The prototype development aims to evaluate the compliance of the basic functionality of the e-inclusion prediction system with the goals set for the system.

4.3. Evaluation of Algorithmic Prediction Model and Prototype

To evaluate the predictive algorithmic model and prototype of e-inclusion, a 160-hour curriculum was used to improve teachers' digital skills. Educators specialized in one of three topics: robotics technology, video technology, mobile technology.

A data set of 65 student records has been prepared for the evaluation of the model and prototype. The data set has not previously been used for predictive model training. To check the correspondence of the model prediction to the real situation, 4 to 6 months after the acquisition of the program, information was obtained from the learners whether they use the newly acquired skills.

Evaluation of the Model Predicting E-Inclusion

The study aims to evaluate the model predicting e-inclusion by comparing the model prediction on the use of digital skills with the real use of newly acquired digital skills for VET teachers.

The evaluation of a model based on a machine learning approach includes its evaluation according to the requirements of the machine learning task (performance and robustness assessment) (see Section 2) and business objectives (Ashmore et al., 2019; Studer et al., 2021).

To determine the performance of the model predicting e-inclusion and whether the model meets its objectives, the following research questions have been studied:

1. What percentage of all VET teachers at-risk of digital exclusion who acquire digital skills can the model predict as belonging to the group at risk?

2. What percentage of predicted VET teachers at-risk of digital exclusion belong to the risk group?

3. To what extent is it possible to predict using model (without significantly lowering the performance indicators of the model) e-inclusion for individuals who acquire digital skills in courses that are different from the courses with which data the predictive model is trained?

Results. The prediction of students who will or will not use the newly acquired skills and the actual number of skills users are shown in the confusion matrix (Table 4.1). As can be seen in the table, the model did not recognize only 3 students (out of 31 students) who are at risk and have not used the newly acquired skills.

Table 4.1

Confusion Matrix of the Model Predicting E-Inclusion: Model Predictions and the Observed Situation Regarding the Use of Newly Acquired Skills

n = 65	Prediction that the newly acquired skills will not be used	Prediction that the newly acquired skills will be used	
Do not really use newly acquired skills	28	3	31
Really use newly acquired skills	13	21	34
	41	24	

The recall of the model predicting the e-inclusion 0.903, which means that the model is able to predict as belonging to the risk group 90.30 % of all teachers of vocational education institutions at risk of digital exclusion who acquire digital skills. The precision of the model is 0.683, which means that in reality, 68.3 % of the educators of vocational education institutions in the digital exclusion risk group belong to the risk group. As the model predicts at-risk students, it can be concluded that the model fulfills its business objectives.

The value of the F2 measure is 0.848 and the accuracy of the model is 0.754. The performance of the model meets the above requirements and can be considered high enough to be used in prediction. The performance indicators of the model correspond to the indicators of student achievement prediction models in various machine learning techniques indicated in the scientific literature.

Evaluation of the Drift of the Model Predicting E-Inclusion

The study aims to evaluate the drift of the e-inclusion prediction model and its constituent models by comparing the performance metrics of training and test data sets. F measure, recall and precision of the model are evaluated.

To evaluate the drift of the model predicting e-inclusion, the following research questions have been studied:

1. By what percentage does the recall of the model and base models change?
2. By what percentage does the precision of the model and base models change?
3. By what percentage does the F measure of the model and the base models change?

F2 measures for training and test data sets were compared. The F2 measure was higher for the training data set, but the difference was small – it decreased by 1.47 %. The F2 measure of the training set is 86.31 %, for the test set – 84.83 %. The difference for the recall is as follows: for the training set the recall value is 95.63 % and for the test set the recall value is 90.30%. The recall of the test set has decreased by 5.33 %. The precision for the test set is 68.30%, but for the training data set it is 62.09 %. In contrast to the recall, the precision rate has increased by 6.21 %.

In the case of a combination of Models M1 and M2, the F2 measure for the training set is 82.35 %, but for the test set – 82.28 %. For the combination of M1 & M2 models, the recall value of the training set is 83.45 %, but for the test set it is 83.87 %. The recall has increased by 0.42 % for the test set. For the combination of M1 & M2 models, the precision for the test set is 76.47 %, but for the training set the precision is 78.23 %, the precision value for the test set has decreased by 1.76 %.

In the case of Model M1, the F2 measure is 79.98 % for the training set and 79.59 % for the test set. The F2 measure has decreased by 0.39 %. The difference for the recall is as follows: for Model M1 the recall for the training set is 79.50 %, but for the test set the recall value is higher – 80.60 %. The recall value for the test set has increased by 1.10 %. The precision of Model M1 for the test set is 75.80 %, but for the training set the precision is 81.80 %. The precision value for the test set has decreased by 6.00 %.

Comparisons of model performance with training and test data sets reveal small differences. The test data resulted in both higher and lower performance compared to the training data. The total model quality indicator F2 measure has reduced its value for the models by 0.07 %, 0.39 %, or 1.47 %, these are small changes that do not significantly affect the model result. Thus, it can be concluded that the model has maintained its quality.

E-Inclusion Prediction Prototype Evaluation

The compliance of the prototype predicting the e-inclusion with the functional requirements is given in Table 4.2.

Table 4.2

Compliance of the Prototype Predicting the E-Inclusion with the Functional Requirements

Requirement	**Evaluation**
Predict students at risk of digital exclusion	Evaluation of prediction results confirms the performance of the model, that is, the ability to predict students at risk.
Ensure the quality of results – maintain performance of prediction model	The model drift evaluation confirms the quality of the model performance.
The interface must be simple and easy to use	Compliance is justified by the fact that the prototype interface provides easy navigation between menus (prediction pages, etc.).

Evaluating the functionality of the prototype, it can be concluded that it meets the requirements.

RESULTS AND CONCLUSIONS

The Doctoral Thesis aimed to develop a model that predicts an individual's e-inclusion in the e-learning environment.

To achieve the goal of the Doctoral Thesis, the following tasks were set.

1. To develop an algorithmic model predicting e-inclusion of individuals:

A. To analyze the scientific literature and other sources related to the e-inclusion processes.

B. To analyze the scientific literature and other sources related to the predictive technologies and methods.

C. To create an algorithmic model predicting e-inclusion of individuals.

2. To create a technological model (prototype) predicting e-inclusion of individuals.

3. To evaluate the technological model predicting the e-inclusion of the teachers of the vocational education institutions.

The following **theoretical results** were obtained by performing the set tasks:

- the theoretical basis for the development of the predictive model of individual e-inclusion has been developed;
- the algorithm and model for predicting the e-inclusion of an individual and the risk factors influencing it have been developed;
- the technological model of e-inclusion has been developed for the prediction of an individual's e-inclusion risk, based on data characterizing the individual.

The following **practical result** was obtained:

- the model predicting an individual's e-inclusion has been developed and evaluated, which allows it to be used for further research in the field of e-inclusion.

Performing the tasks set in the Doctoral Thesis and approbating the obtained results, the following **conclusions** have emerged:

- Digital skills are an essential precondition for the e-inclusion of the individual. Improving individuals' digital skills also contributes to other e-inclusion policy goals. However, access to technology and the existence of digital skills do not guarantee that an individual will use these technologies. Only the meaningful use of digital skills indicates an individual's inclusion.
- E-inclusion is important for everyone, including young people, employees, people planning to change occupations and educators who need technology in their teaching process.
- Learning analytics capabilities contribute to digital education and to ensuring the e-inclusion of the individual in the context of digital skills acquisition.
- The e-inclusion of an individual in the context of digital skills acquisition is influenced by the following factors:
  - student's willingness to learn,
  - student's ability to learn,
  - instructor's willingness to share knowledge,
  - satisfaction with e-learning environment,
  - satisfaction with e-learning materials.
- Individual’s e-inclusion can be predicted by a linear regression model, a set of classifiers, and by using cluster analysis. Combining of e-inclusion prediction models improves the model performance.
- If the goal is to cover as many at-risk students as possible in the prediction, then the classification ensemble method with a majority voting approach combined with a linear regression model or with cluster analysis and linear regression model is appropriate. Such a model can recognize 98.40 % of those who are digitally excluded, but at the same time only 58.20 % of those who are predicted to be digitally excluded are true students at risk. Testing the e-inclusion system prototype, the model was able to predict as belonging to the risk group 90.30 % of all VET teachers in the digital exclusion risk group who acquired digital skills. The precision of the model was 68.3% – so many of the educators at risk of digital exclusion belonged to the risk group.
- If precision and recall are equally important, then the highest performance can be obtained by combining a classifier ensemble model and a cluster analysis model. This

prediction model recognizes 82.80 % of digitally excluded students and can correctly predict 79.60 % of those students who are projected as at-risk students.

- The degree of e-inclusion of an individual affects the recall and precision values of the prediction model. Changes in the e-inclusion threshold in the linear regression model affect the recall and precision values. The higher the e-inclusion threshold, the more at-risk students the model will be able to recognize, but at the same time the model will become less accurate in identifying at-risk students.
- Using the cluster analysis and linear regression model, it is possible to determine the risk factors influencing the e-inclusion of a particular student and their values that correspond to the class of e-included individuals.
- A model for predicting an individual's e-inclusion has been trained with a data set containing data from various digital skills courses. The model can be used for predicting within courses other than the training data set.

The obtained conclusions **confirm the theses**:

1. The e-inclusion of an individual can be predicted using linear regression, cluster analysis, classifiers and artificial intelligence methods.

The obtained results show that the combination of e-inclusion prediction models developed with linear regression, cluster analysis, classifier algorithms improves the model performance indicators.

2. The degree of e-inclusion of an individual can be predicted technologically by the following factors: the level of individual's satisfaction with the e-learning environment and e-learning materials used by the individual to acquire new digital skills, the individual's ability and interest in learning of the new digital skills, and the instructor's willingness to share knowledge.

The obtained results show that the factors differ in the forecasting models. The linear regression model uses the following factors:

- student's willingness to learn,
- student's ability to learn,
- satisfaction with e-learning environment,
- satisfaction with e-learning materials.

The correlation coefficients of the linear regression model indicate that the student's characteristics have a different effect on the prognosis. E-learning materials and student willingness to learn have a greater impact, the e-learning environment and the student's ability to learn have a smaller impact.

A prediction model based on cluster analysis or developed by an ensemble of classifiers uses the following factors:

- student's willingness to learn,
- student's ability to learn,
- satisfaction with e-learning environment,
- satisfaction with e-learning materials,
- instructor's willingness to share knowledge,
- student’s general digital skills.

3. The e-inclusion prediction model can be used for the prediction of the VET teachers’ degree of e-inclusion during the acquisition of digital skills with a model recall of 90.3 % and

F measure of 84.8 %.

Directions for further research:

- Improvement of the individual’s e-inclusion prediction model by obtaining the data from the e-learning system log files.
- Making a prediction as quickly as possible, based on the individual's previous learning outcomes in digital skills courses.
- Improvement of the individual e-inclusion prediction prototype to ensure its robustness.

Novelty, theoretical and practical significance of the Doctoral Thesis:

1. Theoretical and practical aspects of e-inclusion prediction model development have been studied.
2. A technological model of e-inclusion has been developed for the prediction of an individual's e-inclusion risk, based on data characterizing the individual in the process of acquiring digital skills.
3. A model for predicting the inclusion of an individual and the risk factors influencing it has been developed. The model contains a new technology (algorithm) based on linear regressions, cluster analysis, classification methods to determine the individual's e-inclusion risk and the factors influencing it.
4. The result of the Doctoral Thesis is practically usable for instructors in the blended learning course of digital skills acquisition.
5. The result of the Doctoral Thesis is practically usable for information system developers in the development of e-learning systems, student behavior analysis tools, learning analytics tool.

BIBLIOGRAPHY

1. Abad, L. (2014). Media Literacy for Older People facing the Digital Divide: The e-inclusion Programmes Design. Comunicar, 21(42).
2. Achituv, N., Raban, Y., & Soffer, T. (2008). D6.1 & D6.2Policy recommendations for e-inclusion of low socioeconomic status groups (LSG) in e-government services. Retrieved March 10, 2009, from E-government for low socio-economic status groups project website:www.elost.org/D6-2.pdf
3. Aerschot, L. V., & Rodousakis, N. (2008). The link between socio-economic background and Internet use: Barriers faced by low socio-economic status groups and possible solutions. *Innovation: the European journal of social science research*, *21*(4), 317-351.
4. Akhtar, S., Warburton, S., & Xu, W. (2017). The use of an online learning and teaching system for monitoring computer aided design student participation and predicting student success. International Journal of Technology and Design Education, 27(2), 251–270.
5. Alamri A. et al. (2019) Predicting MOOCs Dropout Using Only Two Easily Obtainable Features from the First Week's Activities. In: Coy A., Hayashi Y., Chang M. (eds) Intelligent Tutoring Systems. ITS 2019. Lecture Notes in Computer Science, vol 11528. Springer,
6. Ala-Mutka, K., Malanowski, N., Punie, Y., & Cabrera, M. (2008). Active Ageing and the Potential of ICT for Learning. Institute for Prospective Technological Studies (IPTS) y European Commission. (http://ftp.jrc.es/EURdoc/JRC45209.pdf)
7. Altun, D. (2019). Investigating Pre-Service Early Childhood Education Teachers' Technological Pedagogical Content Knowledge (TPACK) Competencies Regarding Digital Literacy Skills and Their Technology Attitudes and Usage. *Journal of Education and Learning*, *8*(1), 249–263.
8. Amy, H. (2011). The Rural Digital Divide: Exploring Differences in the Health Information Seeking Behaviors of Internet Users. Franklin Business & Law Journal, (2), 65–77.
9. Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. (2012). Quantitative methods for business. Cengage Learning.
10. Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics surveys, 4, 40–79.
11. Ashmore, R., Calinescu, R., & Paterson, C. (2019). Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *arXiv preprint arXiv:1905.04223*.
12. Atallah, R. & Al-Mousa, A. (2019, October). Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method. In *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)* (pp. 1–6). IEEE.
13. Azcona, D., Hsiao, I. H., & Smeaton, A. F. (2019). Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints. *User Modeling and User-Adapted Interaction*, *29*(4), 759–788.
14. Baksa-Haskó G., Baranyai B. (2018). Data-Mining Possibilities in Blended Learning. In: Auer M., Guralnick D., Simonics I. (eds) Teaching and Learning in a Digital

World. ICL 2017. Advances in Intelligent Systems and Computing, vol 716.

15. Basili, C. (2013). Information Literacy Policies from the Perspective of the European Commission. In Worldwide Commonalities and Challenges in Information Literacy Research and Practice (pp. 61–69). Springer International Publishing.
16. Becker, J., Niehaves, B., Bergener B., Räckers, M. (2008). Digital divide in eGovernment: the eInclusion gap model. EGOV 2008. LNCS 5184; 2008. pp. 231–242.
17. Bélanger, F. & Carter, L. (2008). Trust and risk in e-government adoption. *The Journal of Strategic Information Systems*, *17*(2), 165–176.
18. Benda, P. P., Havlíček, Z. Z., Lohr, V. V., & Havránek, M. M. (2011). ICT helps to overcome disabilities. Agris On-Line Papers in Economics & Informatics, 3(4), 63–69.
19. Berkowsky, R. W., Sharit, J., & Czaja, S. J. (2017). Factors predicting decisions about technology adoption among older adults. *Innovation in aging*, *1*(3), igy002.
20. Berral, J. L., Goiri, Í., Nou, R., Julià, F., Guitart, J., Gavaldà, R., & Torres, J. (2010, April). Towards energy-aware scheduling in data centers using machine learning. In *Proceedings of the 1st International Conference on energy-Efficient Computing and Networking* (pp. 215–224).
21. Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M.: The balanced accuracy and its posterior distribution. In: 2010 20th International Conference on Pattern Recognition, pp. 3121-3124. IEEE, (2010, August).
22. Bubenko J. A. (2007) From Information Algebra to Enterprise Modelling and Ontologies — a Historical Perspective on Modelling for Information Systems. In: Krogstie J., Opdahl A. L., Brinkkemper S. (eds) Conceptual Modelling in Information Systems Engineering. Springer, Berlin, Heidelberg
23. Buraimoh, E., Ajoodha, R., & Padayachee, K. (2021, April). Application of Machine Learning Techniques to the Prediction of Student Success. In *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)* (pp. 1–6). IEEE.
24. Caraciolo, M. (2011). Machine Learning with Python – Linear Regression, http://aimotion.blogspot.com/2011/10/machine-learning-with-python-linear.html
25. Casacuberta, D. (2007). Digital inclusion: best practices from eLearning. eLearning papers, 6. Retrieved March 15, 2009, from eLearning papers website: http://www.elearningeuropa.info/files/media/media14197.pdf
26. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321–357.
27. Chena, R.-S., Liu, I.-F. (2013). Research on the effectiveness of information technology in reducing the Rural–Urban Knowledge Divide. Computers & Education, Volume 63, April 2013, Pages 437–445
28. Cios K. J., Kurgan L. A., Hybrid Inductive Machine Learning: An Overview of CLIP Algorithms, In New Learning Paradigms in Soft Computing. 2002, Physica – Verlag GmbH: Heidelberg, Germany. pp. 276–321
29. Cobos, R. & Olmos, L. (2018, December). A learning analytics tool for predictive modeling of dropout and certificate acquisition on MOOCs for professional learning.

In *2018 IEEE international conference on industrial engineering and engineering management (IEEM)* (pp. 1533–1537). IEEE.

30. Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS, in IEEE Transactions on Learning Technologies, vol. 10, no. 1, pp. 17–29, 1 Jan.-March 2017, doi: 10.1109/TLT.2016.2616312.
31. Csordás, A. (2020). Diversifying Effect of Digital Competence. *AGRIS on-line Papers in Economics and Informatics*, *12*(665-2020-1220), 3–13.
32. De Haan, J. (2004). A multifaceted dynamic model of the digital divide. It & Society, 1(7), 66–88.
33. de Hoyos, M., Green, A. E., Barnes, S. A., Behle, H., Baldauf, B., & Owen, D. (2013). Literature Review on Employability, Inclusion and ICT, Report 2.
34. Dewey, J. Interest and effort in education. Houghton Mifflin. Boston; 1913. p. 102. Retrieved: 12.12.2012, URL: http://openlibrary.org/books/OL7141097M/Interest_and_effort_in_education.
35. DiMaggio, P. & Bonikowski, B. (2008). Make money surfing the web? The impact of Internet use on the earnings of US workers. *American Sociological Review*, *73*(2), 227–250.
36. Drabowicz, T. (2014). Gender and digital usage inequality among adolescents: A comparative study of 39 countries. *Computers & Education*, *74*98–111. doi:10.1016/j.compedu.2014.01.016
37. Driessen, M., van Emmerik, J., Fuhri, K., Nygren-Junkin, L., Spotti, M. (2011). ICT Use in L2 Education for Adult Migrants – A qualitative study in the Netherlands and Sweden. Technical Note: JRC59774, http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=4539
38. ECAR-ANALYTICS Working Group. The Predictive Learning Analytics Revolution: Leveraging Learning Data for Student Success. ECAR working group paper. Louisville, CO: ECAR, October 7, 2015.
39. Eiropas Padome (2006). Eiropas Parlamenta un Padomes Ieteikums (2006. gada 18. decembris) par pamatprasmēm mūžizglītībā, OV L 394, 30.12.2006., 10.–18. lpp.
40. Eiropas Padome (2018) Padomes Ieteikums (2018. gada 22. maijs) par pamatkompetencēm mūžizglītībā, OV C 189, 04.06.2018., 1.–13. lpp.
41. Eiropas Revīzijas palāta (2021). ES rīcība nolūkā palielināt digitālās prasmes.
42. Ekubo, E. A. (2020). *Predictive system for characterizing low performance of Undergraduate students using machine learning techniques* (Doctoral dissertation, North-West University (South Africa)).
43. European Commission (2006). Riga Ministerial Declaration on e-inclusion. Riga: European Commission. Retrieved April 7, 2021 from *https://ec.europa.eu/information_society/activities/ict_psp/documents/declaration_riga.pdf*
44. European Commission (2010). A Digital Agenda for Europe: European Commission. Retrieved April 7, 2021
45. European Commission. (2012). Digital Agenda Scoreboard 2011, Pillar 6: Digital Competence in the Digital Agenda https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/digitalliteracy.pdf

46. European Commission (2017). Eurostat. https://ec.europa.eu/eurostat
47. European Commission (2018). Digital Education Action Plan (2018-2020): European Commission. Retrieved April 7, 2021 from *https://eur-lex.europa.eu/legal-content/LV/TXT/PDF/?uri=CELEX:52018DC0022&from=EN*
48. European Commission (2020). Europe Fit Digital Age: European Commission. Retrieved April 7, 2021 from *https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/shaping-europe-digital-future_en*
49. European Commission (2020b). Digital Education Action Plan (2021-2027): European Commission. Retrieved April 7, 2021 from *https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_lv*
50. European Commission (2020c). Eurostat. https://ec.europa.eu/eurostat
51. European Commission (2021). 2030 Digital Compass: the European way for the Digital Decade: European Commission. Retrieved April 7, 2021 from *https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52021DC0118*
52. Farbeh-Tabrizi, K. (2012). Effective Computer Training for People with Disability. Journal Of Applied Computing & Information Technology, 16(1), 1–5.
53. Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., Vuorikari, R. (2016). Research Evidence on the Use of Learning Analytics -Implications for Education Policy. R. Vuorikari, J. Castaño Muñoz(Eds.). Joint Research Centre Science for Policy Report; EUR 28294 EN; doi:10.2791/955210.
54. Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., & Trigg, L. (2009). Weka-a machine learning workbench for data mining. In *Data mining and knowledge discovery handbook* (pp. 1269–1277). Springer, Boston, MA.
55. FreshMinds, & UK Online Centres. (2008). Economic benefits of digital inclusion: buil-ding the evidence. Retrieved October 31, 2015, from http://www.tinderfoundation.org
56. Future Learn. Using Open Data for Digital Business, 2020 https://www.futurelearn.com/courses/open-data-business/4/steps/659724
57. Galván, I. M., Valls, J. M., García, M., & Isasi, P. (2011). A lazy learning approach for building classification models. *International journal of intelligent systems*, *26*(8), 773–786.
58. Gudmundsdottir, G. B., & Vasbø, K. B. (2017, March). Toward improved professional digital competence: The use of blended learning in teacher education in Norway. In *Society for Information Technology & Teacher Education International Conference* (pp. 499-509). Association for the Advancement of Computing in Education (AACE).
59. Guillén-Gámez, F. D., Mayorga-Fernández, M. J., Bravo-Agapito, J., & Escribano-Ortiz, D. (2020a). Analysis of teachers' pedagogical digital competence: Identification of factors predicting their acquisition. *Technology, Knowledge and Learning*, 1–18.
60. Guillen-Gamez, F. D., Mayorga-Fernández, M. J., & Del Moral, M. T. (2020). Comparative research in the digital competence of the pre-service education teacher: face-to-face vs blended education and gender. *Journal of e-Learning and Knowledge*

Society, *16*(3), 1–9.

61. Haight, M., Quan-Haase, A., & Corbett, B. (2014). Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet, level of online activity, and social networking site usage. *Information, Communication & Society*, *17*(4), 503-519. doi:10.1080/1369118X.2014.891633

62. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, *73*, 220–239.

63. Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. Journal of Intelligent Information Systems, 17(2-3), 107–145.

64. Hamerly, G., & Elkan, C. (2003). Learning the k in k-means. In *NIPS* (Vol. 3, pp. 281–288).

65. Hatlevik, O. E., Guðmundsdóttir, G. B., & Loi, M. (2015). Examining factors predicting students' digital competence. *Journal of Information Technology Education: Research*, *14*(14), 123–137.

66. Herrera, F. A. S., Crespo, R. G., Baena, L. R., & Burgos, D. (2019). A solution to manage the full life cycle of learning analytics in a learning management system: Analytic. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, *14*(4), 127-134.

67. Hidalgo, A., Gabaly, S., Morales-Alonso, G., & Urueña, A. (2020). The digital divide in light of sustainable development: An approach through advanced machine learning techniques. *Technological Forecasting and Social Change*, *150*, 119754.

68. Hung, J. L., Shelton, B. E., Yang, J., & Du, X. (2019). Improving predictive modeling for at-risk student identification: a multistage approach. *IEEE Transactions on Learning Technologies*, *12*(2), 148–157.

69. Instefjord, E. J., & Munthe, E. (2017). Educating digitally competent teachers: A study of integration of professional digital competence in teacher education. *Teaching and teacher education*, *67*, 37–45.

70. ITU. (2006). World Telecommunication/ICT Development Report 2006: Measuring ICT for social and economic development, https://www.itu.int/pub/D-IND-WTDR-2006

71. IZM (Izglītības un zinātnes ministrija). (2014). Izglītības attīstības pamatnostādnes 2014.-2020.gadam, http://www.mk.gov.lv/lv/mk/tap/?pid=40305684

72. Jerrim, J. & Sims, S. (2019). The Teaching and Learning International Survey (TALIS) 2018: June 2019.

73. Johansson, L. & Tjäder, C. (2013). IT-Support Direct from Project to a National Service. *Assistive Technology: From Research to Practice: AAATE 2013*, *33*, 399.

74. Khalil Moghaddam, B. & Khatoon-Abadi, A. (2013). Factors affecting ICT adoption among rural users: A case study of ICT Center in Iran. Telecommunications Policy, 37(11), 1083–1094.

75. Kreijns, K., Vermeulen, M., Van Acker, F., & Van Buuren, H. (2014). Predicting teachers' use of digital learning materials: combining self-determination theory and the integrative model of behaviour prediction. *European Journal of Teacher Education*, *37*(4), 465–478.

76. Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). New York: Springer.

77. LAK. 1st International Conference on Learning Analytics, 2011, https://tekri.athabascau.ca/analytics/

78. Lenar, G., Jamila, M., Egor, P., & Rustem, V. (2019, October). Application of Learning Analytics Tools in Learning Management Systems. In *2019 12th International Conference on Developments in eSystems Engineering (DeSE)* (pp. 221–224). IEEE.

79. Lerchner, A., La Camera, G., & Richmond, B. (2007). Knowing without doing. *Nature neuroscience*, *10*(1), 15-17.

80. LR Izglītības un zinātnes ministrija (2021). DigComp 2.1: iedzīvotāju digitālo kompetenču ietvars. Ar astoņiem apguves līmeņiem un lietošanas piemēriem, e-publikācija: LR Izglītības un zinātnes ministrija.

81. Luo, Y., Chen, N., & Han, X. (2020, December). Students' Online Behavior Patterns Impact on Final Grades Prediction in Blended Courses. In *2020 Ninth International Conference of Educational Innovation through Technology (EITT)* (pp. 154–158). IEEE.

82. Lupiañez, F., Codagnone, C., Dalet, R. (2015). ICT for the Employability and Integration of Immigrants in the European Union. Results from a Survey in Three Member States, Luxembourg: Publications Office of the European Union, 2015–237pp.

83. Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, *53*(3), 950–965.

84. Machine Learning and Learning Analytics Workshop. 2014. Learning Analytics & Machine Learning, http://machineanalytics.org/

85. Maennel, K. (2020, September). Learning Analytics Perspective: Evidencing Learning from Digital Datasets in Cybersecurity Exercises. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (pp. 27–36). IEEE.

86. Mahboob, T., Irfan, S., & Karamat, A. (2016, December). A machine learning approach for student assessment in E-learning using Quinlan's C4. 5, Naive Bayes and Random Forest algorithms. In *Multi-Topic Conference (INMIC), 2016 19th International* (pp. 1–8). IEEE.

87. Márquez-Vera, C., Cano, A., Romero, C., & Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied intelligence*, *38*(3), 315–330.

88. Martínez-Alcalá, C. I., Rosales-Lagarde, A., Alonso-Lavernia, M. D. L. Á., Ramírez-Salvador, J. Á., Jiménez-Rodríguez, B., Cepeda-Rebollar, R. M., ... & Agis-Juárez, R. A. (2018). Digital inclusion in older adults: a comparison between face-to-face and blended digital literacy workshops. *Frontiers in ICT*, *5*, 21.

89. Maskey, M., Ramachandran, R., Gurung, I., Freitag, B., Miller, J. J., Ramasubramanian, M., ... & Hain, C. (2019, July). Machine learning lifecycle for earth science application: A practical insight into production deployment. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 10043–10046). IEEE

90. Mason, R., Rennie, F. e-Learning and social networking handbook: resources for higher education; 2008.
91. Miteva, D. & Stefanova, E. (2020). Design of Learning Analytics Tool: The Experts' Eyes View. In *CSEDU (2)* (pp. 307–314).
92. Moncada, I. L. R. Data Literacy and Confidence for Building Learning Analytics Solutions in Higher Education Institutions. A review. *CEUR Workshop Proc.* 2018, 2218, 293–299.
93. Mulyani, E., Hidayah, I., & Fauziati, S. (2019, December). Dropout Prediction Optimization through SMOTE and Ensemble Learning. In *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 516–521). IEEE.
94. Niet, Y. V., Díaz, V. G., & Montenegro, C. E. (2016, September). Academic decision making model for higher education institutions using learning analytics. In *2016 4th International Symposium on Computational and Business Intelligence (ISCBI)* (pp. 27–32). IEEE.
95. Ning Chen; Hoi, S.C.H.; Xiaokui Xiao, "Software process evaluation: A machine learning approach," Automated Software Engineering (ASE), 2011 26th IEEE/ACM International Conference , vol., no., pp. 333, 342, 6–10 Nov. 2011
96. Nissen, M., E. Harnessing knowledge dynamics: Principled organizational knowing & learning; 2006. p. 278.
97. Nistor, N., M. Derntl, and R. Klamma. Learning analytics: trends and issues of the empirical research of the years 2011–2014, in Design for Teaching and Learning in a Networked World, G. Conole et al., Editors. 2015, Springer International Publishing. pp. 453-459.
98. Novaković, J. D., Veljović, A., Ilić, S. S., Papić, Ž., & Milica, T. (2017). Evaluation of classification models in machine learning. *Theory and Applications of Mathematics & Computer Science*, *7*(1), 39–46.
99. Nyce, C. & Cpcu, A. (2007). Predictive analytics white paper. *American Institute for CPCU. Insurance Institute of America*, 9-10.
100. Ono, H. (2005). Digital Inequality in East Asia: Evidence from Japan, South Korea, and Singapore. *Asian Economic Papers*, *4*(3), 116–139.
101. Ortiz-Rojas, M., Maya, R., Jimenez, A., Hilliger, I., & Chiluiza, K. (2019, October). A step by step methodology for software design of a learning analytics tool in Latin America: A case study in Ecuador. In *2019 XIV Latin American Conference on Learning Technologies (LACLO)* (pp. 116–122). IEEE.
102. Osamor, V., Adebiyi, E., Oyelade, J., & Doumbia, S. (2012). Reducing the time requirement of k-means algorithm. *Plos One*, *7*(12), e49946. doi:10.1371/journal.pone.0049946
103. Oye, N. D., Salleh, M. M., Iahad, N. A. E-Learning Barriers and Solutions to Knowledge Management and Transfer. Information Management & Business Review. 3(6); 2011. p. 366–372.
104. Patmanthara, S., & Hidayat, W. N. (2018, June). Improving vocational high school students digital literacy skill through blended learning model. In *Journal of Physics: Conference Series* (Vol. 1028, No. 1, p. 012076). IOP Publishing.
105. Pereira F. D. et al. (2019) Early Dropout Prediction for Programming Courses

Supported by Online Judges. In: Isotani S., Millán E., Ogan A., Hastings P., McLaren B., Luckin R. (eds) Artificial Intelligence in Education. AIED 2019. Lecture Notes in Computer Science, vol. 11626. Springer, Cham

106. Pfeffer, J. & Sutton, R. I. (1999). Knowing "what" to do is not enough: Turning knowledge into action. *California management review*, *42*(1), 83–108.

107. Powell, K. C., Kalina, C. J. Cognitive and Social Constructivism: Developing Tools for an Effective Classroom. Education. 130(2); 2009. pp. 241–250.

108. Prokofyeva, N., Zavjalova, O., & Boltunova, V. (2019, May). Feedback Method in Lecturer-Student Interaction. In *Proceedings of the International Scientific Conference. Volume I* (Vol. 442, p. 448).

109. Quintana, M. G. B. & Zambrano, E. P. (2014). E-mentoring: The effects on pedagogical training of rural teachers with complex geographical accesses. Computers in Human Behavior, 30, 629–636.

110. Rana, N. P., Dwivedi, Y. K., & Williams, M. D. (2013). Analysing challenges, barriers and CSF of e-gov adoption. Transforming Government: People, Process and Policy, 7(2), 177–198. Retrieved from www.scopus.com

111. Rapaport, R. (2009). The new literacy: Scenes from the digital divide 2.0. RetrievedOctober 31, 2015, from http://www.edutopia.org/digital-generation-divide-literacy

112. Ribeiro De Carvalho Martinho, V., Nunes, C., Minussi, C. R., "An Intelligent System for Prediction of School Dropout Risk Group in Higher Education Classroom Based on Artificial Neural Networks," Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference, pp. 159, 166, 4–6 Nov. 2013

113. Rintamäki, K. & Lehto, A. (2018). A digital information literacy course for university teachers: challenges and possibilities.

114. Robles, J. M. & Torres-Albero, C. (2012). Digital Divide and the Information and Communication Society in Spain. Sociologija i prostor/Sociology & Space, 50(3).

115. Romero, C. & Ventura, S. (2019). Guest Editorial: Special Issue on Early Prediction and Supporting of Learning Performance, in *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 145–147, 1 April-June 2019, doi: 10.1109/TLT.2019.2908106

116. Romero, C. & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355.

117. Rudin, C. (2014). Prediction: Machine Learning and Statistics, (MIT OpenCourseWare: Massachusetts Institute of Technology), http://ocw.mit.edu/courses/sloan-school-of-management/15-097-prediction-machine-learning-and-statistics-spring-2012 (Accessed 2 Jun, 2014). License: Creative Commons BY-NC-SA

118. Samuelson, P. (2003). Mapping the digital public domain: Threats and opportunities. *Law & Contemp. Probs.*, *66*, 147.

119. Santos, R., Azevedo, J., & Pedro, L. (2013). Digital Divide in Higher Education Students' Digital Literacy. In Worldwide Commonalities and Challenges in Information Literacy Research and Practice (pp. 178–183). Springer International Publishing.

120. Sanz, E. & Turlea, G. (2012). Downloading inclusion: a statistical analysis of young people's digital communication inequalities. *Innovation: The European Journal off Social Sciences*, *25*(3), 337-353. doi:10.1080/13511610.2012.699652

121. Scheerder, A., van Deursen, A., & van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and informatics*, *34*(8), 1607–1624.

122. Schuh, G., Scholz, P., Leich, T., & May, R. (2020, October). Identifying and Analyzing Data Model Requirements and Technology Potentials of Machine Learning Systems in the Manufacturing Industry of the Future. In *2020 61st International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)* (pp. 1–10). IEEE.

123. Seliya, N., Khoshgoftaar, T.M., and Hulse, J. V. (2009): A study on the relationships of classifier performance metrics. In: 21st IEEE International Conference on Tools with Artificial Intelligence, pp. 59–66. Newark, NJ, Seville: JRC-IPTS. Retrieved from http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=4699.

124. Shearer C., *The CRISP-DM model: the new blueprint for data mining*, J Data Warehousing (2000); 5:13–22

125. Skujiņa, V. Pedagoģijas terminu skaidrojošā vārdnīca. Termini latviešu, angļu, vācu, krievu valodā. Sast. I. Beļickis, D. Blūma, T. Koķe, D. Markus, V. Skujiņa (vad.), A. Šalme. – Rīga: "Zvaigzne ABC", 2000. 248 lpp.

126. Snee, R. D. (1977). Validation of regression models: methods and examples. Technometrics, 19(4), 415–428.

127. Soja, E., Soja, P., Kolkowska, E., & Kirikova, M. (2019, September). Supporting active and healthy ageing by ICT solutions: preliminary lessons learned from Polish, Swedish and Latvian older adults. In *EuroSymposium on Systems Analysis and Design* (pp. 48–61). Springer, Cham.

128. Sorour, S. E., Mine, T., Goda, K., & Hirokawa, S. (2014, April). Efficiency of LSA and K-means in Predicting Students' Academic Performance Based on Their Comments Data. In *CSEDU (1)* (pp. 63–74).

129. Stirna, J., & Persson, A. (2018). Enterprise modeling. *Cham: Springer*.

130. Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K. R. (2021). Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology. *Machine Learning and Knowledge Extraction*, *3*(2), 392–413.

131. Subramaniam, P., R. Motivational Effects of Interest on Student Engagement and Learning in Physical Education: A Review. Int. Journal Phys Educ. Vol. 46(2); 2009. pp. 11–19. Retrieved: 12.12.2012

132. Sundqvist, K., Korhonen, J., & Eklund, G. (2020). Predicting Finnish subject-teachers' ICT use in Home Economics based on teacher-and school-level factors. *Education Inquiry*, 1–21.

133. Suresh A., Sushma Rao H.S., Hegde V. (2017) Academic Dashboard –Prediction of Institutional Student Dropout Numbers Using a Naïve Bayesian Algorithm. In: Vishwakarma H., Akashe S. (eds) Computing and Network Sustainability. Lecture Notes in Networks and Systems, vol 12. Springer, Singapore

134. Tilde. (2014). Letonika, letonika.lv

135. Townsend, L., Sathiaseelan, A., Fairhurst, G., & Wallace, C. (2013). Enhanced broadband access as a solution to the social and economic problems of the rural digital divide. Local Economy, 28(6), 580–595.

136. van Deursen, A. J. & Van Dijk, J. A. (2009). Using the Internet: Skill related problems in users' online behavior. *Interacting with computers*, *21*(5–6), 393–402.

137. van Deursen, A. J. & van Dijk, J. A. (2015). Internet skill levels increase, but gaps widen: A longitudinal cross-sectional analysis (2010–2013) among the Dutch population. *Information, Communication & Society*, *18*(7), 782–797.

138. van Dijk, J. A. (2006). Digital divide research, achievements and shortcomings. *Poetics*, *34*(4–5), 221–235.

139. Vedins, I. Macīšanās māksla. In Latvian. Riga: Avots; 2011.

140. Verhoeven, J. C., Heerwegh, D., & De Wit, K. (2020). Predicting ICT skills and ICT use of University students. *Encyclopedia of Education and Information Technologies*, 1286–1304.

141. Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, *89*, 98–110.

142. Xu, Z., Yuan, H., & Liu, Q. (2020). Student Performance Prediction Based on Blended Learning. *IEEE Transactions on Education*.

143. Yadav, S. & Shukla, S. (2016, February). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In *2016 IEEE 6th International conference on advanced computing (IACC)* (pp. 78–83). IEEE.

144. Yu, B., Ndumu, A., Mon, L. M., & Fan, Z. (2018). e-inclusion or digital divide: an integrated model of digital inequality. Journal of Documentation.

145. Yu, T. K., Lin, M. L., & Liao, Y. K. (2017). Understanding factors influencing information communication technology adoption behavior: The moderators of information literacy and digital skills. *Computers in Human Behavior*, *71*, 196–208.

146. Záhorec, J., Hašková, A., & Munk, M. (2019). Teachers' Professional Digital Literacy Skills and Their Upgrade. *European Journal of Contemporary Education*, *8*(2), 378–393.

147. Zhuge, H. The Knowledge Grid; 2004.

148. Zillien, N. & Hargittai, E. (2009). Digital distinction: Status-specific types of internet usage. *Social Science Quarterly*, *90*(2), 274–291.

Ieva Vītoliņa was born in 1970, in Ventspils. She graduated with an applied mathematics degree from the University of Latvia in 1993, a Master's degree in Social Sciences in 1997, and a Master's degree in Computer Science in 2002. Since 2008, she has been a researcher with the Distance Education Study Centre of Riga Technical University. Her research interests are in the field of e-inclusion and e-learning.