

**Publisher**<http://jssidoi.org/esc/home>

AI-DRIVEN STUDENT PROFILING: A CROSS-INDUSTRY REVIEW AND FUTURE RESEARCH DIRECTIONS ON MACHINE LEARNING FOR ADMISSIONS AND RETENTION*

Oksana Ivashchenko ¹, Mário Lelovský ², Andrii Kopp ³, Oleksandr Shmatko ⁴, Luboš Cibák ⁵

^{1,3,4} National Technical University "Kharkiv Polytechnic Institute", 2, Kyrpychova str., 61002 Kharkiv, Ukraine,

^{2,5} Bratislava University of Economics and Management, Furdekova 16, 851 04 Bratislava, Slovakia

E-mails: ¹Oksana.Ivashchenko@khpi.edu.ua; ²lelovsky@digitalnakoalicia.sk; ³Andrii.Kopp@khpi.edu.ua;

⁴oleksandr.shmatko@khpi.edu.ua; ⁵Lubos.Cibak@vsemba.sk

Received 11 January 2026; accepted 19 March 2026; published 30 June 2026

Abstract. Higher education institutions are increasingly facing challenges in student recruitment and retention, necessitating the development of data-driven segmentation and profiling models. Traditional segmentation approaches rely heavily on historical data and often fail to capture shifts in applicant behaviour. In contrast, AI-based techniques enable real-time analysis of interactions, supporting more adaptive and personalised recruitment strategies. Across fields like marketing, finance, and e-commerce, researchers have shown that clustering, supervised learning, and hybrid ML models can meaningfully improve segmentation tasks. This article explores the application of such methodologies in the context of higher education. A systematic literature review (SLR) and a Weighted Sum Model (WSM) were applied to evaluate and rank the relevance and adaptability of segmentation techniques used in the analysed studies. Results indicate that K-Means and Random Forest are widely applied in student segmentation and performance prediction. At the same time, hybrid models, such as K-Means combined with decision trees, offer improved accuracy and interpretability. However, real-time behavioural analytics (e.g., Google Analytics) and natural language processing (NLP) methods remain underexplored. Overall, the findings suggest that AI-driven segmentation models can enhance student engagement, academic success prediction, and strategic enrolment management. Future work will focus on integrating real-time behavioural data, refining segmentation models, and improving predictive analytics for university admissions and retention strategies.

Keywords: student segmentation; student profiling; higher education; machine learning; AI; clustering; predictive analytics

Reference to this paper should be made as follows: Ivashchenko, O., Lelovský, M., Kopp, A., Shmatko, O., Cibák, L. 2026. AI-driven student profiling: a cross-industry review and future research directions on machine learning for admissions and retention. *Entrepreneurship and Sustainability Issues*, 13(4), 461-478. <http://doi.org/10.9770/f7937756383>

JEL Classifications: I21, I23, C38

Additional disciplines: educology; information and communication; informatics

1. Introduction

Digital transformation is impacting various sectors of the economy and aspects of public life. Educational institutions are increasingly facing challenges not only in understanding and engaging with potential applicants during admissions campaigns, but also in retaining and adapting to the needs of current students. Creating profiles of these individuals, sometimes referred to as "customer identity creation," will improve student enrolment strategies, increase student retention, and create a personalised educational path for students.

* This research was funded by International grant project of VŠEM No. 1/2024_M – Management of the processes of recovery and sustainable development of public and economic law entities of the EU Member States, in particular Slovakia and Ukraine.

Traditional customer segmentation methods are static, relying on historical data rather than capturing current behaviour. Additionally, they lack predictive capabilities and primarily work with structured data. In contrast, the modern digital era generates vast amounts of unstructured data that are crucial for segmentation, predicting customer behaviour, and retaining customers.

Traditional data analysis methods in education, such as demographic segmentation or survey-based data analysis, do not provide effective results, given the large volumes of data about students and the fact that this data comes from different sources, has a different structure - student applications, academic records, activity on social networks. Machine learning methods address these challenges by integrating data from multiple sources and identifying behavioural patterns that are difficult to detect manually. These models support tasks such as profiling applicants, forecasting academic risks, and estimating the likelihood of student withdrawal. Methods such as clustering, supervised learning, and deep learning models are effective for creating student profiles and uncovering useful insights from the data. Additionally, combining machine learning methods with tools like Google Analytics allows universities to obtain useful statistics on applicants' engagement in real time and to personalise approaches to the recruitment of prospective students.

The purpose of this article is to investigate the application of machine learning methods to create customer profiles in various sectors of the economy, with an emphasis on their relevance and existing solutions for higher education institutions. The review seeks to identify potential gaps in the use of machine learning methods to create applicant and student profiles. These findings will form the basis for further research aimed at developing university applicant and student profiles, which will allow universities to conduct more effective, personalised admission campaigns and possibly increase student retention rates.

2. Literature review

Customer relationship management (CRM) is an integral part of any company's marketing strategy. CRM is a strategic approach to managing customer interactions, which aims to increase customer engagement, satisfaction, and long-term retention. CRM is most often used in business sectors to optimise marketing strategies. Still, in recent years, CRM principles have been increasingly applied by universities to improve the results of admissions campaigns, attract and retain students.

In the higher education sector, the use of CRM enables optimising the admissions process and tracking interactions with students at different stages of the admissions and study process, ultimately improving the performance of the higher education institution. In addition, combining CRM with machine learning and predictive analytics enables universities to identify student groups and personalise interactions with them.

Consider two CRM approaches for better understanding and engaging customers: customer segmentation and customer profiling. Customer segmentation is an analytical method for dividing customers into groups based on common characteristics, widely used in business and marketing. It allows companies to adapt their development strategies and promote products to different segments, thereby increasing efficiency, business profitability and customer engagement (Das et al., 2022).

For universities, segmentation enables you to optimise academic and administrative processes for different student groups. Traditionally, segmentation in higher education can be based on various factors: demographic data (age, nationality, socio-economic background), geographical data, behavioural engagement, academic performance and preferences when choosing disciplines; now, activity in social networks is also added to these data. Segmentation helps universities optimise marketing campaigns for personalised communication with different student groups, taking into account each group's unique needs and interests. Additionally, the use of this tool allows for the development of personalised academic support and individual communication strategies.

Segmentation techniques such as clustering and supervised learning can be effectively applied to student data to identify patterns in enrolment, engagement, and retention, according to research on predictive analytics and customer segmentation tools.

For universities, it is best to start analysing student data by segmenting students, given that universities attract students from different demographic groups, academic interests, and career goals, as well as behavioural differences, geographical, and socio-economic factors. The conducted student segmentation serves as the basis for student profiling, providing a deeper understanding of individual students within each segment. Student profiling is an approach that segments students by creating detailed profiles based on a combination of data, such as behavioural, demographic, geographic, and social network data. In this context, segmentation divides students into large groups, and profiling allows you to create more detailed student profiles within these groups (Das et al., 2022). Such deeper data analysis enables you to obtain more holistic information about students, for example, enabling early identification of students at risk of expulsion and individual career guidance.

The application of machine learning methods, such as clustering, deep learning, and predictive modelling, provides valuable insights from large datasets. In addition, tools such as Google Analytics help educational institutions analyse student interactions in real time, increasing the effectiveness of recruitment and engagement. The integration of CRM, segmentation, and profiling provides universities with a robust framework for enhancing student engagement and institutional efficiency, which, at the current stage of higher education institutions, especially in Ukraine, is an important factor in ensuring sustainable development and resilience to new challenges. Studies have shown that universities that implement data-driven segmentation and profiling methods have higher student retention rates and improve overall academic performance (Das et al., 2022). The combination of segmentation and student profiling allows universities to apply not universal strategies, but to develop measures and strategies that are tailored to the individual needs of students.

As previously mentioned, segmentation and student profiling are related but serve different purposes in analysing student data. While segmentation provides a macro-level view by grouping students into broader categories, student profiling focuses on micro-level insights by analysing specific characteristics and behaviours of individual students (Hemsley-Brown, 2016). The key differences between segmentation and profiling are summarised as follows in Table 1.

Table 1. The key differences between segmentation and profiling

Aspect	Segmentation	Profiling
Purpose	Group students into broad categories based on common characteristics	Creates detailed, personalised insights for individual students
Scope	Macro-level analysis	Micro-level analysis
Methods Used	Clustering, RFM analysis, demographic segmentation	Predictive modelling, supervised learning, deep learning
Application	Targeted marketing, student recruitment, personalised communication	Personalised learning, academic support, dropout prediction

Source: Own work

According to the study (Das et al., 2022), segmentation should be the first step in the analytical process, providing a structured foundation for student profiling. As a result, by first identifying broad student categories, educational institutions can refine their strategies using detailed profiling methods to ensure services are tailored to individual needs. The integration of CRM, segmentation and profiling approaches into university operations is essential for the effective functioning of educational institutions in an environment of intense competition. CRM is the basis for managing student interactions; segmentation allows you to identify significant groups of students for targeted student engagement and retention; and creating student profiles enables you to implement personalised approaches to student retention and interaction. As segmentation and profiling research continue to evolve, it is important to examine how machine learning (ML) approaches are applied across industries and how these methodologies can be adapted to higher education. This study aims to systematically review the existing literature on machine learning-based segmentation and profiling. To achieve this, a structured methodology for identifying, evaluating, and synthesising relevant scientific articles, previously used in previous studies, was adapted and refined. The following section describes the systematic review process, including database selection, search queries, inclusion and exclusion criteria, and the consensus approach used to ensure robust results.

3. Methodology of review

This study follows a systematic literature review (SLR) approach, with particular attention to accuracy, transparency, and replicability. We build on the structured review process outlined in our earlier work (Cherednichenko et al., 2023) on entity resolution and item matching, adapting it here to examine how AI and ML techniques are applied to segmentation and customer profiling in the education sector. To structure the review, we organised the process into several stages, illustrated in Figure 1.

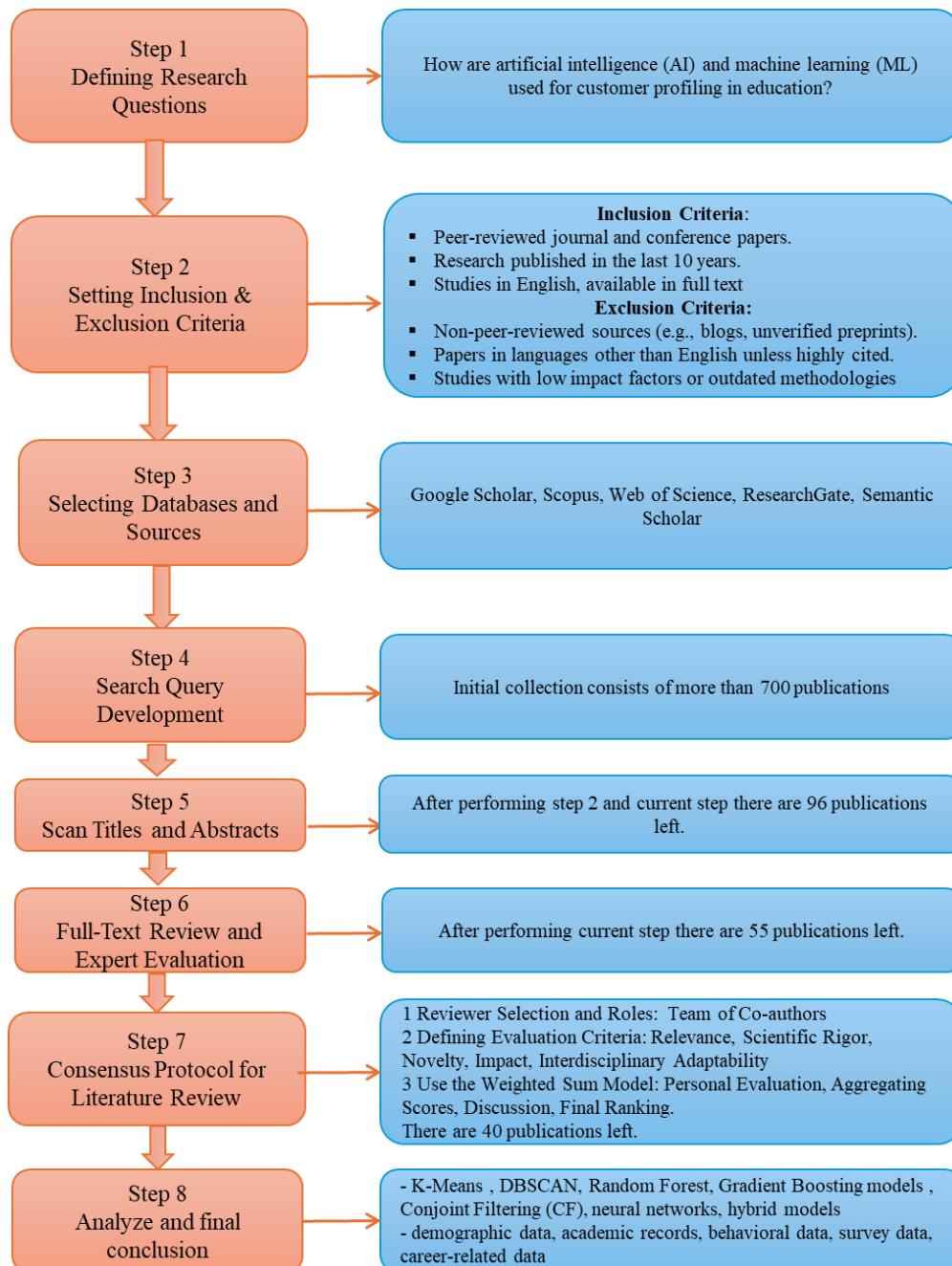


Figure 1. Step by step, the review methodology
Source: Own work

1. Defining the research questions. The central research question will be:

“How are artificial intelligence (AI) and machine learning (ML) used for customer profiling in education?”.

To explore the main question more fully, we also considered several supporting questions:

“What AI/ML methods are most commonly used for customer profiling?”

“What datasets and features are used for profiling in education?”

“What are the main advantages and limitations of AI-based profiling in this context?”

2. Establishing criteria for including and excluding articles from the review. To ensure the material reviewed is relevant, we applied the following inclusion criteria: peer-reviewed journals and conference publications; studies published primarily within the last five years; publications in English; and publications available to read in full. We also formulated exclusion criteria to filter out studies that did not meet the scope or quality requirements of the review: unverified sources; articles not in English; dissertations; studies with low impact factors or outdated methodologies.

3. Selection of databases and sources. Several bibliometric databases were selected for a comprehensive search of publications, taking into account criteria such as peer-reviewed status, broad coverage, and accessibility: Google Scholar, Scopus, Web of Science, ResearchGate, and Semantic Scholar.

“How are artificial intelligence (AI) and machine learning (ML) used for customer profiling in education?”.

To explore the main question more fully, we also considered several supporting questions:

“What AI/ML methods are most commonly used for customer profiling?”

“What datasets and features are used for profiling in education?”

“What are the main advantages and limitations of AI-based profiling in this context?”

2. Establishing criteria for including and excluding articles from the review. To ensure the material reviewed is relevant, we applied the following inclusion criteria: peer-reviewed journals and conference publications; studies published primarily within the last five years; publications in English; and available for full reading. We also formulated exclusion criteria to filter out studies that did not meet the scope or quality requirements of the review: unverified sources; articles not in English; dissertations; studies with low impact factors or outdated methodologies.

3. Selection of databases and sources. Several bibliometric databases were selected for a comprehensive search of publications, taking into account criteria such as peer-reviewed publications, broad coverage, and availability of access: Google Scholar, Scopus, Web of Science, ResearchGate, Semantic Scholar.

4. Developing a search query. Because segmentation and profiling in higher education is still emerging, we initially adopted a broad search strategy, including information from business, marketing, finance and healthcare. As a result, several search directions were identified, and corresponding queries were developed:

a) General customer profiling using machine learning:

- "machine learning" AND "customer profiling" AND "identity resolution";
- "customer segmentation" AND "predictive analytics" AND "machine learning";
- "data-driven customer profiling" AND "AI";
- "personalisation in customer experience" AND "machine learning";
- "customer behaviour modelling" AND "clustering algorithms";

b) Education-specific profiling (students & university):

- "student profiling" AND "machine learning" AND "university admissions";
- "AI in higher education" AND "student segmentation";
- "predictive analytics" AND "student behaviour" AND "academic performance";
- "university recruitment" AND "AI-driven student profiles";
- "personalised learning" AND "AI" AND "student engagement";
- "student success prediction" AND "data analytics";
- "dropout prediction" AND "machine learning" AND "higher education";

c) Cross-sector adaptation: finding solutions from other fields for universities:

- "customer insights" AND "AI" AND "higher education";
- "marketing analytics" AND "student enrollment";
- "customer relationship management" AND "education sector";
- "recommendation systems" AND "student retention";
- "customer loyalty prediction" AND "student engagement";
- "AI-driven user profiling" AND "education technology".

After testing multiple query combinations, we selected the formulation that yielded the most relevant results across databases: (*"machine learning" OR "AI" OR "artificial intelligence"*)AND (*"customer segmentation" OR "customer profiling" OR "student profiling"*)AND (*"higher education" OR "educational sector" OR "universities"*) AND (*"business" OR "industry" OR "corporate sector"*). Therefore, after the execution of the second and current steps, more than 700 publications were obtained.

5. Title and abstract comparison. Once the search was completed, we screened the retrieved publications against the inclusion and exclusion criteria. After this, the retrieved articles were additionally checked based on:

- exclusion of duplicates, i.e., repeated studies from different databases;
- the relationship between the title and the abstract, with priority given to those studies where machine learning was mentioned.

This step removed duplicates and clearly irrelevant studies. Before proceeding with further evaluation, implementing the above steps reduced the number of publications to 96.

6. Full-text review and peer review. Following the preliminary screening, the co-authors conducted a more detailed assessment of the remaining publications to meet the following criteria:

- citation availability – priority was given to well-cited sources;
- methodological rigour – ensuring the use of sound empirical approaches in research;
- taking into account the presence of segmentation and profiling tasks in the research.

Final decisions were made through discussion and consensus among co-authors. This step reduced the number of publications for further in-depth analysis to 55.

7. Consensus protocol for literature review. The Weighted Sum Model (WSM) was used to evaluate the selected articles. A team of co-authors with expertise in machine learning, educational technology and customer profiling carried out the review process. To ensure consistency, all publications were evaluated according to a structured system of weighted criteria, with weights reflecting the importance of each criterion:

- relevance – alignment with the research area: segmentation and profiling of university students;
- scientific rigour – quality of research methods used;
- novelty – introduction of new concepts or models;
- impact – potential impact on scientific and business communities;
- interdisciplinary adaptability – can the concepts be applied in different fields.

Each co-author assigned a weight to each of the criteria independently. Each publication was evaluated according to these criteria using the following scale: 1 - very low, 2 - low, 3 - moderate, 4 - high, 5 - very high.

After all publications were evaluated for each co-author, a weighted score was calculated for each article using their individual scores and weights. The next step was to calculate the average of the three weighted scores for each article. This approach allows for the individual weighting of each reviewer's criteria to be taken into account.

After the individual evaluation of each publication, a meeting was held between the co-authors to discuss existing discrepancies in scores and weights. This iterative process ensured a balanced and consistent evaluation across all publications. This step allowed us to reduce the number of publications for further in-depth analysis to 40.

The WSM evaluation allowed us to classify the reviewed studies into three relevance tiers, reflecting their methodological quality and their potential applicability to higher-education profiling tasks. Figure 2 presents

the studies from the first two groups. Studies with the highest rating (4.30+) offer methodologies that are most relevant to our research topic, while studies with lower ratings (3.50) offer AI and clustering methods from other fields that can be adapted to enhance university student segmentation models.

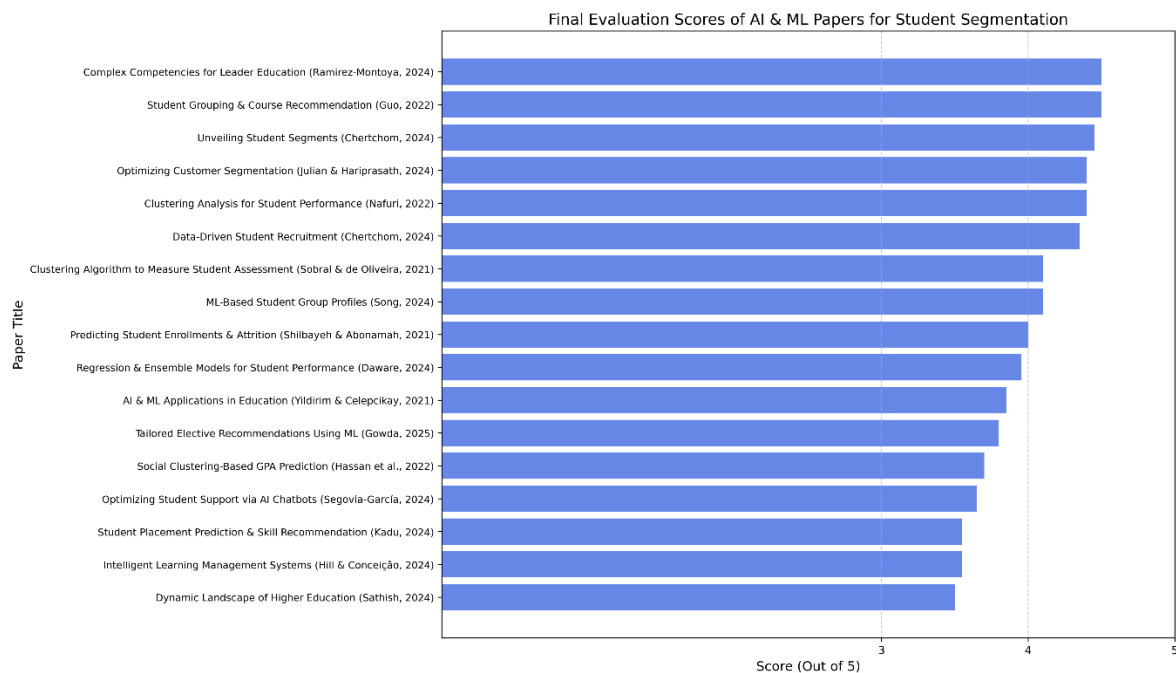


Figure 2. The studies with ranks higher than 3.50

Source: Own work

The articles with the highest rating (4.30+) represent, from the authors' perspective, the most effective methodologies for student segmentation and profiling. These studies directly contribute to the development of AI-driven segmentation and profiling models that can be applied in university marketing and admissions strategies. These studies suggest clustering methods, predictive analytics, and machine learning approaches that can improve admissions personalisation processes and facilitate the development of more effective student engagement and retention strategies.

Studies within the 4.30–3.50 rating range include studies that offer valuable approaches and models that can be adapted for profiling and segmenting students during admissions campaigns. Based on the analysis of student data, these studies offer tools for predicting student success and developing personalised learning strategies. Articles with a rating below 3.50 focus mainly on customer segmentation and profiling in non-education industries such as banking, retail, and marketing. Despite their industry-specific focus, these studies are valuable because customer segmentation has been extensively applied in these fields for many years with proven, effective implementations. In this regard, they offer artificial intelligence methodologies and clustering infrastructure to improve student profiling models. Their adaptation can facilitate the adoption of innovative techniques and retention strategies by applying business intelligence methods to analyse student data.

The reviewed studies address our research questions and highlight several methodological directions for future work. Less relevant articles, while not directly focused on higher education, offer valuable insights that can help develop models for student segmentation and profiling. In particular, AI-based segmentation techniques obtained from other business areas can be adapted to improve student recruitment strategies, predict enrolment trends, and enhance retention efforts at universities.

4. Findings and discussion

Artificial intelligence (AI) and machine learning (ML) methods are widely used in higher education, particularly for student profiling, segmentation, predictive analytics, and clustering. We will examine the methods and approaches proposed by the authors in 18 publications that received a score higher than 3.50, and assess their appropriateness for creating student profiles, segmenting students during the admission campaign, and developing personalised learning. In addition, earlier foundational studies are referenced selectively to contextualise the evolution of student profiling approaches; however, these works were not included in the final evaluated set.

In several studies, the authors used ML models to predict student success, retention, and dropout risks. Research on predicting future employment using machine learning methods is highly relevant, given the competition and pressure university graduates face in the modern world. For instance, in (Kadu et al., 2024), the Random Forest machine learning method was used in combination with feature selection to predict employment and provide recommendations on professional skills, based on the following set of data available in the university database: academic performance (CGPA, points), internship experience, completed projects, certificates obtained, aptitude test results, communication skills ratings and extracurricular activities. Another area of application of machine learning methods is to determine the structure of student recruitment and attrition. According to (Shilbayeh & Abonamah, 2021) predictive models such as Random Forest and Decision Trees were applied to identify groups of students that the university should attract to comply with academic standards, and risk groups, i.e. students who are most likely to be expelled. As in previous work, the authors emphasise the importance of properly preparing input data: selecting characteristics, filling in missing data, and removing outliers or incorrectly entered data. The dataset used by the authors for this model included students' demographic data and academic records (e.g., English proficiency, academic performance). In Ramirez-Montoya et al. (2024), the researchers propose using Support Vector Regression (SVR) combined with Recursive Feature Elimination (RFE) to predict student performance. The use of RFE enables the identification of the most relevant predictors for the model, thereby ensuring its interpretability and computational efficiency. Predictive models help identify students who may need additional support and develop targeted interventions. A study (Daware, 2024) explores machine learning methods to analyse key factors influencing student achievement and to develop predictive models. Several algorithms were used, including linear regression, random forest regressor, gradient boosting regressor, support vector regressor (SVR), and K-nearest neighbours regressor, each chosen for its ability to capture different patterns in the data, whether linear or complex. The models were trained using characteristics such as attendance, socio-economic background, parental education, and academic history to determine their impact on student achievement. The best results were achieved with ensemble learning methods, including gradient boosting regressor and support vector regressor (SVR), which effectively captured nonlinear relationships in student achievement data. Using such a model will allow identifying students at risk for academic success and implementing personalised support.

The research (Hassan et al., 2022) integrates clustering (K-Means) with regression (LASSO regression) to predict GPA based on social and academic factors. Students with similar engagement behaviours were grouped, and regression models were applied to each cluster to predict final GPA. Unlike previous works, the authors used not only demographic data but also information on students' social behaviour; this data for their dataset was collected by analysing mobile sensor data and by conducting surveys of students before and after the courses. The findings in (Abyaa et al., 2019) indicate that the characteristics of students, which are used to create a student profile within the framework of the active development of e-learning: a static student profile, knowledge, cognitive abilities, social characteristics and motivation, while these characteristics are used both independently in machine learning models and in combination, as was done in (Hassan et al., 2022). The authors of (Rai et al., 2024) propose the introduction of data analytics into the work of the marketing department and the student recruitment department, for this purpose a survey was developed and conducted among potential applicants, the survey results were used to determine the most potential candidates for enrolment in the university, for this purpose it was proposed to use the following algorithms: Logistic Regression, K-Nearest Neighbor, and Artificial Neural Networks. Artificial Neural Networks demonstrated the highest accuracy.

Predictive analytics can improve the process of creating a student profile by identifying key characteristics to include. Combining predictive models with clustering can improve approaches to student segmentation.

Another set of studies focuses on clustering techniques, using indicators such as academic results, engagement metrics, and behavioural attributes to form meaningful student segments. For example, in (Mohamed Nafuri et al., 2022), the following clustering methods K-Means, DBSCAN, BIRCH are proposed to identify groups of students based on their performance indicators. In this study, to solve the problem of "large amount of data", which leads to difficulties in data processing and analysis, feature selection methods were applied to find the best attributes for this study. Both supervised (using random forest, extra tree, info gain, and chi-square techniques) and unsupervised (applied variance thresholding (≥ 0.3) to identify high-variance features and remove redundant ones.) feature selection methods showed the same result, so the following 12 features were selected: place of birth, income groups, high school, university, registration age, entry qualification, field of study, sponsorship, CGPA, number of activities, and employment status. The application of this approach is relevant to university CRM, as it can provide student segmentation for personalised learning. Traditional market analysis methods often fail to handle a large, diverse customer base, a limitation that is also common in higher education institutions due to the volume of student data. The solution to this problem may lie in implementing big data analytics and machine learning, which offer more efficient, automated segmentation methods. Among such approaches, the most popular and effective approach is K-Means clustering (Julian et al., 2024). The researchers (Julian et al., 2024) believe that this technique is particularly useful in retail. Still, from our point of view, it can also be adapted for the higher education sector, thereby improving segmentation and profiling efficiency. In Song et al. (2024), the authors use survey results as data about students. The survey was conducted among 2,492 students, and data were collected on students' demographics, educational status, personal morality, political ideology, interests, and hobbies. To assess the questionnaire's validity and reliability, the Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were used. Based on the clustering results using the K-Means clustering method, the authors built a classification prediction model for university student groups based on the BP neural network. This combination of two machine learning methods helped overcome the limitations of traditional survey result processing, yielding student profiles that support adaptive learning paths and the development of personalised services for students. The K-Means clustering method was used by researchers (Sobral & de Oliveira, 2021) to test whether it is possible to take into account students' self-assessment during the learning process and to assess the correlation between self-assessment and other student characteristics, the data source was the results of a student survey and university academic records. In the study (Chertchom, 2024), clustering-based segmentation using the K-Means algorithm is used to identify patterns of applicant registration to improve student recruitment strategies. The data source was the university's applicant registration system. In this study, the characteristics that determine the choice of applicants were "Province", "Parent Occupation", and "Parent Revenue". The results of this study were used to develop marketing strategies for student recruitment in different regions. The authors' focus (Guo et al., 2022) is on developing an improved K-means clustering algorithm for student segmentation, and they introduce a group-oriented course recommendation model that takes into account both statistical and semantic dimensions. Word2Vec Embeddings was applied for feature extraction to better capture semantic relationships among student characteristics. For further student segmentation was used K-Means Clustering (Improved Version) based on their learning preferences and academic performance. The use of Word2Vec for feature extraction offers an innovative way to analyse unstructured student data, which could be useful in processing text-based academic records, student feedback, and learning preferences.

According to Guan et al. (2020), between 2010 and 2019, the number of studies on student profiling and learning analytics using AI increased significantly, with a focus on learning processes and the application of AI in education. The most common themes in published research focused on learning models, student engagement, and predictive analytics for academic success and dropout prevention. A key feature of this research period was the shift to AI-powered predictive modelling to track student progress, academic performance, and dropout risk. Additionally, learning analytics, big data, the use of AI techniques, and ethical considerations around the use of student data for institutional analytics became key research areas during that period. These ideas are relevant to our research as they highlight how AI-driven segmentation and profiling models can be improved to enhance student support, retention prediction, and personalised academic guidance.

Artificial intelligence has played a significant role in increasing the integration of personalisation into university learning processes, thereby improving student engagement, academic advising, and real-time learning adaptation (Yildirim & Celepcikay, 2021). AI applications on EdTech platforms play an important role in tracking student behaviour, providing personalised learning experiences, and offering intelligent recommendations. A study by Segovia-García (2024) highlights the role of AI-developed chatbots in student-institution interactions, providing students with instant academic support and guidance. These chatbots generate real-time interaction data that can be used to profile behaviour and clustering patterns to improve student retention strategies. Additionally, Gowda et al. (2025) demonstrate how collaborative filtering and genetic algorithms optimise course recommendations based on choice, ensuring that students receive personalised guidance on their academic trajectory. While these AI-based systems do not directly segment students, they provide important behavioural insights that can improve profiling models for data-driven student support. Modern learning management systems (LMSs) have evolved into sophisticated AI-powered platforms that personalise learning paths. According to Hill and Conceição (2024), intelligent learning management systems (ILMSs) use machine learning algorithms to track student engagement, dynamically adapt course content, and segment students based on real-time interactions. By analysing LMS activity data, institutions can proactively identify students who are struggling, optimise curriculum development, and implement personalised academic interventions. These AI-driven tools create an adaptive education system and also facilitate student segmentation and profiling.

The paper of Sathish et al. (2024) highlights the development of a job recommendation system for students that continuously improves its recommendations based on user feedback (rejection of a job offer or acceptance). This mechanism is similar to adaptive student segmentation models in education, where machine learning algorithms adjust student groupings based on behaviour, activity, and academic performance. In this way, the concept of adaptive learning from recommender systems can be used to improve real-time student segmentation by incorporating dynamic feedback loops, making segmentation models more responsive to students' learning behaviour and performance trends.

The graph (Figure 3) below summarises the frequency of machine learning techniques applied in the reviewed articles. The graph below shows the frequency of use of machine learning methods in the reviewed articles. According to this graph, the following conclusions can be made:

- K-Means clustering is the most widely used technique for student segmentation, often combined with feature selection methods.
- Random Forest and Gradient Boosting models are common for predicting student performance due to their ability to handle multidimensional learning data.
- Conjoint Filtering (CF) and neural networks are gaining popularity for personalised course recommendations and adaptive learning systems.

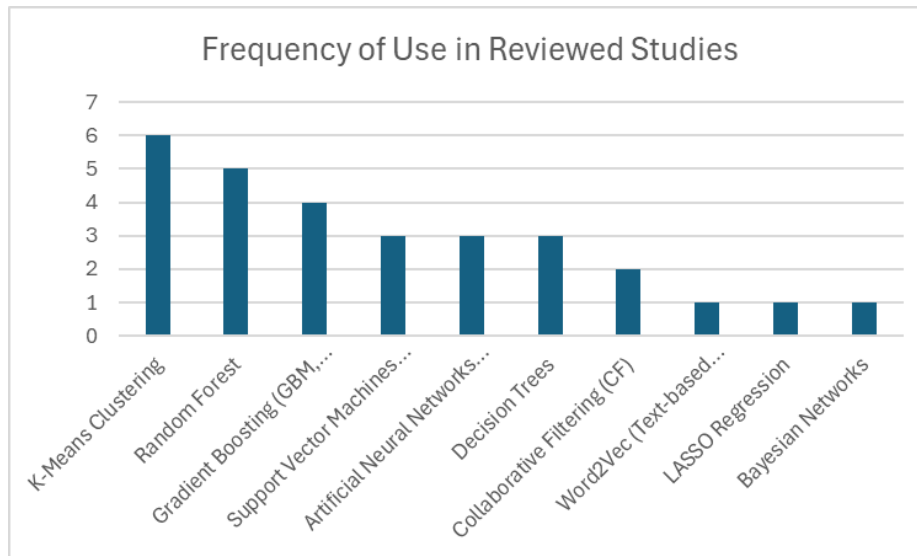


Figure 3. The frequency of machine learning techniques applied in the reviewed articles
Source: Own work

Some interim conclusions can be drawn: first of all: machine learning is widely used for student profiling and retention prediction, using clustering models to segment students based on academic, behavioural, and socio-economic factors; second, course recommendation systems developed using AI methods are becoming a powerful tool to improve learning personalisation; third, hybrid models combining clustering + supervised learning (e.g., K-Means + Decision Trees) are effective in improving segmentation accuracy. It should also be emphasised that although our study did not pay much attention to this issue, ethical issues regarding AI-based analytics remain a key concern, especially in the use of student data for predictive decision-making. Table 2 summarises the primary focus areas, ML techniques used, and potential future applications of these techniques.

Table 2. The primary focus areas, ML techniques used, and potential future applications of these techniques

Focus Area	Key Techniques Identified	Future Applications
Student Profiling & Segmentation	K-Means, DBSCAN, Improved K-Means, BIRCH	Adaptive learning pathways, micro-segmentation
Predicting Student Success & Retention	Random Forest, Decision Trees, SVM, GBM, LASSO Regression	Early intervention, targeted academic support
Personalised Course Recommendation	Collaborative Filtering, Word2Vec, Hybrid AI models	AI-driven academic advising, automated course recommendations
Dropout Prevention & Risk Prediction	Random Forest, Gradient Boosting, Bayesian Networks	AI-powered early warning systems
Learning Analytics & Adaptive LMS	AI-enhanced Learning Management Systems (ILMS), Neural Networks	Dynamic content adaptation, student engagement tracking
University Admissions & Marketing	Logistic Regression, K-Means, Neural Networks	AI-based recruitment strategies, data-driven admissions decision-making
Employment Prediction & Career Matching	Decision Trees, Random Forest, Feature Engineering	AI-powered job placement support, career path recommendations

Source: Own work

In addition to studies on the application of machine learning methods in the higher education sector, studies on the use of various machine learning (ML) methods for customer segmentation and profiling and data set creation strategies in many other sectors (the retail, finance, marketing, and telecommunications sectors of the economy) were selected for analysis, they were rated by authors below 3.50. The aim is to explore how these methods can be adapted and applied in the higher education sector to improve student profiling, admissions strategies, and retention efforts.

ML-driven segmentation methods can be broadly classified into unsupervised, supervised, and hybrid models. These methods are widely used in industries such as finance, marketing, retail, healthcare, and education, and

provide insights into customer behaviour, preferences, and predictive trends. Our review results are summarised in Tables 3 and 4.

Table 3. Machine learning methods used in various industries to perform segmentation and profiling

Machine Learning Method	Purpose in Customer Segmentation	Relevant Papers & Sources
Unsupervised Learning		
Affinity Propagation	A clustering algorithm that does not require pre-defined cluster numbers.	Mim & Logofătu (2022); Panji (2023)
DBSCAN (Density-Based Spatial Clustering)	Segments high-value customers and filters out noise data.	Hicham & Karim (2022); Kiran et al. (2024)
Fuzzy C-Means Clustering	Used for dynamic segmentation in marketing.	Widianti (2024); Thitipornndharma & Khonthaoagdee (2024)
Hierarchical Clustering	Groups of customers based on purchasing power and product affinity.	Wani (2023); Gupta & Israni (2024)
K-Means Clustering	Groups of customers into similar segments based on pre-defined numerical features.	Mustafa (2023); Wang (2024); Arakh (2024)
	Identifies customer groups based on spending habits, loyalty, and preferences.	Julian & Hariprasath (2024); Saxena (2022); Arakh (2024)
Spectral Clustering	Used in banking to classify credit risk groups	Shaikh & Khedkar (2023)
Principal Component Analysis (PCA)	Reduces data dimensionality to improve segmentation accuracy.	Widianti (2024); Alkhayrat (2020); Wani (2023)
Supervised Learning		
Decision Trees (DT)	Rule-based classification that helps define customer segments.	Mustafa (2023); Thitipomndharma & Khonthaoagdee (2024); Wang (2024)
Random Forest (RF)	Ensemble method that improves accuracy and identifies key features for segmentation.	Mustafa (2023); Cao (2024); Wang (2024)
Support Vector Machines (SVM)	Finds optimal hyperplanes to classify customers into different segments.	Mustafa (2023); Cao (2024); Arakh (2024)
Optimisation Techniques		
Genetic Algorithms	Optimisation technique that enhances clustering accuracy.	Hamidi & Haghi (2024)

Source: Own work

ML-based segmentation has many advantages, but there are still challenges such as data quality, algorithm selection, and interpretation, which can be overcome by hybrid models that combine machine learning with traditional statistical methods, improving both segmentation accuracy and interpretability. These hybrid models often integrate clustering algorithms with dimensionality reduction or ensemble methods, balancing performance and transparency in the segmentation process. For example, using such a combination of K-Means + Decision Trees increases the accuracy of the model and the usefulness of the model in real-world e-commerce scenarios, while universities can use such a hybrid model to group applicants based on academic records and then predict their probability of success.

Table 4. Hybrid models utilised in various industries to perform segmentation and profiling

Hybrid Model Used	Industry	Paper
Gaussian Mixture Model (GMM) + Logistic Regression	Retail / Marketing	John et al. (2023)
PCA + K-Means + Hierarchical Clustering	Finance	Wani (2023)
K-Means + (K-nearest neighbours, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine)		Hoang Dang Tran (2023)
K-Means + Decision Tree	Credit card companies	Thitipornndharma & Khonthaoagdee (2024)
K-Means + PCA + Decision Trees	Retail	Rahmet & Gürkan (2024)
K-Means + Apriori		Phyo et al. (2023)
K-Means + SVM	E-commerce	Arakh (2024)
Hybrid Clustering (K-Means + PCA + Decision Trees)	Various industries	Rahmet & Gürkan (2024)

Source: Own work

A major challenge in segmentation is data collection and preprocessing. Industries utilise various dataset creation strategies to build high-quality customer and student profiles (Table 5).

As a result of analysing strategies used in other business sectors, we can distinguish between data fusion and integration and feature development and dimensionality reduction (PCA, LDA). The first approach is effectively used in industries such as telecommunications and retail, combining structured (purchase history) and unstructured data (social media activity) to create a complete customer profile. In higher education, artificial intelligence methods can also be used to combine LMS data, academic records, customer surveys, social media activity to increase segmentation accuracy, and improve the learning experience and admission strategies. The dimensionality reduction approach is actively used in marketing and finance, PCA (principal component analysis) and LDA (linear discriminant analysis) reduce the complexity of the data while preserving the main attributes.

Using this approach in higher education will ensure the selection of the most relevant student attributes for profile development and noise reduction in student data, resulting in a more accurate student profile and improved personalised support system.

Table 5. Dataset creation strategies used in various industries

Paper	Industry	Dataset Used	Key Features
Alkhayrat (2020)	Telecom	Mobile usage data, call detail records	Call frequency, data consumption, billing history
Widianti (2024)		Call detail records, mobile data usage	Call duration, data consumption, network preferences
Shaikh & Khedkar (2023)	Banking	Account activity, credit card transactions	Spending behaviour, credit score, loan history
Wang (2024)	Retail & Finance	Transaction records & customer demographics	Age, income, purchase frequency, payment method
Lewaaelhamd (2023)	Retail & E-commerce	Purchase records, browsing data	Spending frequency, recency, product categories
Mustafa (2023)		Online purchase logs, browsing data	Purchase frequency, spending habits, product categories
Katragadda (2023)	Marketing	Website interactions, ad engagement, online behaviour	Clickstream data, social media engagement, conversion rates
Thitiporntharma & Khonthaoagdee (2024)	Finance	Transaction patterns, credit card usage	Spending clusters, income levels, payment preferences

Source: Own work

Table 6 highlights how ML-based clustering and classification models, leveraging different business areas, can significantly enhance student recruitment strategies by segmenting applicants, identifying outliers, optimising marketing, and improving scholarship distribution.

Table 6. Applications of business model in higher education and their benefits

Business Model	Application in Admissions	Benefits
K-Means Clustering	Groups applicants based on academic performance, location, interests	Helps target marketing efforts and scholarship offers
DBSCAN (Density-Based Clustering)	Identifies outliers (applicants with unique backgrounds)	Ensures diversity and inclusion in recruitment
Hierarchical Clustering	Creates detailed applicant segments based on multiple features	Customises admission messaging for different applicant types
Support Vector Machines (SVM)	Classifies students as high, medium, or low probability enrollees	Optimises recruitment resources by focusing on likely candidates
Neural Networks (Deep Learning)	Predicts which students are most likely to succeed	Improves scholarship and admissions decisions

Source: Own work

To sum up, customer segmentation models from various industries offer valuable methodologies that can be adapted for student profiling and recruitment strategies in higher education. Unsupervised clustering techniques provide dynamic segmentation, supervised models enable academic performance prediction, and hybrid

approaches combine the strengths of both for enhanced student profiling. AI-driven dataset creation techniques, such as data fusion and feature selection, ensure more reliable insights for university decision-making.

Our research aimed to answer the following key question: How are artificial intelligence (AI) and machine learning (ML) used for customer profiling in education? AI and ML are widely used in student segmentation, profiling, predictive analytics, and clustering models; various predictive models are used to predict student performance, dropout probability, and future employability. Learning management systems (LMS), recommender systems, and chatbots powered by AI allow universities to improve student engagement and personalise their learning trajectory. The most common methods used for segmentation are K-Means, DBSCAN, hierarchical clustering, in turn, for predicting student success and dropout rates, such methods as random forest, decision trees, gradient boosting, logistic regression are most often used, various hybrid models are successfully used for both segmentation and prediction, for example, K-means + decision trees, clustering + regression (for example, LASSO regression). In addition to these widely used ML methods, the use of Natural Language Processing (NLP) technique Word2Vec is also being implemented to process student feedback and academic records; joint filtering, neural networks and reinforcement learning are used for personalised recommendations and personalisation of students' learning trajectory. The key issue for the effective and high-quality work of AI/ML models is the selection of student characteristics, the data used in the studied sources can be divided into the following groups:

- demographic data: age, gender, nationality, socio-economic background;
- academic records: GPA, course performance, attendance;
- behavioural data: online engagement, social media activity, LMS interaction;
- survey data: student self-assessment, satisfaction surveys;
- career-related data: internship experience, employment status, skill gaps.

Despite covering many aspects of ML-driven student profiling, some areas remain understudied. First, the integration of Google Analytics to analyse applicant or student behaviour in real time is not actively explored, i.e. most studies use static datasets and do not examine real-time web interactions. Accordingly, future work should investigate how website visitation patterns influence enrolment decisions. Second, there is little use of natural language processing (NLP) methods to process, for example, student feedback, chatbot interactions, and emails, and incorporate this information into student characteristics used in creating profiles. Third, improving hybrid models. Fourth, implementing real-time adaptive segmentation. Current segmentation methods classify students using static datasets, but it is possible, for example, to develop and implement reinforcement learning models that adjust engagement strategies based on real-time student interactions.

Conclusions

This study examined AI and machine learning models and techniques for customer segmentation and profiling in higher education and business, identifying key clustering, predictive analytics, and hybrid models that can be adapted to university admission and retention strategies, and outlining issues that require further research.

Let us outline some future directions for our research on student segmentation and profiling.

Stage 1. Data collection and feature development for student admissions:

- 1) Collect structured demographic and academic data and unstructured data on applicant behaviour, such as social media activity, university chats, and correspondence;
- 2) Feature selection methods such as random forest feature importance ranking and principal component analysis (PCA) will be used to optimise data quality;
- 3) Perform research and application of approaches to combining data from different information sources;
- 4) Explore the feasibility of using cloud computing solutions such as AWS, Google Cloud or Microsoft Azure for data storage, integration and real-time analytics while complying with GDPR (General Data Protection Regulation) and FERPA (Family Educational Rights and Privacy Act).

Stage 2. Integrating real-time behavioural data and using NLP techniques. As part of Stage 2, future work should focus on integrating real-time tracking of applicant behaviour using web analytics tools such as Google

Analytics. In addition, enrichment of the dataset using NLP techniques (e.g., Word2Vec for text analysis of student feedback) will be explored to improve profiling accuracy and segmentation accuracy:

- 1) Collect and analyse Google Analytics data to track applicant interactions with the university website;
- 2) Apply NLP techniques to analyse student feedback, chat interactions, and support requests to improve profiling accuracy;
- 3) Explore the use of reinforcement learning models for adaptive segmentation.

Stage 3. Selecting and testing AI/ML models for student profiling:

- 1) Conduct experiments on the use of segmentation methods (K-Means, DBSCAN) to cluster students based on admission;
- 2) Develop predictive models (random forest, gradient boosting, neural networks) to predict student retention and academic achievement;
- 3) Model evaluation metrics such as F1 score, recall accuracy, and root mean square error (RMSE) will be used to evaluate the effectiveness of different approaches.

This requires considering ethical and data security issues. For AI, machine learning, and data analytics to be fully effective in education, access to student data is essential. However, it should be noted that the rights to educational data should remain with students, and educational institutions should implement robust data protection measures to prevent unauthorised access. Transparency in data collection, storage, and use is essential for building trust in AI-powered educational technologies while protecting students' digital rights, and these requirements need to be considered when collecting data and implementing segmentation and profiling models.

By implementing these three phases of research, we plan to develop data-driven segmentation and profiling models using machine learning and AI techniques to develop marketing strategies to attract and retain students, enabling personalised, effective, and ethical decision-making.

References

- Abyaa, A., Khalidi Idrissi, M., & Bennani, S. (2019). Learner modelling: Systematic review of the literature from the last 5 years. *Educational Technology Research and Development*, 67(5), 1105-1143. <https://doi.org/10.1007/s11423-018-09644-1>
- Agustino, D. P., Budaya, I. G. B. A., Harsemadi, I. G., Dharmendra, I. K., & Pande, I. M. S. A. (2023). Perbandingan algoritma DBSCAN dan affinity propagation dalam segmentasi pelanggan tenant inkubator bisnis. *Jurnal Sistem Informasi dan Komputer (SISFOKOM)*, 12(2). <https://doi.org/10.32736/sisfokom.v12i2.1682>
- Alkhayrat, M., Aljnidi, M., & Aljoumaa, K. (2020). A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA. *Journal of Big Data*, 7, Article 9. <https://doi.org/10.1186/s40537-020-0286-0>
- Arakh, R., Shoaib, R., Kriplani, S., Yadav, V., & Agrawal, S. (2024). Customer segmentation using machine learning and hybrid model. *International Journal of Scientific Development and Research*, 9(6). <https://ijsdr.org/papers/IJSDR2406110.pdf>
- Cao, L. (2024). Comparison of Machine Learning Methods in Customer Segment. *Highlights in Science, Engineering and Technology*, 92, 133-137. <https://doi.org/10.54097/c2s9m434>
- Cherednichenko, O., Ivashchenko, O., & Kirchmayer J. (2023). Towards item matching task: systematic review/ Smerom k úlohe priad'ovania položiek: systematická kontrola. *Public Administration and Regional Development*, 19(1), 81-98. <https://www.vsemba.sk/portals/0/Subory/vedecky%20casopis%2001%20-%202023%20-%2017102023.pdf>
- Chertchom, P. (2024). Unveiling Student Segments: Leveraging Clustering Analysis of Registration Data for Enhanced Recruitment Strategies. *Economics and Business Administration Journal Thaksin University*, 16(3), 181-196. <https://doi.org/10.55164/ecbajournal.v16i3.268195>
- Das, S., Nayak, J. (2022). Customer Segmentation via Data Mining Techniques: State-of-the-Art Review. In: Nayak, J., Behera, H., Naik, B., Vimal, S., Pelusi, D. (eds) *Computational Intelligence in Data Mining. Smart Innovation, Systems and Technologies*, 281. Springer, Singapore. https://doi.org/10.1007/978-981-16-9447-9_38

Daware, P. (2024). Regression and ensemble models for analysing students' key performance factors. *International Journal for Research in Applied Science & Engineering Technology*, 12(11). <https://doi.org/10.22214/ijraset.2024.65404>

Gowda, A. R. et al. (2025). Tailored elective recommendations using collaborative filtering and machine learning. *International Journal of Scientific Research in Engineering and Management*, 9(1), 1-9. <https://doi.org/10.55041/IJSREM40661>

Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, 4(4), 134-147. <https://doi.org/10.1016/j.ijis.2020.09.001>

Guo, Y., Chen, Y., Xie, Y., & Ban, X. (2022). An effective student grouping and course recommendation strategy based on big data in education. *Information*, 13(4), 197. <https://doi.org/10.3390/info13040197>

Hamidi, H., & Haghi, B. (2024). An approach based on data mining and genetic algorithm to optimising time series clustering for efficient segmentation of customer behavior. *Computers in Human Behavior Reports*, 16, 100520 <https://doi.org/10.1016/j.chbr.2024.100520>

Hassan, Y. M., Elkorany, A., & Wassif, K. (2022). Utilising social clustering-based regression model for predicting student's GPA. *IEEE Access*, 10, 48948-48963. <https://doi.org/10.1109/ACCESS.2022.3172438>

Hemsley-Brown, J. (2016). Higher Education Market Segmentation. In: Shin, J., Teixeira, P. (eds) *Encyclopedia of International Higher Education Systems and Institutions*. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9553-1_33-1

Hicham, N., & Karim, S. (2022). Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation using Clustering Ensemble and Spectral Clustering. *International Journal of Advanced Computer Science and Applications*. <https://doi.org/10.14569/ijacsa.2022.0131016>

Hill, L. H., & Conceição, S. C. O. (2024). Intelligent learning management systems (ILMS): Implications for teaching and learning. *eLearn*, 2024(12), Article 1. <https://doi.org/10.1145/3709442.3702011>

John, J. M., Shobayo, O., & Ogunleye, B. (2023). An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics*, 2(4), 809-823. <https://doi.org/10.3390/analytics2040042>

Julian, A., & R, H. S. (2024). Optimising customer segmentation through machine learning. In *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)* (413–416). IEEE. <https://doi.org/10.1109/IC2PCT60090.2024.10486699>

Kadu, R. K., Assudani, P. J., Mukewar, T., Kapgate, J., & Bijekar, R. (2024). Student placement prediction and skill recommendation system using machine learning algorithms. In *2024 International Conference on Inventive Computation Technologies (ICICT)*. IEEE. <https://doi.org/10.1109/ICICT60155.2024.10544738>

Katragadda, V. (2022). Dynamic customer segmentation: Using machine learning to identify and address diverse customer needs in real-time. *Iconic Research and Engineering Journals*, 5(10). <https://www.irejournals.com/formatedpaper/1703349.pdf>

Kiran, D., & Ashwini, C. (2024). Customer segmentation using machine learning. *International Journal of Scientific Research in Engineering and Management*, 8(7), 1–13. <https://doi.org/10.55041/IJSREM36658>

Lewaaelhamd, I. (2023). Customer Segmentation Using Machine Learning Model: An Application of RFM Analysis. *Journal of Data Science and Intelligent Systems*, 2(1), 29-36. <https://doi.org/10.47852/bonviewJDSIS32021293>

Mim, S. S., & Logofatu, D. (2022). A cluster-based analysis for targeting potential customers in a real-world marketing system. In *2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP)* (159–166). IEEE. <https://doi.org/10.1109/ICCP56966.2022.10053985>

Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering Analysis for Classifying Student Academic Performance in Higher Education. *Applied Sciences*, 12(19), 9467. <https://doi.org/10.3390/app12199467>

Nabeel Mustafa, S.M., Akhtar, A., Peter Noronha, J.T., Salman, M., & Baig, M.A. (2023). Customer Segmentation using Machine learning Techniques. *2023 International Multi-disciplinary Conference in Emerging Research Trends (IMCERT)*, 1, 1-7. <https://doi.org/10.1109/IMCERT57083.2023.10075194>

Phyo, T., & Uttama, S. (2023). Unveiling Patterns in the Night Market: A Machine Learning Approach to Customer Segmentation and Market Basket Analysis. *2023 7th International Conference on Information Technology (InCIT)*, 514-519. <https://doi.org/10.1109/InCIT60207.2023.10413047>

Potluri, C. S., Rao, G. S., Kumar, L. V. R. M., Allo, K. G., Awoke, Y., & Seman, A. A. (2024). Machine learning-based customer segmentation and personalised marketing in financial services. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)* (1570–1574). IEEE. <https://doi.org/10.1109/IC3SE62002.2024.10593143>

Rahmet, S., & Gürkan, F. (2024). Integration of pre-processing and machine learning methods for improved customer segmentation. In *2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP)* (1–6). IEEE. <https://doi.org/10.1109/IDAP64064.2024.10710767>

Rai, K. D., Kasinathan, V., Mustapha, A. (2024). Data-Driven Student Recruitment Strategies in Private Higher Educational Institutions. In: Uden, L., Ting, I.H. (eds) *Knowledge Management in Organisations. KMO 2024. Communications in Computer and Information Science*, vol 2152. Springer, Cham. https://doi.org/10.1007/978-3-031-63269-3_14

Ramirez-Montoya, M. S., Morales-Menendez, R., Tworek, M., Escobar, C. A., Tariq, R., & Tenorio-Sepulveda, G. C. (2024). Complex competencies for leader education: artificial intelligence analysis in student achievement profiling. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2378508>

Sathish, C., Shanmugam, K., Kumar, K. P., Sivamoorthy, S., & Ajay, K. S. (2024). Dynamic landscape of higher education: The intersection of student skills and industry demand. *Indian Journal of Computer Science and Technology*, 3, 73-79. <https://doi.org/10.59256/indjcs.20240302009>

Segovia-García, N. (2024). Optimising student support: A review of the use of AI chatbots in higher education. *European Public & Social Innovation Review*. <https://doi.org/10.31637/epsir-2024-324>

Shaikh, A. F., & Khedkar, S. A. (2023). Clustering approach to high-dimensional data for banking customer segmentation. *International Journal of Advanced Research in Computer and Communication Engineering*, 12(8). <https://doi.org/10.17148/IJARCC.2023.12805>

Shilbayeh, S.A., & Abonamah, A.A. (2021). Predicting student enrollments and attrition patterns in higher educational institutions using machine learning. *Int. Arab J. Inf. Technol.*, 18, 562-567. <https://doi.org/10.34028/18/4/8>

Sobral, S. R., & de Oliveira, C. F. (2021). Clustering algorithm to measure student assessment accuracy: A double study. *Big Data and Cognitive Computing*, 5(4), 81. <https://doi.org/10.3390/bdcc5040081>

Song, R., Pang, F., Jiang, H., & Zhu, H. (2024). A machine learning-based method for constructing group profiles of university students. *Heliyon*, 10(7), e29181. <https://doi.org/10.1016/j.heliyon.2024.e29181>

Thitipomdharma, T., & Khonthaoagdee, S. (2024). K-Means and Decision Trees in Action: Unveiling Customer Segments and Personalised Promotions for Credit Card Companies. *2024 8th International Conference on Business and Information Management (ICBIM)*, 85-89. <https://doi.org/10.1109/ICBIM63313.2024.10823628>

Tran, H. D., Le, N., & Nguyen, V.-H. (2023). Customer churn prediction in the banking sector using machine learning-based classification models. *Interdisciplinary Journal of Information, Knowledge, and Management*, 18, 87-105. <https://doi.org/10.28945/5086>

Wang, Z. (2024). Customer Segmentation Based on Machine Learning Methods. *Highlights in Science, Engineering and Technology*, 92, 126-132. <https://doi.org/10.54097/g70xqb16>

Wani, A., Priyanka, M., & Prasath, R. (2023). Unleashing customer insights: Segmentation through machine learning. In *2023 World Conference on Communication & Computing (WCONF)* (1–5). IEEE. <https://doi.org/10.1109/WCONF58270.2023.10235136>

Widianti, R., Surono, S., & Ibraheem, K. I. (2024). Handling Noise Data with PCA Method and Optimization Using Hybrid Fuzzy C-Means and Genetic Algorithm. *JUITA: Jurnal Informatika*, 12(2), 141-147. <https://doi.org/10.30595/juita.v12i2.21765>

Yildirim, Y., & Celepcikay, A. (2021). Artificial Intelligence and Machine Learning Applications in Education. *Eurasian Journal of Higher Education*, (4), 1-1. <https://doi.org/10.31039/ejohe.2021.4.49>

Funding: This research was funded by International grant project of VŠEM No. 1/2024_M – Management of the processes of recovery and sustainable development of public and economic law entities of the EU Member States, in particular Slovakia and Ukraine.

Author Contributions: Conceptualisation: *Ivashchenko O., Lelovský M., Kopp A.*; methodology: *Ivashchenko O., Kopp A., Shmatko O.*; data analysis: *Ivashchenko O., Shmatko O., Cibák L.*, writing—original draft preparation: *Ivashchenko O., Lelovský M.*; writing; review and editing: *Ivashchenko O., Lelovský M., Cibák L.*; visualisation: *Ivashchenko O., Kopp A.*. All authors have read and agreed to the published version of the manuscript

Oksana IVASHCHENKO, PhD is an Associate Professor at the Department of Software Engineering and Management Intelligent Technologies of National Technical University “Kharkiv Polytechnic Institute” (Ukraine). Research interests: data mining, artificial intelligence and machine learning.

ORCID: <https://orcid.org/0000-0003-3636-3914>

Dr. h. c. Ing. Mário LELOVSKÝ is a doctoral student at the Department of E-government and Digital Technologies at Bratislava University of Economics and Management (BUEM) (Slovakia). He is the First Vice-President of ITAS, First Vice-President of the National Union of Employers, and Chairman of the Digital Coalition – National Coalition for Digital Skills and Jobs of the Slovak Republic. Research interests: digitalisationdigitalisation in the business and public administration.

ORCID: <https://orcid.org/0009-0003-7163-9602>

Andrii KOPP, PhD is an Associate Professor and Head of the Department of Software Engineering and Management Intelligent Technologies of National Technical University “Kharkiv Polytechnic Institute” (Ukraine). Research interests: models and information technologies for analysing and improving the quality of the business process structure of complex systems.

ORCID: <https://orcid.org/0000-0002-3189-5623>

Oleksandr SHMATKO, PhD is an Associate Professor and Deputy Head of the Department of Software Engineering and Management Intelligent Technologies of National Technical University “Kharkiv Polytechnic Institute” (Ukraine). Research interests: cyber security, system analysis, risk management, strategical management.

ORCID: <https://orcid.org/0000-0002-2426-900X>

Dr. h. c. doc. Ing. Ľuboš CIBÁK, PhD., MBA is an associate professor and Head of the Department of E-government and Digital Technologies at Bratislava University of Economics and Management (BUEM) (Slovakia). Research interests: e-government, digital technologies, information systems, application of information technologies in education and public administration.

ORCID: <https://orcid.org/0000-0003-3881-7924>

Copyright © 2026 by author(s). Publishing rights by Entrepreneurship and Sustainability Center

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access