

Educational Data Mining Methods for the Analysis of Student's Digital Footprint*

Evgenia Baranova
ev_baranova@mail.ru

German Shvetsov
shvetzoff.german@yandex.ru

Tatiana Noskova
noskovatn@gmail.com

Herzen State Pedagogical University of Russia,
St. Petersburg, Russian Federation

Abstract

The collection and analysis of a student's digital footprint is an integral part of the development of education in the Russian Federation in the context of digitalization of the economy. Among other things, such practice is associated with the assessment of the formation of students' competencies in blended learning. The study aims to develop approaches, methods and tools for analyzing data generated in the digital education environment (DEE), in order to identify correlations between the structure and content of educational programs and students' performance.

Educational data mining (EDM) methods as well as methods of design and development of information systems and databases are relied upon. The following methods of data processing were used: correlation analysis and distillation of data for human judgement.

The experimental base of the study is the DEE of the Herzen State Pedagogical University of Russia, which ensures the formation of data on educational programs, individual learning paths of students and their successful development. A large amount of detailed data collected for over 10 years on various aspects of the educational process was used for the analysis.

As a result, the authors have developed a tool for collecting and analyzing the digital footprint of students in the context of blended learning using EDM methods. In accordance with these methods, calculations were made that allow to draw preliminary conclusions about the nature of relationships between the selected indicators (the number of activities on the course, the participation of students in the course chat(s), the final assessment for the course, the time spent on the course) obtained as part of the students' work in the Moodle distance learning system (Moodle DLS) and the results of midterm assessment. For example, the value of relationships between the indicator of the final assessment for online education courses (OEC) and results of the midterm assessment is 0.7108, which indicates a strong correlation. The value of relationships between the number of activities on the course and the grade for the midterm assessment (0.314) shows a weak correlation. On the basis of such calculations, it is possible to formulate assumptions on how particular activities may help in mastering various disciplines and on the OEC effectiveness, as well as prepare recommendations for updating educational resources in order to improve their quality. The experience, approaches, and results presented by the authors demonstrate the feasibility of EDM methods for the analysis of educational process.

Keywords: *student's digital footprint, data mining, educational data mining, big data, online learning, digital education environment*

*Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1 Introduction

Digitalization of the economy presents a significant challenge for the entire education system. Digital economy requires the education system not just to “digitize” individual processes, but to create an integrated approach that would set new goals, change the structure and content of the educational process.

An effective education involves not only the transfer of information from professor to student, it also requires complex social interactions and adaptation to the needs of each student and their cultural and social context. Digital transformation changes the content, forms and methods of education. “Digitization” primarily affects school and university education systems. AI-driven mass online courses, chatbots and lesson plans are just a few examples of digital transformation in higher education [Kurbatsky, 2019].

Rapid introduction of information technologies in various spheres of our life, including education, leads to a partial or complete transfer of human activity to the virtual environment. Information posted by users on their pages in the digital environment can be considered as a product of activity that allows for identifying the individuality of a particular person, making assumptions about their psychological characteristics and cognitive abilities [Tulupyeva et al., 2015].

The continuous digital transformation of society makes its own adjustments to traditional learning models, as well as to online education models [Shamsutdinova, 2020]. Building student digital profiles based on analysis of the so-called digital footprint is one of the recent trends. In the latest studies, the concept of digital footprint of a student is interpreted in different ways, but there is a general perspective — a transition from a narrow, technical understanding of the footprint as a set of IP addresses and reports on the network activity of students to a broader consideration.

According to O. P. Zhigalova: “Digital footprints considered as the result of educational and professional activities in a digital format represent data that allow for determining the level of competencies, forming a learning path, evaluating opportunities and strategies for further improvement and professional development in a particular field.” [Zhigalova, 2019] According to V. N. Kurbatsky: “The digital footprint is an array of data on the results of a student’s educational and project activities, including all the materials that the student creates: presentations, prototypes, audio and video recordings, road maps, etc.” [Kurbatsky, 2019]

The prevalence and influence of social networks on the digital image of students is growing. In the works of S. A. Vartanov [Vartanov, 2018]; G. A. Nikolaenko [Nikolaenko, 2019]; T. V. Tulupyeva [Tulupyeva et al., 2016] and others, we can find that such information can be very different and include hundreds of footprints: from photos and videos to comments, likes, reposts and other virtual activities.

The digital footprint of a student starts from the moment of entering the university and contains personal data, information about admission, study department, educational program, academic performance during the entire period of study, as well as educational analytical data collected automatically when students use distance learning systems and open education platforms for online courses.

The digital footprint in the educational space is an alienated result of a person’s educational activity. The entire educational space that has a digital footprint must be created with the use of pedagogical design.

The methods of processing information about the educational process, in particular the digital footprint of its subjects, include educational data mining (EDM) methods, which provide extensive opportunities for analyzing and interpreting the results obtained. However, despite the obvious advantages of the EDM methods, research related to their use is clearly insufficient for simulating higher educational processes.

There are no developed methodology, tools for data analysis and data interpretation based on these methods and designed for:

- identification of critical points in the process of mastering educational programs, disciplines and modules that cause the greatest difficulties for students,
- determining the relationships and mutual influence of traditional and distance learning components on the success of students in mastering educational programs,
- early identification of students in the group of risk who can be expelled, etc.

The study targets the development of methods and tools for analyzing data generated in the digital education environment (DEE) to identify links between the structure and content of educational programs and results of students' learning activities in order to optimize the educational process.

2 Literature analysis

In the Russian and foreign literature data mining (DM) is interpreted as a collective name for a set of methods, that is also referred to as “information extraction”, “data excavation”, “knowledge extraction”, “template analysis”, “knowledge discovery in databases” in other papers and books [Ovsyanitskaya, 2013].

DM approaches are based on classification, clustering, simulation and forecasts, decision tree construction, evolutionary programming, and fuzzy logic. Initially, the DM methods were based on research in the field of applied statistics. To date, the development of such methods is significantly associated with research of artificial intelligence (pattern recognition, search for knowledge representation models, neural networks, genetic algorithms), papers on information visualization, machine learning, etc. The DM technology is supported by the concept of patterns which describe regularities expressed in human-readable forms.

The modern research related to the application of methods of data mining, machine learning and statistics on various types of information about the educational process is called educational data mining (EDM) [Belonozhko et al., 2017][Zorić, et al., 2020]. Within the framework of this field of study, models and methods of educational data processing aimed at improving the educational process are developed. EDM specifies the source area of large data sets (Big Data, hereinafter referred to as the BD) in relation to the educational process; aims to find patterns (samples, templates, schemes, regularities) that are characteristic of this subject area. A detailed overview of the key EDM components was presented in [Peña-Ayala, 2014]. A. Peña-Ayala identified three categories of main research modules related to EDM: problems, methods, and algorithms. The use of EDM is also considered in studies conducted to understand how students interact with educational technologies [Angeli et al., 2017].

The paper [Manyika et al., 2011] states that EDM methods make it possible to build models for describing various aspects of the educational process in order to identify factors that affect the success of mastering of learning paths by students, including teaching, administrative and organizational activities. The paper [Castro, 2007] suggests that clustering and visualization methods can be applied to analyze educational data and thus improve the process of electronic learning. Following the study [Zamora-Musa & Velez, 2017] with the use of EDM, association rule method, based on the data obtained from the survey, connections between the categorical variables of the survey (environmental efficiency, environmental influence, environmental utility, control of the environment, ease of learning in the environment) are identified to further improve the immersive environment (augmented reality technology, provide the effect of full or partial presence in an alternative space). In the study [Razaque & Alajlan, 2020], the influence of external factors (relationships, alcohol consumption, parents' education, frequency of meetings with friends, place of residence) on student academic performance is determined using EDM methods.

According to V. A. Larionova and A. A. Karasik, data mining allows to study the behavioral pattern of students; to learn how to take into account individual characteristics of each in the formation

of educational paths [Larionova & Karasik, 2019].

The study of A. Abdulmohsen suggests that the main data sources used in EDM nowadays include the following categories:

1. Traditional learning, when the knowledge is passed on to students through personal contact. Data can be collected by traditional methods, such as observation and questionnaires, which allow to study cognitive skills of students and determine the success of their learning. The statistical method and psychometry can be applied to the data.
2. Online learning and education management systems (EMS) provide students with materials, instructions, communication, and reporting tools that enable them to learn independently. Data mining techniques can be applied to data stored in databases.
3. Systems of Intelligent learning and adaptive educational hypermedia provide content to students based on their profiles, so it is necessary to use data mining methods to analyze the created user profiles.

Specifically, the authors of the following researches [Ezekiel & Mogorosi, 2020], [Rawad & Rémi, 2020] considered the use of data mining methods to identify relationships between the interaction of students with the DLS elements (files, tests, forums) and their final academic performance, as well as relationships between the data of students of the mass open online course (MOOC) — age, gender, geographical region, education, socio-professional status, and their academic performance.

Our literature analysis: [Abdulmohsen, 2016], [Baker & Yacef, 2009], [Bowers et al., 2012], [Bowers, 2010], [Shrestha & Pokharel, 2021], allows to identify the following areas of use of EDM methods in educational practice, the most important for higher education:

- modeling the behavior of students in the learning process in order to predict the development of their cognitive abilities, to identify the contingent of students at risk, with a high probability of being expelled;
- development of new models and ways of presenting knowledge in the subject area that would correspond to diverse learning styles and cognitive capabilities identified in students using EDM methods;
- studying the processes of interaction of students with the digital environment, identifying the components of the environment, studying the effects that the educational environment has on the results of learning.

3 Research Methods

Within the framework of this study, the focus is on the components of digital footprint associated with the development of learning paths by students.

The target of the study is the components of digital footprint, represented by midterm academic performance of students in the form of grades obtained during the study of disciplines, practical studies and modules, as well as results of students' educational activities in Moodle DLS in the process of mastering the educational program and their correlation. The research is carried out with a view to identifying critical stages for students and optimization of the structure, content and methods of educational programs implementation in high education.

Research tools include information systems and databases for generating and collecting data, and EDM methods for analyzing, processing, and measuring the data.

The method of analyzing the student digital footprint developed by the authors is based on the formulated in [Lunkov & Kharlamov, 2014] stages of EDM methods application for data analysis

and consists of the content of the stages specified in relation to the subject area in accordance with the study's aim, the characteristics of means and tools for obtaining, processing, analyzing data and presenting results.

The first stage involves setting an objective and identifying entities of the subject area under consideration. The need to integrate various technologies of distance learning and traditional forms of implementation of the educational process is dictated by the modern world. In this environment, analyzing and identifying the relationships of student performance indicators in the implementation of various forms of education becomes relevant. Let us highlight the main information sources of data:

- an educational program, which is a set of key features of education (volume, content, targets), organizational and pedagogical conditions and forms of assessment, presented as a curriculum, a calendar curriculum, work programs of academic disciplines, courses, disciplines (modules) and other components, including evaluation and methodological materials;
- the activity of the faculty, including development of teaching materials (lectures, practical classes, tests), including in a digital format, assessment of students' current and midterm academic performance, etc.;
- students' learning activity, which is evaluated in the course of current, midterm and final academic performance.

Specification of the relationships between these information sources of the educational process allowed to develop a chart of information flows, shown in Figure 1.

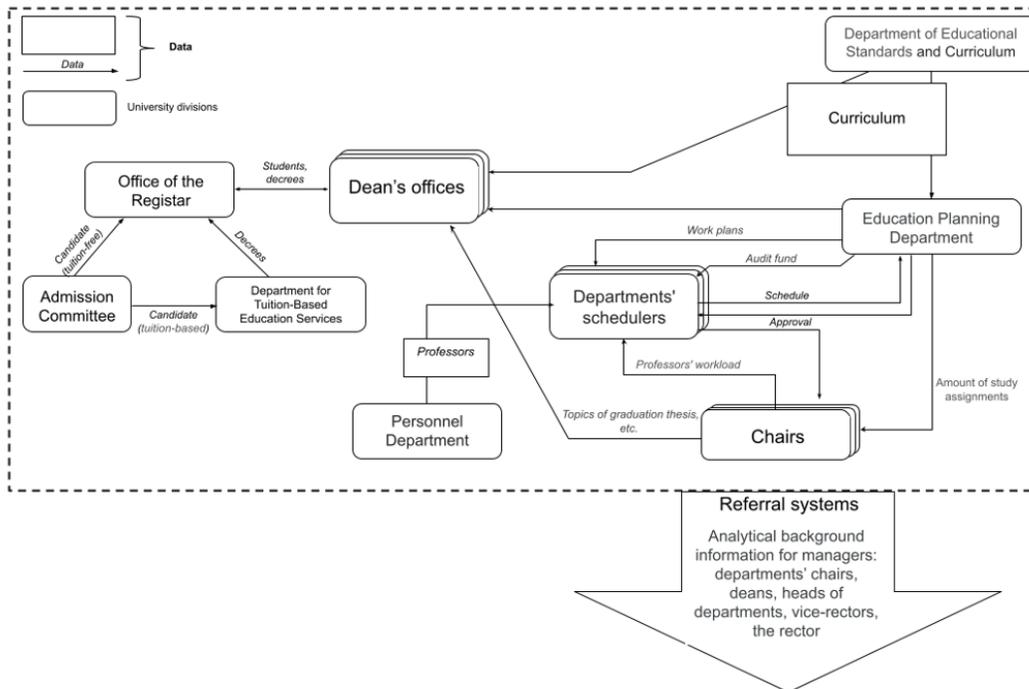


Figure 1: Information Flow Chart

The second stage involves the preparation of data for EDM and it is based on the methods of designing and developing information systems [Pedersen, 2019] and databases [Campbell & Majors, 2017]. At this stage, data modeling is carried out, i.e. the definition and analysis of data requirements that are necessary for the implementation of EDM, and the development of structures for operational, reference and archival databases (DB) to represent data sources generated using internal and external information systems.

To implement the information flow chart presented in Figure 1, the Herzen State Pedagogical University has developed and rolled out an integrated educational management information system (EMIS) in the educational process <http://oio.herzen.edu.ru>. The EMIS is integrated with various DEE components, including Moodle DLS. It includes a set of information systems (IS) and web resources based on the distributed database HERZEN, which provides storage of complete information about various aspects of the educational process, including

- data on the main professional educational programs carried out;
- academic performance of students;
- information about the academic load of faculties;
- schedules of classes and exams;
- data on e-learning courses conducted by teachers used in the implementation of educational programs (disciplines, modules, practical classes), etc.

Data is formed in the database for subsequent analysis by means of certain tools (information systems) – the EMIS components, developed together with the authors. IS “Training and Work Plans” is the tool for generating data on implemented main professional educational programs. It provides the creation of electronic curricula of educational programs in accordance with the requirements of the Federal State Educational Standard of Higher Professional Education and the Federal State Educational Standard of Higher Education, the automatic formation of work plans and the volume of training assignments of departments for all levels of education and forms of training.

To collect data on the midterm academic performance of students, the IS “Dean’s Office” is used. It provides the formation of students’ personal data, information about characteristics of the student’s educational program, information about the structure and composition of the educational program, and data on individual learning paths as well as the success of their mastering [Baranova et al., 2020]. Academic load is distributed between teachers using the IS tool “Department Load”.

A system designed to create an electronic schedule of classes and exams for students in line with work plans and distributed load of teachers is the IS “Schedule”. The generated schedule is checked for correctness: compliance with the restrictions on the number of training hours per day for the group and the teacher; the teacher can not be assigned classes at the same time with different groups; the classroom can not be assigned to conduct different classes at the same time; lectures should be conducted only by teachers with a degree, etc.

All these tools allows for generating information in databases about the core vocational education programs (CVEP), students, the workload of the faculty for data analysis using EDM methods.

Nowadays, electronic learning and distance learning technologies are widely used. To obtain complete and diverse information about the educational process for EDM, it is necessary to generate data on the content and results of electronic learning in DLS. Today, one of the most popular systems for supporting the educational process in the context of distance education is Moodle DLS. The platform’s most important advantages that make it popular are openness, mobility, portability, extensibility, widespread use, etc.

Moodle DLS contains a wide set of tools for creating and maintaining electronic learning courses, various opportunities for presenting educational materials, monitoring the progress, ensuring the

feedback between teachers and students, which allows students to post their work in DLS, receive reviews and teacher's advice [Zaitseva, 2012].

Training courses, simple in structure, contain educational and methodological materials for the development of the discipline in the form of presentations, lecture texts, lists of literature, exam questions, etc. Most of the courses published in the university's Moodle DLS involve a wider use of various DLS tools and include, as a rule, a working program of the discipline; educational and methodological materials for mastering the discipline; methodological recommendations on studying the e-learning course and preparing for various types of classes and assessment for students; a forum for communication and discussion of questions on the course for students and the teacher; assessment funds for monitoring students' knowledge.

Moodle DLS, based on MariaDB DBMS (a branch of MySQL DBMS), provides storage of various data on students' interaction with the digital environment during the learning process, including current academic performance. Moodle database query engine allows getting various data sets, including

- general information about the course, teachers, and students;
- the amount of time spent by the student in the online course;
- the number of changes made by students in the various modules of the course;
- the results of completed tasks by students in the course;
- information about student activity on course forums, etc.

The tools that provide a link between the data sources presented in EMIS and Moodle DLS include IS "Electronic Monitoring" and "Electronic Atlas" web resource — components for monitoring the activities of the faculty related to online learning.

The systems provide an opportunity for the teacher to form an application for the creation of an e-learning course in Moodle DLS, as well as the ability to process such an application. Such resources allow organizing a link between the data on disciplines, courses in Moodle DLS and the teachers who implement education. The list of disciplines and the teacher's e-learning course are displayed on their personal page in the "Electronic Atlas" (Fig. 2).

This connection allows establishing a correspondence between the body (students, faculty) and their activities in Moodle DLS, provides an opportunity to obtain more detailed information about the structure of specific courses.

The third stage includes the collection and analysis of the final data, the definition of EDM methods for their processing and visualization of the results. Correlation analysis and the method of Distillation of Data for Human Judgment were used as EDM methods to identify the relationships between the entities of the subject area.

Disciplines of the core curriculum read by professor in the 2019–2020 academic year	Form of education	Study year	Semester	Request for the opening of OEC
Disciplines (modules) for selection. Design and Development of Online Education Resources. Education and Pedagogical Science — Theory and Methods of Education and Mentoring (Information Technologies)	Classroom learning	3	1	<u>OEC is available</u>
<u>Module: Content of the Subject. Algorithms and Methods of Programming.</u> Pedagogical Education — Information Technologies in Education	Classroom learning	1	1	<u>OEC is available</u>

Figure 2: Web resource “Electronic Atlas”. The section “Teacher’s Disciplines”

The term “distillation of data” was introduced by D. Siemens and R. Baker and involves data clustering, i.e. automatic division of elements into groups based on the “similarity” of a set of characteristics in order to reduce the dimension of the original data set, and data visualization, i.e. representation in a form that promotes human perception [Siemens & Baker, 2012].

At this stage, it is planned to use tools designed to obtain summary final data for analyzing the quality of the educational program as a whole according to various criteria, such as the success of the program mastering by students, the qualification characteristics of the faculty, the most difficult components of the program to master, the demand for electronic educational courses in disciplines, etc.

The web resource “Digital Passport CVEP” developed by the authors allows obtaining quantitative values of the indicators of the above criteria for analyzing the structure, composition, and various aspects of the implementation of educational programs. For example, to generate for each educational program:

- data on the structure of the personnel implementing the program: academic degrees and positions of teachers, participation of teachers in the implementation of the CVEP (disciplines/modules, practical classes, management of the thesis and SFE), departments’ participation in the implementation of the CVEP, teacher activity in Moodle DLS (number of hours, structure of developed courses), etc.
- data on the structure of the student body: the percentage of foreign citizens, the share of students on a contractual basis, the movement of the student body (admission, expulsion, graduation), employment of graduates [Shrestha & Pokharel, 2021] (availability of work, job in one’s degree field, employment according to the field of activity);

- summary data on student attendance and academic performance, student ratings in the group, learning activities in Moodle DLS, etc.

For visual representation, the generated data can be aggregated by the selected attributes: form of study, level of education, field of study/major, orientation (scope of education), on a budgetary/contractual basis, etc.

The method proposed by the authors allows determining the nature of the relationships between the selected entities of the subject area, analyzing the student's digital footprint in order to make changes to the content of educational programs and predicting the success of students' mastering of learning paths.

4 Results

The authors developed the tool named "Digital Passport of Core Vocational Education Program" (DP CVEP). This tool provides access to analytical data on the CVEP and its implementation, allows to track the dynamics of student expulsions, and to identify the categories of students who are at risk of being expelled, educational programs, and training areas that are difficult to master at early stages.

The tool allows to automatically cluster data on the CVEP (faculty, student body) by certain characteristics (staff of the faculty, academic degrees, workload by type of contract, summary student performance, current students) and visualize them. The resource is a component of EMIS, created on the basis of modern technologies for developing web solutions [Stauffer, 2019], currently includes a set of reports for the university management, which allows to generate and visualize analytical data on the implementation of CVEP, teaching staff, student body, and class attendance.

To assess the intensity of students' work in Moodle DLS, indicators were identified, for example, the number of actions performed with the course elements, including the number of clicks on the course pages, loading resources, clicking on links, sending messages in forums and chats. Highlighted indicators also included: the activity of students in the chat(s) of the course (the number of messages sent in the chat(s)); the final grade for the course (formed on the basis of the points received when completing tasks or tests); the time spent on the course by students.

Next, a number of disciplines were selected that play a key role in training for various educational programs, the development of which is based on the extensive use of e-learning course in Moodle DLS: geometry, organization theory and management of organizational policy in education, plant systematics, ancient world history, protection and preservation of cultural heritage, etc.

Information about the success of students' mastering of the educational program is built in the form of grades for midterm assessment in modules, disciplines and practical classes in the Herzen database. To connect data from two databases, an integration model was developed (Figure 3) and implemented in the DP CVEP.

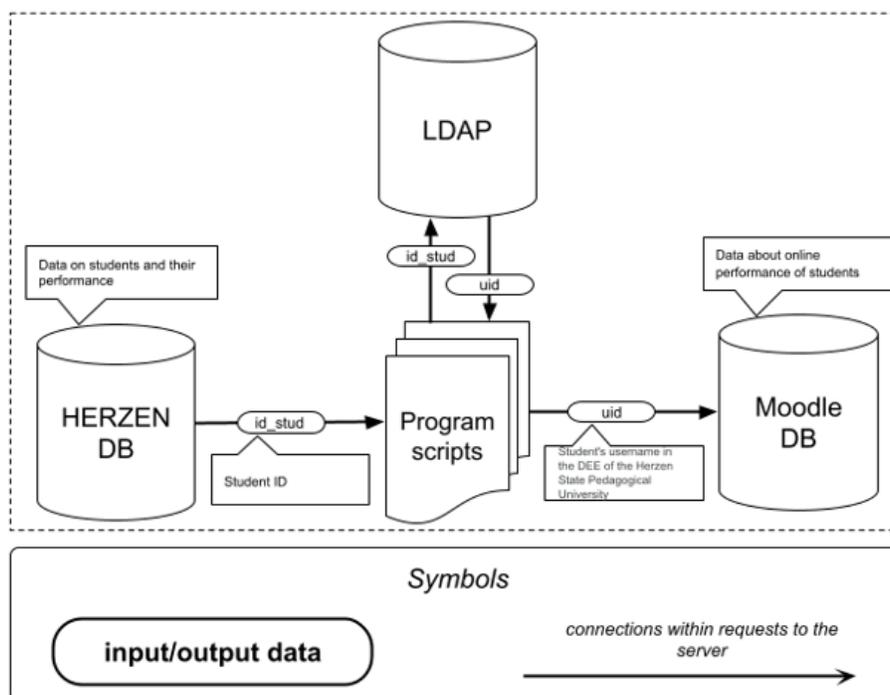


Figure 3: Model for integration of HERZEN DB and Moodle DB

The tool allows identifying the correlation between the selected indicators of students' performance in Moodle DLS and the successful results of their training. The calculation of the indicators' values is automatic and possible thanks to the queries to MariaDB. The DP CVEP tool implementation enabled us to obtain the data on the training of five groups, each containing 20 students on average. Correlation coefficients were calculated for each discipline, student group, and selected indicator, reflecting the dependence of the indicator values on students' grades for the disciplines midterm assessments (Table 1).

Indicators of the number of students' actions and their activity in the chat(s) of the course characterize the additional activity of students on the way to achieve the final result. The final grade for the course is formed on the basis of the quality of the student's task performance throughout the completion of the online course.

The calculations presented in Table 1 are part of the results of the study. Although these results are based on a relatively small amount of data, they make it possible to formulate assumptions that, in the future study, will be tested on large amounts of data by EDM methods.

For instance, based on the data obtained, one can assume that the values of the third indicator have the most consistent high correlation, with an average value of 0.7108. Thus, the availability of tools providing an evaluation of the current performance of students in the OEC (assignments, tests) with substantial certainty allows predicting the success of the student's mastering of the discipline.

At the same time, the first indicator has the smallest correlation, with an average value of 0.314. It is possible to draw a preliminary conclusion that this indicator doesn't present a significant factor for characterizing the quality of student's learning process in the course.

Table 1: Correlation r-coefficients for selected indicators

Discipline	Number	Value of r for indicator			
		of student in the group	Number of course activities	Students' activity in chat(s)	Final course grade
Geometry	20	0.2582	0.051	0.4014	0.5269
Organization theory and organizational policy management in education	27	0.6349	0.0308	0.917	0.5615
Plant Systematics	16	0.1388	0.4070	0.9369	0.3882
History of the Ancient World	29	0.1935	0.0494	0.6303	0.4290
Protection and Preservation of Cultural Heritage	27	0.3447	0.5205	0.6959	0.4753

The analysis of data in the context of disciplines showed significant differences in the values of indicators, related, according to the authors, to the specificity of the subject area and the structure of the OEC of disciplines. On the basis of the analysis of the results of users' interaction with educational and methodological materials of the OEC, it is possible to formulate assumptions about the effectiveness of the course and recommendations for its further improvement.

5 Discussion

Therefore, during the analysis of educational data, the authors identified the stages of application of EDM methods which include: study of the nature of the subject area; design and development of tools for the implementation of data collection and data generation; selection of approaches to identify the relationships between these data, and implementation of developed data processing tools and tools for a visual presentation of the results. Experimental data were obtained in DEE of the Herzen State Pedagogical University of Russia.

Calculations of dependencies between the highlighted indicators of the student's digital footprint confirmed the feasibility and importance of EDM methods implementation, as it allowed to draw conclusions about the impact of such indicators on academic results of students within the framework of current and midterm academic performance. It would be fair to conclude that the use of EDM methods is directly dependent on the preparedness of the educational organization: availability of appropriate equipment and skills, as well as digital and IT competence of the students. The results

of further research aimed at identifying factors influencing the success of students' mastering of the courses will allow, according to the authors, to create recommendations for an upgrade of the structure and content of educational programs in order to improve the quality of training.

6 Conclusion

The obtained results allow authors to move towards the next stage of the study, based on an expanded database, and analyze the students' performance within all the educational programs available at the university.

Based on such data, models (patterns) describing students' learning activities within the framework of blended learning will be developed, which, according to the authors, will allow predicting the success of students mastering of the learning path, identifying students at risk and "critical points" of educational programs that cause students difficulties in learning, and developing recommendations to improve the online education courses, etc.

Thus, the use of EDM methods in the framework of the proposed methodology and implementation of the developed tools for the analysis and interpretation of detailed data on various aspects of the learning process will improve the efficiency and effectiveness of education.

The method of analysis of the digital footprint of students suggested by the authors can be used for informed managerial decision-making during the design, editing, and implementing university educational programs with a view to improving the quality of training.

References

- [Abdulmohsen, 2016] Abdulmohsen A. (2016) Data Mining in Education. (IJACSA) International Journal of Advanced Computer Science and Applications, 2016, Vol. 7, No. 6.
- [Angeli et al., 2017] Angeli, C., Howard, S., Ma, J., Yang, J., Kirschner, P. A.(2017) Data mining in educational technology classroom research: can it make a contribution? Computers Education, 2017, No. 113, pp. 226–242.
- [Baranova et al., 2020] Baranova E. V., Vereshchagina N. O., Shvetsov V. (2020) Digital Tools for Learning Activities Analysis // Izvestia: Herzen University Journal of Humanities Science, Saint Petersburg, 2020, No. 198, pp. 56–65.
- [Baker & Yacef, 2009] Baker R., Yacef K.(2009) The state of educational data mining in 2009: A review and future visions // Journal of Educational Data Mining, 2009, Vol. 1, No. 1, pp. 3–17. doi: 10.5281/ZENODO.3554657.
- [Belonozhko et al., 2017] Belonozhko P. P., Karpenko A. P., Khramov D. A.(2017) Analysis of Educational Data: Directions and Perspectives of Application // Internet-journal NAUKOVEDENIE, 2017, Vol. 9, No. 4. Available online: <https://cyberleninka.ru/article/n/analiz-obrazovatelnyh-dannyh-napravleniya-i-perspektivy-primeneniya/viewer> (accessed on 5 July 2020).
- [Bowers et al., 2012] Bowers A. J., Spratt R., Taff S. A.(2012) Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity // The High School Journal, 2012, Vol. 96, No. 2, pp. 77–100, doi: 10.1353/hsj.2013.0000.
- [Bowers, 2010] Bowers A. J.(2010) Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data-driven decision making, dropping out and hierarchical cluster analysis // Practical Assessment Research and Evaluation, 2010, Vol. 15, Article 7, doi: 10.7275/r4zq-9c31.

- [Campbell & Majors, 2017] Campbell L., Majors C. (2017) Database Reliability Engineering: Designing and Operating Resilient Database Systems, O'Reilly Media, 2017.
- [Castro, 2007] Castro, F.(2007) Applying Data Mining Techniques to e-Learning Problems / F. Castro, A., Vellido, A., Nebot, F. Mugica// Studies in Computational Intelligence, 2007, Vol. 62, pp. 183–221.
- [Ezekiel & Mogorosi, 2020] Ezekiel U. O., Mogorosi, M.(2020) Educational Data Mining for Monitoring and Improving Academic Performance at University Levels. (IJACSA) International Journal of Advanced Computer Science and Applications, 2020, Vol. 11, No. 11.
- [Manyika et al., 2011] Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A.(2011) Big Data: The Next Frontier for Innovation, Competition, and Productivity / Ed. McKinsey Global Institute, 2011.
- [Kurbatsky, 2019] Kurbatsky V. N. (2019) Digital Footprint in the Educational Space as the Basis of Transformation of a Modern University // HSE: scientific-methodical and journalistic journal, 2019, No. 5, pp. 40–45.
- [Larionova & Karasik, 2019] Larionova V. A., Karasik A. A.(2019) Digital Transformation of Universities: Notes on the Global Conference EdCrunch Ural on Technologies in Education // University Management: Practice and Analysis, 2019, No. 23, pp. 130–135.
- [Lunkov & Kharlamov, 2014] Lunkov A. D., Kharlamov A. V.(2014) Intelligent Data Analysis // Textbook. National Research Saratov State University, 2014. Available online: <http://elibrary.sgu.ru/uchit/1141.pdf> (accessed on 3 October 2020).
- [Nikolaenko, 2019] Nikolaenko G. A.(2019) The Perspectives of Using Digital Traces of Researchers for Analyzing their Communication Strategies (by the Example of the Social Network ResearchGate) // Sociology of Science and Technology, 2019,
- [Ovsyanitskaya, 2013] Ovsyanitskaya L. Y.(2013) Intelligent Data Analysis as a Component of Pedagogical Management // Education and science, 2013, No. 10. Available online: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-kak-sostavlyayuschaya-pedagogicheskogo-upravleniya> (accessed on 14 October 2020).
- [Pedersen, 2019] Pedersen K.(2019) Modern Science Proves Intelligent Design: The Information System Worldview, Archway Publishing.
- [Peña-Ayala, 2014] Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. Expert systems with applications, 2014, 41 (4), pp. 1432–1462.
- [Razaque & Alajlan, 2020] Razaque A., Alajlan A.(2020) Supervised Machine Learning Model-Based Approach for Performance Prediction of Students. Journal of Computer Science, 2020, Vol. 16, 1150–1162.
- [Rawad & Rémi, 2020] Rawad C., Rémi B. (2020) Internationalizing Professional Development: Using Educational Data Mining to Analyze Learners' Performance and Dropouts in a French MOOC. International Review of Research in Open and Distributed Learning, 2020, Vol. 21, No. 4.
- [Shrestha & Pokharel, 2021] Shrestha S., Pokharel M. (2021). Educational data mining in moodle data. International Journal of Informatics and Communication Technology (IJ-ICT). 10. 9. 10.11591/ijict.v10i1.pp9-18.

- [Siemens & Baker, 2012] Siemens G., Baker R. S. (2012) Learning analytics and educational data mining: towards communication and collaboration // LAK'12. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. New York: Association for Computing Machinery, 2012, pp. 252–254, doi: 10.1145/2330601.2330661.
- [Shamsutdinova, 2020] Shamsutdinova T. M. (2020) Cognitive Model of Electronic Learning Trajectories Based on Digital Footprint // Otkrytoye Obrazovaniye, 2020, No 2. Available online: <https://cyberleninka.ru/article/n/kognitivnaya-model-traektorii-elektronnogo-obucheniya-na-osnove-tsifrovogo-sleda> (accessed on 21 September 2020).
- [Stauffer, 2019] Stauffer M. (2019) Laravel: Up Running: A Framework for Building Modern PHP Apps 2nd Edition, O'REILLY, 2019.
- [Tulupyeva et al., 2015] Tulupyeva T. V., Suvorova A. V., Azarov A. A., Tulupyev A. L., Bordovskaya N. V. (2015) Computer Tools in the Analysis of Students' Digital Footprints in Social Network: Possibilities and Primary Results // Computer Tools in Education, 2015, No. 5, pp. 3–13.
- [Tulupyeva et al., 2016] Tulupyeva T. V., Tafinyseva A. S., Tulupyev A. L. Approach to the Analysis of Personal Traits Reflection in Digital Footprint // Vestnik psihoterapii, 2016, No. 60 (65), pp. 124–137.
- [Vartanov, 2018] Vartanov S. A. (2018) Digital Media and Big Data: a Mathematical Approach to Media Environment Analysis // Vek Informatsii, 2018, Vol. 1, No. 2, pp. 211–213.
- [Zaitseva, 2012] Zaitseva O. (2012) Use of LMS Moodle in Education // Education and Mentoring: Methods and Practice, 2012, No. 2, pp. 59–64.
- [Zamora-Musa & Velez, 2017] Zamora-Musa R. Velez J. (2017) Use of Data Mining to Identify Trends between Variables to Improve Implementation of an Immersive Environment, Journal of Engineering and Applied Sciences, 2017, Vol. 12, No. 22, pp. 5944–5948.
- [Zhigalova, 2019] Zhigalova O. P. (2019) Formation of the Educational Environment in the Conditions of Digital Transformation of Society // Scholarly Notes of Transbaikal State University, 2019, Vol. 14, No. 2, pp. 69–74.
- [Zorić, et al., 2020] Zorić A., Obrenovic B., Akhunjonov U. (2020). Benefits of Educational Data Mining. 2020.