How Data Mining and Artificial Intelligence can Contribute to Increasing Academic Performance

Corina Simionescu[1], Mirela Danubianu[2], Marius Silviu Maciuca[3]

Abstract: In the increasingly advanced information age we live in, data is a precious resource in all fields, and education is no exception. The use of data mining methods and techniques and artificial intelligence in education can contribute to improving the learning process and the academic performance of students. The objective of data mining is to discover patterns from large volumes of data for the personalization of learning, the prediction of student performance, the prevention of school dropout, the evaluation of the effectiveness of educational programs, the creation of automated learning systems, the improvement of teaching processes, and innovation in education. In our paper, we aim to analyze the impact of the use of data mining and artificial intelligence to identify opportunities for their use to improve the educational system. In a future project, we propose to apply data mining techniques and methods to original data sets that lead to the best possible quality of pre-university education in Romania.

Keywords: educational data mining; data mining methods and techniques

1. Introduction

Data mining and artificial intelligence are complementary fields, both aiming to explore valuable knowledge from available data. The quality and quantity of the data directly influence the performance in the two fields because the decisions to be taken must lead to the improvement of the processes (Fischer et al., 2020).

[1]PhD in progress, Stefan cel Mare University of Suceava, Romania, Address: University Street 13, Suceava 720229, Romania, Corresponding author: corina.simionescu@usm.ro.
[2] PhD, Stefan cel Mare University of Suceava, Romania, Address: University Street 13, Suceava 720229, Romania, E-mail: mirela.danubianu@usm.ro.
[3] Ph.D. student, Stefan cel Mare University of Suceava, Romania, Address: University Street 13, Suceava 720229, Romania, E-mail: mmarmms@gmail.com.

Data mining is a field that is increasingly capturing the interest of researchers due to its ability to discover hidden and often interesting patterns within considerable volumes of information. These large data sets come from various real-world sources, so they are susceptible to being incomplete, noisy, or inconsistent (Danubianu, 2015).

The goal of discovering patterns in large volumes of data is to make the best decisions by gaining a correct understanding of phenomena or behaviors. Data mining uses techniques and algorithms to analyze existing data and identify patterns and relationships useful for making decisions or making predictions.

The use of data mining in education gave birth in 2005 to the field of *Educational Data Mining*, which can be described as a relatively new multidisciplinary field. In 2008, the first annual conference was held, resulting in The Journal of Educational Data Mining (2009) and a society, the International Educational Data Mining Society, in 2011.

Data mining combines techniques from statistics, artificial intelligence, machine learning, neural networks, database systems, and data visualization. (Figure 1)

Figure 1. Relationship between Data Mining and Related Fields (Simionescu et al., 2021)

Artificial intelligence (AI) is the concept and technology that al.ow systems to make decisions, learn from experiences, and solve problems in a way similar to human thinking. To make autonomous decisions based on the patterns identified in the data, artificial intelligence uses techniques such as Machine learning (ML), Artificial intelligence, Supervised learning, Unsupervised learning, Support Vector Machine (SVM), K- Nearest Neighbor (KNN) Artificial Neural Network (ANN), FUZZY, etc. (Varshney, et al., 2019).

While data mining uses machine learning techniques to extract patterns from data, helping to discover relevant knowledge and identify trends, artificial intelligence is built on machine learning, where artificial intelligence models are trained using data to perform various tasks, such as image classification, voice recognition, or decision-making.

Because the process of implementing artificial intelligence is relatively new, we need to ensure the accuracy of the processed data so that decision-making is not risky for the problem to be solved (Sharma, et al., 2023).

Both data mining and artificial intelligence have applications in education, but also in various fields such as market analysis, personalization of recommendations, fraud detection, optimization of production processes, and many others.

Data processed by data mining can be used as input for artificial intelligence systems, and AI models can be used to improve the performance of data mining processes.

Thus, data mining and artificial intelligence complement each other, contributing to the development and application of intelligent solutions for solving complex problems in different fields.

2. Data Mining Methods and Techniques Used in Education

In the field of education, data mining is used to explore data collected from a variety of sources: LMS (Learning Management Systems) platforms, e-learning systems, questionnaires and surveys, data collected at the educational institution level (age, gender, ethnic origin, previous education level, school performance assessment, behavioral characteristics, students' social interactions, curriculum, etc.). (Bayer, et al., 2012), discussion forums (a feature that can distinguish online classrooms from the traditional classroom) (Wong, 2015).

Knowledge extracted from large volumes of data can be used to improve the teaching-learning-assessment process, personalize the learning experience of students, etc.

Data mining for the analysis of academic performance is an important field in education and can provide valuable insights for improving the educational process and supporting student performance. We list some of the data mining methods and techniques used for this purpose:

2.1. Classification

This is a supervised learning method. The classification algorithm consists of building the classifier, and then using the classifier for classification. (Dol, et al., 2021). Classification involves assigning labels to datasets based on their characteristics and attributes. Among the classification algorithms we mention: Logistic Regression, Support Vector Machines, K-Nearest Neighbours, Kernel SVM, Naïve Bayes, Decision Tree Classification, Random Forest Classification.

Classification, one of the most known and used data mining methods (Ahmed et al., 2020) can be used for fraud detection, target marketing, performance prediction, medical diagnosis etc (Jiawei Han et al., 2011).

Other popular situations where classification algorithms can be used are email spam detection, speech recognition and biometric identification.

In education, classification can be used, for example, to identify key factors that influence academic performance and to make predictions about students' future performance.

2.2. Clustering

Is an unsupervised learning method? It involves grouping similar data into clusters so that common patterns or characteristics can be identified between those data. In the context of education, this method can help identify student profiles with similar characteristics: educational behavior, interests, learning styles, etc.

Clustering algorithms can be classified according to the data, the type analyzed, the similarity measure used to group the data, and the theory used to define the cluster (Sivogolovko et al., 2012).

The clustering algorithms used in data mining are: K-Means clustering, Mean Shift, Divisive Hierarchical Clustering, Hierarchical Agglomerative clustering, Gaussian Mixture Model, DBSCAN, OPTICS, BIRCH Algorithm.

Research based on clustering can lead to the personalization of educational content according to the identified needs of a specific group.

2.3. Association Rules

They are used in data mining to identify interesting relationships and correlations between different attributes or data sets. Examples of algorithms used in association rule-based research are: Apriori Algorithm, Eclat Algorithm, F-P Growth Algorithm.

In the context of academic performance, algorithms can be used to identify associations between certain behaviors or study habits and academic results or associations can be identified between the frequency of attendance at classes, the number of hours of individual study, and grades obtained in exams.

2.4. Regression

Is another data mining technique used to identify relationships and trends between the input variables and the target variable?

By using regression, the factors that have the greatest impact on academic performance can be identified and estimates of future performance can be made based on these factors.

2.5. Text Mining

Is a Data Mining technique focused on extracting semantic information, patterns, and knowledge from large collections of texts? This involves applying algorithms and techniques of natural language processing (NLP) to analyze, organize, and understand texts automatically and efficiently.

In the Text Mining process, textual data is transformed into a numerical form, called a vector representation, so that it can be processed by traditional Data Mining algorithms. This representation may include essential terms, word frequency, word pairs, or other text-specific features.

Text mining is used in many fields. Among the text mining methods and techniques, we mention text clustering, extraction of association rules, and text visualization (Salloum, et al., 2018).

These data mining methods and techniques can be applied to academic data to better understand student performance, provide personalized educational support, and improve the learning process in educational institutions.

3. The Use of Artificial Intelligence in Education

Artificial intelligence (AI) can be used in education to improve the learning process and personalize the learning experience of each student. Here are some ways in which AI can be implemented in education:

- *Virtual tutorial systems*: AI can be used to create virtual tutorial systems that can provide personalized explanations and individualized support for each student;
- *Data analysis to improve performance*: AI can analyze data about student performance to identify patterns and trends and identify ways to improve teaching methods;
- *Personalized recommendations for learning materials*: AI can analyze student performance and learning preferences to recommend appropriate learning materials and resources for each student;
- *Adaptive assessment*: AI can be used to create tests and assessments adapted to the level of knowledge and abilities of each student;
- *Teacher assistance*: AI can assist teachers in managing classrooms, organizing learning materials, and providing feedback to students;
- *Learning robots*: AI can be used to create learning robots that can interact with students and provide educational support;
- *Automatic translation*: AI can be used to translate learning materials into different languages, facilitating access to education for students speaking other languages;
- *Learning simulators*: AI can be used to create learning simulators that can help students practice their practical skills in a safe and controlled environment.

By using artificial intelligence in education, advantages can be obtained, such as personalizing learning, improving student performance, and facilitating access to education for all students. As with any new technology, there are risks associated with its use, such as the dehumanization of the learning experience and ethical and security issues. (Ayala-Pazmiño, M.et al., 2023).

4. The Impact of Using Data Mining and Artificial Intelligence in Education

We all use artificial intelligence quite often every time we use GPS systems, voice recognition apps, facial recognition, or filters that create crazy effects on Snapchat or Instagram, etc. (https://er.educause.edu/articles, 2023).

Initially, Artificial Intelligence took shape in the form of computing systems. Later, it evolved towards the development of intelligent educational systems based on the Internet and the online environment.

Currently, however, we are witnessing a transition to the use of integrated IT systems alongside other innovative technologies, including humanoid robots and web chatbots, which can assume the duties and functions of instructors either independently or in collaboration with them (Chen L.et al., 2020).

Using these advanced platforms, instructors have been able to complete their assignments more efficiently and coherently. For example, the process of reviewing and grading student assignments has improved significantly. In addition, the quality of teaching activities has seen a significant improvement.

On the other hand, one of the major advantages of these systems lies in their use of machine learning and adaptability. This has led to the customization and adaptation of curriculum and content to meet individual student needs. Thus, the learning process becomes more efficient and attractive, providing a more relevant educational environment oriented towards the individual development of each student (Upadhyaya, et al., 2023).

Due to the rapid evolution of research and the use of data mining and artificial intelligence in the field of education, it is difficult to mention all the studies conducted. In this paper, we can offer some notable examples in the field of using AI and data mining in education in Romania without pretending to be an exhaustive presentation.

"Stefan cel Mare" University in Suceava, Romania, through dedicated researchers, uses data mining to improve speech therapy therapy in a personalized way. (Danubianu, et al., 2009; Danubianu, et al., 2011; Danubianu, et al., 2012; Danubianu, 2013; Danubianu, et al., 2018).

These studies reveal the fact that data mining methods and techniques represent an extremely valuable tool in the context of speech therapy, having the ability to provide essential information for the implementation of personalized and optimized

therapeutic programs. These programs are specifically adapted to the characteristics of each child, which leads to a reduction in the duration needed for therapy while increasing the chances of obtaining superior results. Finally, the use of this technology can help reduce the costs of speech therapy. The motivation for research using Data Mining methods and techniques in education is supported by its significant potential to revolutionize the learning-teaching-assessment process. With a huge amount of data generated in the educational environment, from student performance to classroom interactions and learning preferences, exploiting this data becomes a valuable opportunity to improve the quality of education.

Data Mining research in education can bring many tangible benefits. First, it allows us to identify patterns and trends that might go unnoticed through traditional methods. Thus, we can discover aspects such as the factors that influence the failure of high school students at the baccalaureate level in Romania (Barila, 2021).

The educational data from the next seven years, regarding the follow-up of the baccalaureate exam of the students who studied at a high school in Suceava, Romania, were analyzed and processed with Data Mining techniques to discover correlations that would be difficult to find in other conditions (Marcu, et al., 2021).

Also, there are studies to identify the factors that cause students to refuse the exam in a certain baccalaureate subject (Maier, et al., 2023).

In addition to identifying patterns, Data Mining helps us personalize the learning experience. We can tailor teaching materials and strategies to individual student needs, resulting in more effective and engaging learning. By improving the interaction between teachers and students, data mining research can help create a stimulating and results-oriented learning environment.

Another crucial aspect is managing the risk of dropping out. Data Mining can flag students who are struggling and may need additional support. By intervening early and adapting instructional strategies, we can improve the chances of retaining these students in the education system.

In addition to educational benefits, Data Mining research in education can also contribute to the efficient management of resources in institutions. This leads to an optimization of the budget and a more accurate allocation of human and material resources.

Regarding the use of AI in education, an example is the use of chatbots by students at the Politehnica University of Timișoara. This is the first higher education

institution in Romania that encourages students to use Chatbot for study under the conditions of a critical approach to the contents provided.

Another very important aspect is related to the observance of the ethical norms of the university. In this context, the Politehnica University of Timișoara informed the students that the exams would be oral, a situation in which robots could not replace the students (Cocea, 2023).

Together with top European universities and industry partners, Babeș-Bolyai University (UBB) will implement, in the next 4 years, a large-scale European project that focuses on the development and implementation of Artificial Intelligence (AI) technologies. Through advanced research in AI, UBB aims to develop technological solutions to solve complex problems in various fields such as medicine, industry, urban development, and more. This endeavor aims to bring innovations that have a significant impact on today's society and contribute to the sustainable development of communities. In addition to research, the project also envisages the development of specialized academic programs in the field of AI. With this, UBB wants to train new generations of specialists in the field of advanced technologies, preparing them for the future challenges of industry and research.

From the study of the specialized literature in the IEEE Xplore database, it can be observed that in the last 5 years, 2018-2022, the interest of researchers from all over the world in the fields of artificial intelligence and data mining in education has greatly increased (Figure 2, Figure 3).

I searched the two databases as follows:

• For the search of specialized literature in the field of artificial intelligence, we used the keywords "artificial intelligence" and metadata "education"; then we ran the search for each of the last 5 years.
• To search for materials in the field of data mining, we used the keywords "data mining" and metadata "education". The search was performed for each of the last 5 years.
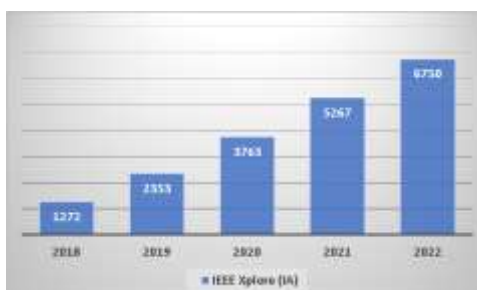
Figure 2 Results for ("All Metadata": Artificial Intelligence) AND ("All Metadata": Education)
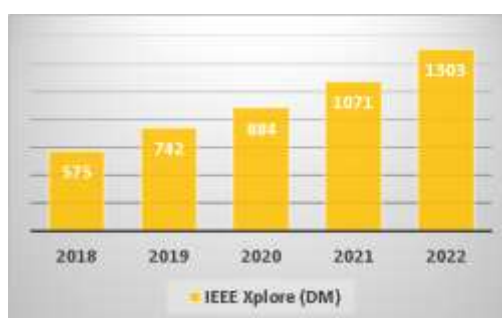

Figure 3. Results for ("All Metadata": Data Mining) AND ("All Metadata": Education)

It is important to note that the field of research in Data Mining and artificial intelligence in education is expanding, and new papers and findings are published frequently. The motivation for researching data mining in education is fueled by an interest in bringing significant innovations to the learning process and making education more personalized and effective.

This approach can transform the way education is delivered and experienced, setting the stage for a generation of better-prepared and motivated students.

The use of Data Mining and AI in universities is an interesting and complex research and development topic worldwide, to improve the academic experience of students and optimize educational processes.

5. Future Perspectives and Directions for the Integration of Data Mining and AI in Academia

The integration of Data Mining and AI in the academic environment has significant potential to transform the learning and research experience, providing several opportunities for improving academic performance and the efficiency of academic

81

institutions. Here are some insights and future directions for the use of Data Mining and AI in academia:

- *Personalization of learning*: continued development of artificial intelligence systems capable of personalizing each student's learning experience based on their abilities, needs, and learning style. By using data and AI algorithms, academic institutions can create personalized learning plans and tailored study materials for each student;
- *Curriculum optimization*: Applying data mining and AI to analyze data about student performance in different study programs and modules to optimize the content and structure of academic programs. This can help identify courses that are less effective and tailor curricula to meet labor market demands;
- *Predicting dropout*: using AI to identify students at high risk of dropping out and develop early interventions to support them. This can help reduce the dropout rate and increase the academic success of students;
- *Development of virtual assistant systems*: Integrating AI in the development of virtual assistants for students and teachers to answer questions, provide guidance, and provide constant and relevant feedback;
- *AI-based research*: expanding the application of data mining and AI in academic research to uncover new trends and patterns, improve research methodologies, and facilitate evidence-based decision-making;
- *Analyzing the effectiveness of teaching methods*: using data mining and AI to analyze data about student performance in different teaching methods to identify the most effective and innovative teaching methods;
- *Creating automated learning systems*: Artificial intelligence can be used to develop automated learning systems that can provide individualized feedback and support for each student. Thus, a more interactive learning process adapted to the specific needs of the student is promoted;
- *Use of AI in the administration of academic institutions*: Implementation of AI in the management of academic institutions, such as student data management, institution performance analysis, and resource planning;
- *Innovation in education*: Research with data mining and artificial intelligence opens up new opportunities for innovation in education. These technologies can be used to develop interactive educational applications, simulations, and other innovative learning tools that improve student engagement and outcomes;
- *Ethics and Security*: Strengthening ethical and security concerns in the use of Data Mining and AI in academia to ensure that student data is protected and that algorithms are transparent and unbiased.

These insights and future directions show that Data mining and AI have the potential to revolutionize academia, transforming learning, research, and management processes innovatively and efficiently. However, academic institutions must be aware of the ethical challenges and responsibilities associated with the use of this technology and develop responsible and sustainable approaches to implementing Data mining and AI in academia.

In conclusion, research with data mining and artificial intelligence in education focuses on creating a more effective educational environment that is more adapted to the individual needs of students.

By understanding and using data responsibly, new technologies can contribute to increased academic performance, student engagement, and the overall quality of learning.

This is an important step towards a more tailored, efficient, and relevant education in today's technological age.

References

*** (2019). https://er.educause.edu/articles/2019/8/artificial-intelligence-in-higher-education-applications-promise-and-perils-and-ethical-questions#fn1. *Web page*, retrieved at 19.07.2023.

Ahmed, M. R.; Tahid, S. T. I.; Mitu, N. A.; Kundu, P. & Yeasmin, S. (2020). A Comprehensive Analysis on Undergraduate Student Academic Performance using Feature Selection Techniques on Classification Algorithms. *11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, pp. 1-6.

Ayala-Pazmiño, Mario (2023). Artificial Intelligence in Education: Exploring the Potential Benefits and Risks. *593 Digital Publisher CEIT*. 8, pp. 892-899. 10.33386/593dp.2023.3.1827.

Barila, Adina; Danubianu, Mirela & Paraschiv, Andrei Marcel (June 2021). Towards a Deep Analysis of High School Students' Outcomes. *IJCSNS International Journal of Computer Science and Network Security*, Vol. 21 No. 6.

Bayer, J.; Bydzovská, H.; Géryk, J.; Obšívač, T. & Popelínský, L. (2012). Predicting drop-out from social behaviour of students. *International Conference on Educational Data Mining*, pp. 103–109.

Chen, L.; Chen, P. & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, vol. 8, pp. 75264-75278.

Cocea, M. (2023). *Web page*. https://mindcraftstories.ro/stiinta/cercetare-romaneasca-05-23-ai-ul-previne-anxietatea-si-consumul-excesiv-de-energie/, retrivied at 27.07.2023.

Danubianu, M. (2015). Step-by-step data preprocessing for data mining. A case study. https://educationaldatamining.org. *Web page*, retrieved at 17.07.2023.

Danubianu, M.; Pentiuc, S. G. & Socaciu, T. (2009). Towards the Optimized Personalized Therapy of Speech Disorders by Data Mining Techniques Fourth, *International Multi-Conference on Computing in the Global Information Technology.* Cannes/La Bocca, France, 2009, pp. 20-25, doi: 10.1109/ICCGI.2009.11.

Danubianu, Mirela & Barila, Adina (20130. Logo-DM – A Speech Therapy Optimization Data Mining System, IMMM. *The Third International Conference on Advances in Information Mining and Management.*

Danubianu, Mirela & Pentiuc, Stefan (2013). Data Dimensionality Reduction Framework for Data Mining. *Electronics and Electrical Engineering* 19. 10.5755/j01.eee.19.4.2043.

Danubianu, Mirela & Tobolcea, Iolanda & Pentiuc, Stefan. (2011). *Data Mining in Personalized Speech Disorders Therapy Optimization.* 10.5772/14129.

Danubianu, Mirela (2013). Combining the Power of Data Mining with Expert System for Efficiently Personalize Speech Therapy. *Stud. Inform. Univ. 11*, pp. 50-64.

Danubianu, Mirela; Pentiuc, Stefan Gheorghe; Tobolcea Iolanda & Socaciu Tiberiu (2018). *Model of a Data Mining System for Personalized Therapy of Speech Disorders.*

Danubianu, Mirela; Popa, Valentin & Iolanda Tobolcea (2012). Unsupervised Information-Based Feature Selection for Speech Therapy, Optimization by Data Mining Techniques, ICCGI. *The Seventh International Multi-Conference on Computing in the Global Information Technology.*

Dol, S. M. (2021). Use of Classification Technique in Educational Data Mining. *2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*, NaviMumbai, India, 2021, pp. 1-7, doi: 10.1109/ICNTE51185.2021.9487739.

Fischer, C.; Pardos, ZA.; Baker, RS.; Williams, JJ.; Smyth, P.; Yu, R. et al., (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, vol. 44, nr. 1, pp. 130-160,

Han, Jiawei; Kamber, Micheline & Pei, Jian (2011). *Data Mining: Concepts and Techniques.*

Maier, Mariana & Czibula, Gabriela & DELEAN, Lavinia-Ruth. (2023). Using Unsupervised Learning for Mining Behavioural Patterns from Data. A Case Study for the Baccalaureate Exam in Romania. *Studies in Informatics and Control*; 32; pp. 73-84. 10.24846/v32i2y202307.

Maier, Mariana & Czibula, Gabriela & DELEAN, Lavinia-Ruth. (2023). Using Unsupervised Learning for Mining Behavioural Patterns from Data. A Case Study for the Baccalaureate Exam in Romania. *Studies in Informatics and Control*, 32, pp. 73-84. 10.24846/v32i2y202307.

Marcu, D.; Danubianu, M.; Barîlă, A. & Simionescu, C. (2021). Algorithms for Classifying the Results at the Baccalaureate Exam - Comparative Analysis of Performances. *IJCSNS International Journal of Computer Science and Network Security*. Coreea de Sud, Seu.

Prashant Upadhyaya, Gaganjot Kaur (2023). Smart Multi-linguistic Health Awareness System using RASA Model*International Conference on Sustainable Computing and Smart Systems* (ICSCSS), pp. 922-927.

Salloum, Said; Al-Emran, Mostafa; Monem, Azza & Shaalan, Khaled (2018). *Using Text Mining Techniques for Extracting Information from Research Articles.*

Sharma, M. & Gupta, R. (2023). The Significance of using Data Extraction Methods for an Effective Big Data Mining Process. *2nd International Conference for Innovation in Technology (INOCON)*, Bangalore, India, pp. 1-4.

Shivi, Varshney; Rajvardhan, Jigyasu; Amandeep, Sharma & Lini, Mathew (2019). Review of various Artificial Intelligence Techniques and its applications. *IOP Conference Series: Materials Science and Engineering*. 594. 012023. 10.1088/1757-899X/594/1/012023.

Simionescu, C.; Danubianu M. & Turcu C.O. (2021). Data mining in educational data – a useful tool for sustainable learning development, *EIRP Proceedings of The International Conference European Integration - Realities and Perspectives 16th Edition*, pp. 349-353. https://dp.univ-danubius.ro/index.php/EIRP/article/view/212/194, 2021

Sivogolovko, Elena & Novikov, Boris (2012). Validating cluster structures in data mining tasks. *In Proceedings of the 2012 Joint EDBT/ICDT Workshops (EDBT-ICDT '12). Association for Computing Machinery*. New York, NY, USA, pp. 245–250.

Wong, J.-S.; Pursel, B.; Divinsky, A. & Jansen, B. J. (2015). An analysis of mooc discussion forum interactions from the most active users, in *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Washington, DC: Springer, pp. 452–457.