

DOI: 10.37943/23JOZI8138**Gulzhan Soltan**

Candidate of Technical Sciences, Associate Professor, Department of Computer Engineering
gulzhan.soltan@astanait.edu.kz; orcid.org/0000-0002-1603-7524
Astana IT University, Kazakhstan

Zharkyn Bekkaliyev

Master's Degree student, Department of Computational and Data Science
z.bekkaliyev.02@gmail.com; orcid.org/0009-0008-2160-5427
Astana IT University, Kazakhstan

Alen Issayev

Bachelor's Degree student, Department of Computational and Data Science
alenkz20001@gmail.com; orcid.org/0009-0006-5185-0250
Astana IT University, Kazakhstan

Nazerke Akzhol

Master's Degree student, Department of Computer Engineering
nmaldybayeva@gmail.com; orcid.org/0009-0002-7323-9760
Astana IT University, Kazakhstan

Zhuldyz Basheyeva

PhD, Assistant professor, Department of Computer Engineering
zhuldyz.basheyeva@astanait.edu.kz; orcid.org/0000-0001-9605-2101
Astana IT University, Kazakhstan

APPLYING DATA ANALYTICS AND BI SYSTEMS TO BUILD A STUDENT DIGITAL PROFILE: THE CASE OF ASTANA IT UNIVERSITY

Abstract: Modern challenges of the digital transformation of education require the development of new approaches to assessing academic success and monitoring students' educational trajectories. This study presents a functional model of the data analytics system and visualization of the digital profile of a graduate of Astana IT University (AITU), based on the Integrated IGPA (Integrated Grade Point Average) indicator, which combines the academic, research, and social achievements of students. The aim of the work is to create a system of analytics, visualization, and interpretation of data reflecting the comprehensive development of students and their readiness for professional activity. The theoretical part examines modern approaches to educational analytics in higher education. A critical analysis of scientific sources, including research on learning analytics, educational data mining, and the formation of digital profiles of students, was carried out. The emphasis on technical aspects and insufficient connection with educational practice reveals the main limitations of the existing models.

The empirical part uses anonymized data from AITU students for 2022–2024, covering the indicators of Grade Point Average (GPA), Indicators of Research-Oriented Study (iROS), and Social Competition Indicators (SSCI). Dashboards built with the help of Power BI made it possible to visualize and interpret educational trajectories. The use of machine learning algorithms (K-means clustering, PCA analysis) ensured the typologization of student profiles. Using Python and the scikit-learn, seaborn, and pandas' libraries allowed us to deeply explore the relationships between IGPA components.

The results of the study demonstrate the possibilities of personalized academic support, strategic management of educational processes, and increased transparency of student achievement. The developed model can serve as a basis for making managerial decisions and improving the quality of educational programs in the context of digital transformation.

The proposed approach can be scaled and adapted to other educational institutions, regardless of their size and specialization. Flexibility in integrating additional indicators reflecting the unique goals and values of a particular educational environment facilitates the model's versatility.

Keywords: digital graduate profile, integral GPA, educational analytics, data visualization, digital transformation of education, automation of data analysis, interactive dashboards.

Introduction

Modern higher education is undergoing a stage of intensive digitalization, in which data processing and analysis technologies play a key role. There is a rapid development of approaches to the analysis of educational data, including using artificial intelligence, machine learning, and business intelligence methods to improve the quality of teaching and academic performance. The integration of learning analytics (LA) and business intelligence (BI) systems allows not only to improve the quality of teaching and personalize learning but also to improve the management of educational processes.

The systematic review carried out in [1], [3] summarizes the use of AI in teacher training, emphasizing its strengths in adaptive learning and automatic feedback. To create intelligent educational platforms, it is necessary to integrate deep learning into educational big data analytics systems. Thus, learning analytics and educational data mining are becoming important tools for monitoring academic performance and improving teaching effectiveness [4], [5], [6]. Despite the favorable theoretical background, most research on this topic is limited to analyzing the accuracy of models, not including their impact on educational decisions. There is no assessment of the stability of algorithms in real educational conditions (for example, with incomplete data, input errors, etc.).

Many papers focus on the technical side (accuracy, metrics, ROC curves), while learning success and the impact on student behavior are often left out of consideration. The prediction methods used are based on a narrow set of metrics (academic performance estimates, LMS completion time), which limits reliability and leads to overfitting models. The authors propose specific models for predicting academic success; however, these models suffer from methodological issues due to insufficient validation: the sample sizes are small, and there is a lack of control over related variables such as motivation, engagement, and course context [9], [10], [13].

Recently, one of the significant trends has been the development of digital student profiles. The formation of a graduate's digital profile is one of the conditions for the implementation of the main directions of the Bologna process and the requirements of the modern labor market. Therefore, the development of a graduate's digital profile system based on integral GPA (GPA) indicators should meet the needs of stakeholders and all interested parties. Studies [2], [7], [8] focus on modeling graduate competencies and student professional mobility, emphasizing the importance of digital tools in shaping the educational trajectory. These jobs connect a digital profile with the possibility of personalized support, career planning, and improving the quality of the educational process.

The implementation of the integrated GPA system for students at Astana IT University (AITU) was launched in 2022. The relevant documents, regulations, and rules have been developed. IGPA reflects the achieved level of students' academic achievements, their aptitude for research, and their increased social responsibility towards society. The IGPA consists of three assessment indicators:

- GPA (Grade Point Average): A weighted average assessment of a student's academic achievement over a given period;
- iROS (Indicators of Research-Oriented Study): Indicators reflecting the student's research skills and achievements acquired during academic and extracurricular activities;
- SCI (Social Competition Indicators): Indicators that characterize social competencies, including participation in community, volunteer, and creative activities.

The introduction of the integral GPA is aimed at the comprehensive development of students and increasing their competitiveness, and there are certainly a number of advantages to implementing the IGPA calculation system for students. Comprehensive assessment takes into account not only academic achievements but also the research and social activity of students. They provide motivation for development and encourage students to participate in scientific research and social activities. They provide an objective assessment of students' achievements in various fields.

However, the successful implementation of this system requires careful consideration of assessment methods and consideration of the views of all stakeholders.

An example of practical implementation is research using Power BI as a business intelligence tool [16], [17] and [20] demonstrate how visualization of key indicators using BI dashboards helps to increase the transparency of management decisions and monitor academic results. Research [11], [12] and [19] supports these approaches by proposing models for the development of KPI-oriented dashboards in universities. Most of the research focuses on administrative analytics (finance, KPIs, attendance) rather than in-depth analysis of the learning process. BI tools often act as a visualization interface rather than a full-fledged analytical module. There is no synergy between BI and AI, and LA systems operate separately, which reduces efficiency.

The presented studies demonstrate the high potential of data analytics and AI in education, but at the same time there are no comprehensive models combining digital profiles, learning analytics, and BI visualization [14], [15] and [18].

Scientific novelty of the project.

The project offers a comprehensive approach to the analysis and visualization of educational data, representing a unique solution for the higher education sector. A functional data analytics model has been developed, incorporating processes for processing and visualizing the integrated GPA (IGPA). This model has no analogues in the current practice of universities in Kazakhstan.

For the first time, an integrated approach is used to analyze the relationships between various components of IGPA (GPA, iROS, SSCI) through mathematical modeling, allowing for the identification of hidden patterns and trends.

The project addresses not only the practical tasks of managing the educational process at AITU but also lays a scientific foundation for further research in the field of educational data analytics. It can be adapted and applied in other educational institutions, thereby contributing to the digital transformation of higher education.

Methods and Materials

This study uses a combination of educational analytics, data visualization, and analysis methods to thoroughly evaluate students' academic success, understand their learning behaviors, and create data-driven profiles of students. The research includes two complementary areas: firstly, the construction of a visual analytical model based on business intelligence tools focused on system aggregation and interpretation of indicators of integral academic success; secondly, the use of quantitative analysis methods and machine learning algorithms to identify patterns in educational data. The research has two main parts: first, creating a visual model

using business intelligence tools to combine and understand indicators of overall academic success; second, applying quantitative analysis methods and machine learning algorithms to find patterns in educational data (Fig.1).

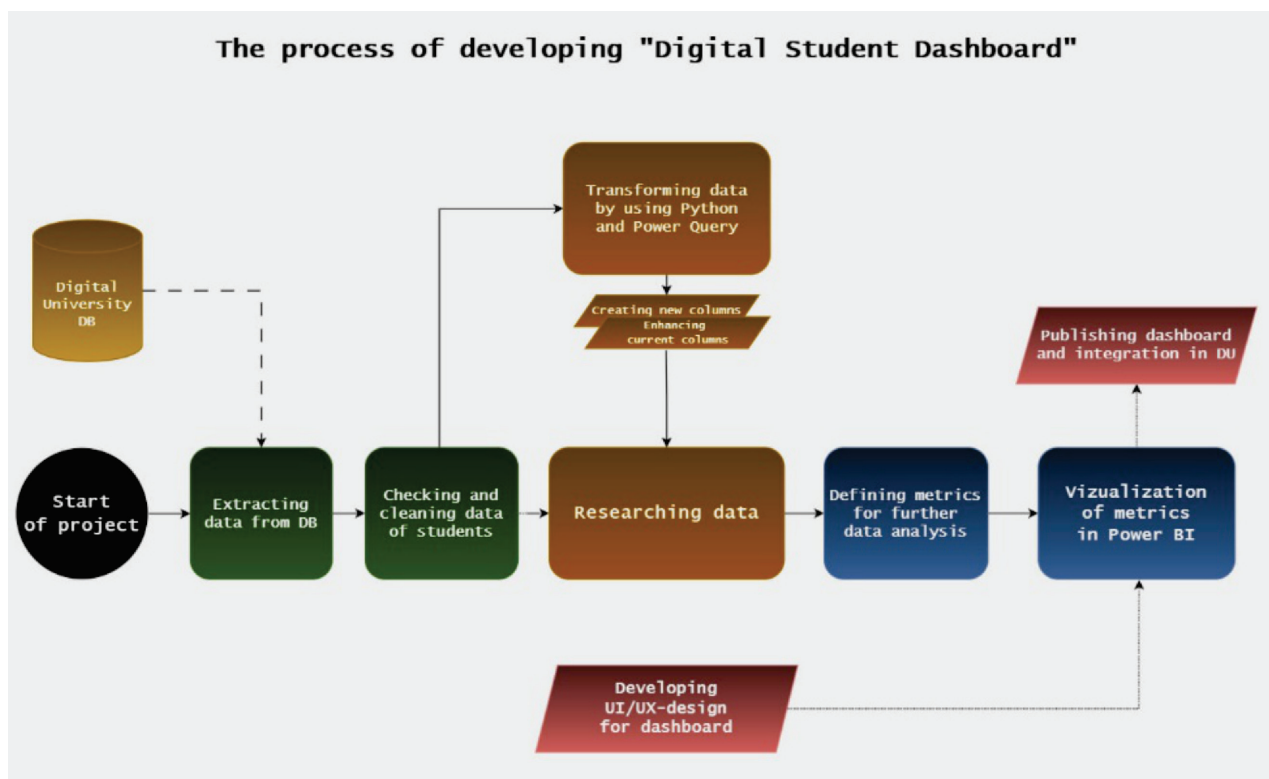


Figure 1. Stages of developing “Digital Student Dashboard”

Anonymized data from Astana IT University students covering academic cohorts in 2022-2024 was used as an empirical base. The data structure included both quantitative performance indicators in key disciplines and metrics characterizing research (iROS) and social (SSCI) activity, as well as an integrated IGPA score calculated according to the approved formula: $IGPA = 0.5 \times GPA + 0.35 \times iROS + 0.15 \times SSCI$. Additionally, the parameters reflecting the educational program, department, year of admission, and student status were involved. The data was extracted from the Digital University information system and provided as structured CSV files, which were integrated using the key student_id identifier.

The initial data processing and visualization were performed in a Power BI environment using Power Query. We implemented the relationships between the tables and formed measures, aggregating and normalizing indicators at the level of individual students and educational programs. The built dashboards included visual representations of academic performance, research engagement, and social engagement, as well as their distribution by year and field of study. This approach allowed us to create a visual basis for preliminary analysis and identification of structural patterns in academic trajectories.

The next stage of the study used descriptive and differential statistics methods, as well as machine learning algorithms. Specifically, we conducted a correlation analysis using the Pearson coefficient to evaluate the connection between academic performance and IGPA components. Further, the students were stratified by the level of final grades (low, medium, high), which revealed differences in indicators of research and social activity between the groups. The main analytical task was to build a cluster model using the K-means algorithm, optimized by the elbow method and verified through silhouette analysis. The values of GPA, iron, SKY, and

final scores were used as input features for clustering. The Principal Component Analysis (PCA) method was used to visualize the multidimensional space and interpret the cluster structure, allowing for the reflection of student segmentation in a two-dimensional space (Fig.2).

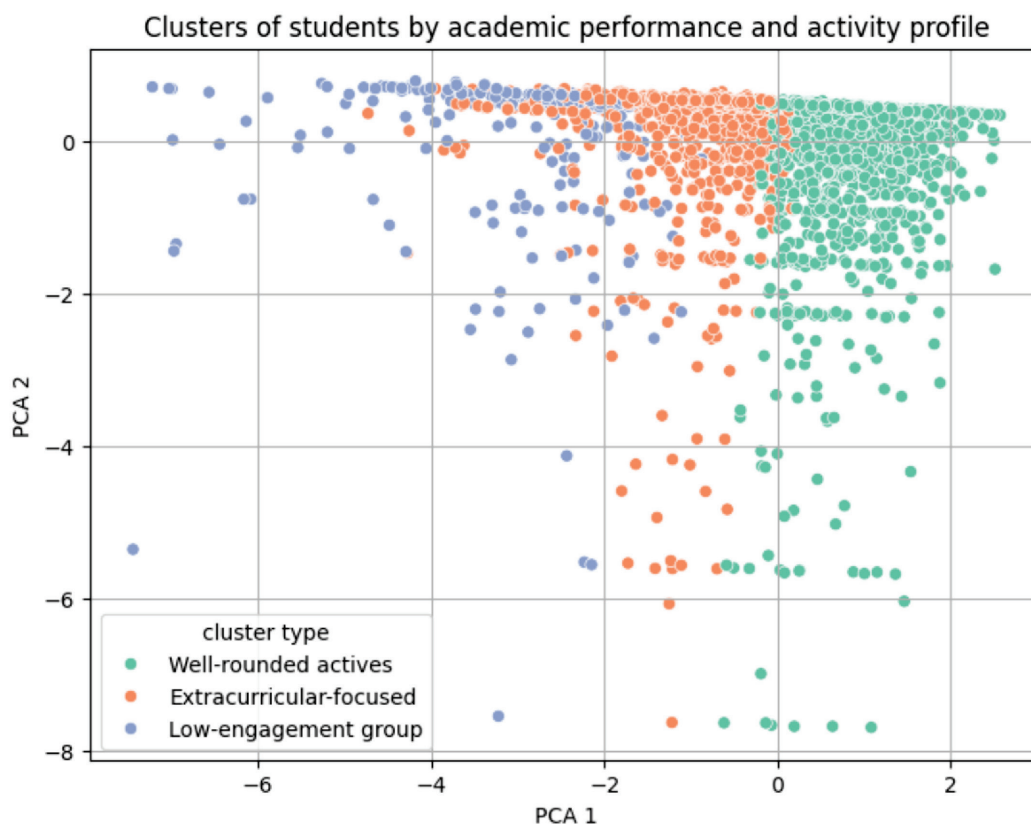


Figure 2. Cluster of students by academic performance and activity profile.

Mathematical model

To enable a deeper and more accurate analysis of student performance and to identify individual strengths in learning, we conducted an analysis of academic disciplines. In order to assess student performance in specific subjects, it is necessary to determine the type of discipline, in other words, the cluster to which it belongs.

The development of natural language processing (NLP) and semantic analysis methods opens up new opportunities for automated and more precise clustering. In this study, word embedding techniques and machine learning algorithms for the semantic clustering of course titles are applied, which allows to uncover hidden relationships between disciplines.

All stages of the analysis were implemented using Python tools (pandas, numpy, matplotlib, seaborn, and scikit-learn libraries) in the PyCharm development environment. When developing analytical procedures, special attention was paid to reproducibility, interpretability, and visual reporting, which meets the requirements for applied research in the field of educational analytics. The approach used provided an opportunity for both aggregated monitoring at the educational program level and personalized analysis of the educational dynamics of individual students. The results obtained formed the basis for the typologization of student profiles and the formulation of recommendations for the individualization of academic support in the digital educational environment.

The development of the digital graduate profile at AITU is based on a formalized model that reflects three key components of a student's success: academic, research, and social. For

comprehensive evaluation, an integrated indicator – **IGPA** – is used, calculated as a weighted sum of three metrics:

$$IGPA_i = \alpha \cdot GPA_i + \beta \cdot iROS_i + \gamma \cdot SSCI_i \quad (1)$$

where:

$IGPA_i$ – integrated performance score of student i ,

GPA_i – weighted average of academic performance,

$iROS_i$ – index of research activity,

$SSCI_i$ – index of social engagement,

$\alpha = 0.5$, $\beta = 0.35$, $\gamma = 0.15$ – weighting coefficients reflecting the priorities of AITU's educational strategy.

Each component is normalized on a 0 to 1 scale before being substituted into the formula, ensuring comparability of metrics with different scales and natures. To support further analysis and student segmentation based on IGPA, a clustering model was built using the **k-means algorithm (KMeans)**. Students were represented as feature vectors consisting of:

$$X_i = [GPA_i, iROS_i, SSCI_i] \quad (2)$$

The course titles underwent basic preprocessing, which included converting text to lower-case and removing special characters. It is important to note that more advanced steps, such as stop-word removal or lemmatization, were not applied, as the **SentenceTransformer** model used in this study effectively captures the semantic meaning of text, including the contextual role of stop words.

A key stage of the process involves transforming the textual course titles into dense numerical vectors, or embeddings. For this purpose, a pre-trained **SentenceTransformer** model called “**all-MiniLM-L6-v2**” was used. This model generates high-quality semantic representations of sentences, where words and phrases with similar meanings are located close to each other in the vector space. Each course embedding serves as its **semantic fingerprint**. Clustering enables the formation of student groups with similar profiles in terms of academic, research, and social activities. The optimal number of clusters **k** was determined using the **elbow method**:

$$W(k) = \sum_{j=1}^k \sum_{x \in C_j} \|x - \mu_j\|^2 \quad (3)$$

where C_j is the set of students in the j -th cluster, and μ_j is the cluster centroid. Minimization of the within-cluster variance $W(k)$ ensured the stability of the clustering structure.

The same algorithm was also used to identify clusters of academic disciplines. In this study, $k = 5$ clusters were defined, which made it possible to divide the courses into five main semantic groups (Fig. 3).

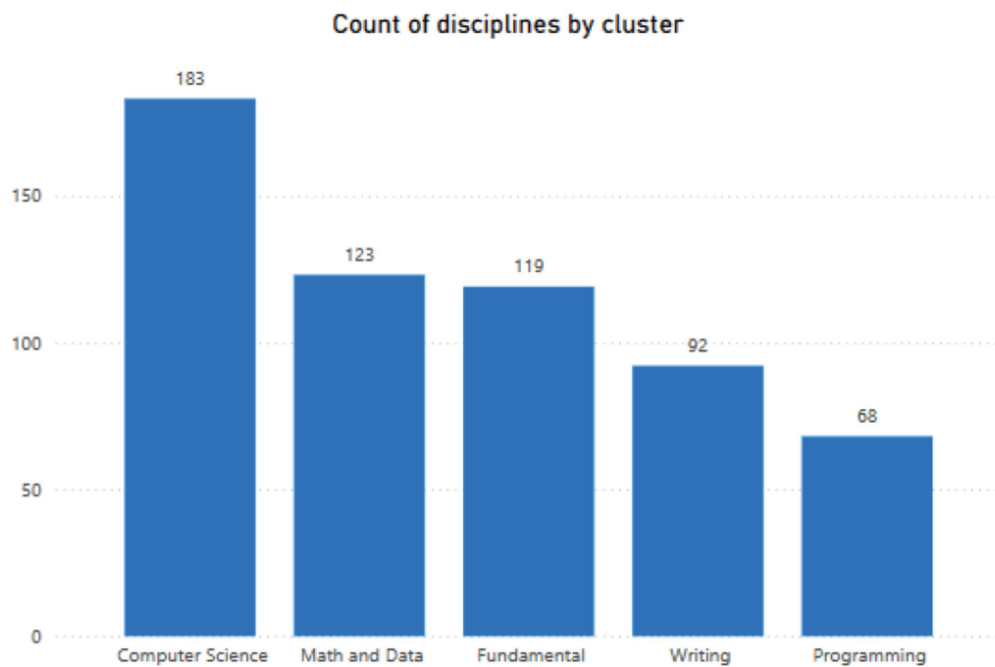


Figure 3. Count of disciplines by cluster.

To interpret the results in a reduced feature space, **Principal Component Analysis (PCA)** was applied. This method reduces the three-dimensional space to two dimensions while preserving the maximum possible variance in the data:

$$Z = X \cdot W \quad (4)$$

where X is the matrix of normalized features, and W is the matrix of eigenvectors corresponding to the largest eigenvalues of the covariance matrix.

The proposed model provides a mathematical foundation for constructing personalized digital student profiles, visualizing their educational trajectories, and generating recommendations for individualized academic support.

The data obtained from the Digital University system was successfully processed and analyzed using Power BI, including the built-in Power Query tool for data transformation. As part of the analysis, various measures were created, such as:

1. 1stCourseCount, 2ndCourseCount, 3rdCourseCount, GraduateCount
2. IROS students count, IROS 2 total points, IROS avg GPA
3. Measures related to individual student selection (Selected Student ID)
4. Indicators for assessing students' social engagement (SSCI), including:
SSCI students, SSCI Avg grade, SSCI AvgGradeByCourse,
SSCI Total hours, SSCI TotalCreditsByCourse,
SSCI TotalHoursByCourse, SSCI TotalPointsByCourse
5. General metrics such as StudentStatusMeasure, Total students, and TotalHoursSSCIper-Student

Initially, data were loaded into Power BI, where relationships between tables were modeled using the student_id column. This allowed for the integration of information from the Transcript, SSCI Data, and IROS Data tables, enabling comprehensive assessment of student academic performance, social engagement, and research activity.

Data transformation through Power Query further enriches the dataset with additional attributes such as year of admission, department affiliation, and educational program, thus enabling deeper analytical insights.

The processed data formed the basis for the creation of interactive reports and visualizations in Power BI, allowing for detailed exploration of the relationships among academic indicators, social activity, and research achievements of students.

Development of Interactive Dashboards

Based on the data obtained from the Digital University system, an interactive dashboard titled "Digital Graduate Profile" was developed. It visualizes key performance indicators related to student activities. The total number of students represented on the dashboard is 9,785.

Overview. The dashboard includes an "Overview" section that provides general information on the distribution of students based on their involvement in research activities (IROS) and social engagement (SSCI). According to the data, 26% of students (2,529 individuals) participate in IROS, while 70% (6,846 individuals) are involved in SSCI. These indicators are presented using a **pie chart**, clearly illustrating the proportions (Fig. 4).

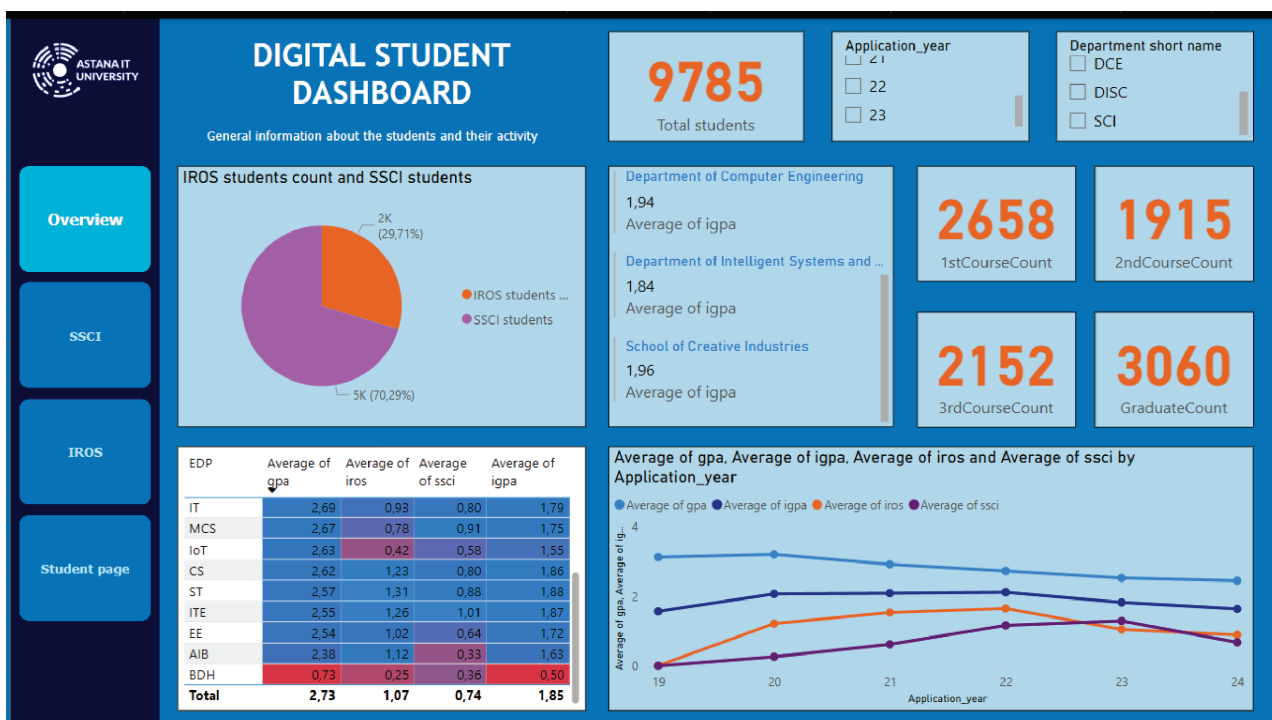


Figure 4. "Overview" page.

Average Indicators by Educational Programs (EDP). The table in the "Overview" section displays the average values of key metrics across different Educational Degree Programs (EDP), providing detailed analytics for each program. It also includes visualizations of the number of enrolled and graduated students.

Dynamics of Indicators by Year of Admission. The dashboard features a chart that displays the **average values of GPA, IGPA, IROS, and SSCI** by year of admission (from 2019 to 2024).

It includes filters by admission year (19, 20, 21, 22, 23, 24) and department (DCDS, DISCB, DCE, SCI), offering flexibility for in-depth analysis. The interactive interface facilitates real-time monitoring and comparison of student performance metrics, supporting data-driven decision-making in the educational process.

Results

The results of the study are based on the analysis of anonymized student data from Astana IT University for the years 2022–2024. Data was collected and processed for the three main components of the integrated IGPA assessment: GPA (academic performance), iROS (research activity), and SSCI (social engagement).

During the initial data processing stage, the information was aggregated and visualized using Power BI, enabling the tracking of indicator changes across cohorts and educational programs.

The dashboard includes individual pages for each key indicator, such as SSCI, which reflects students' social engagement. This section presents both historical and current performance metrics, providing departments with valuable insights into strategic planning and informed decision-making regarding future activities (Fig. 5).

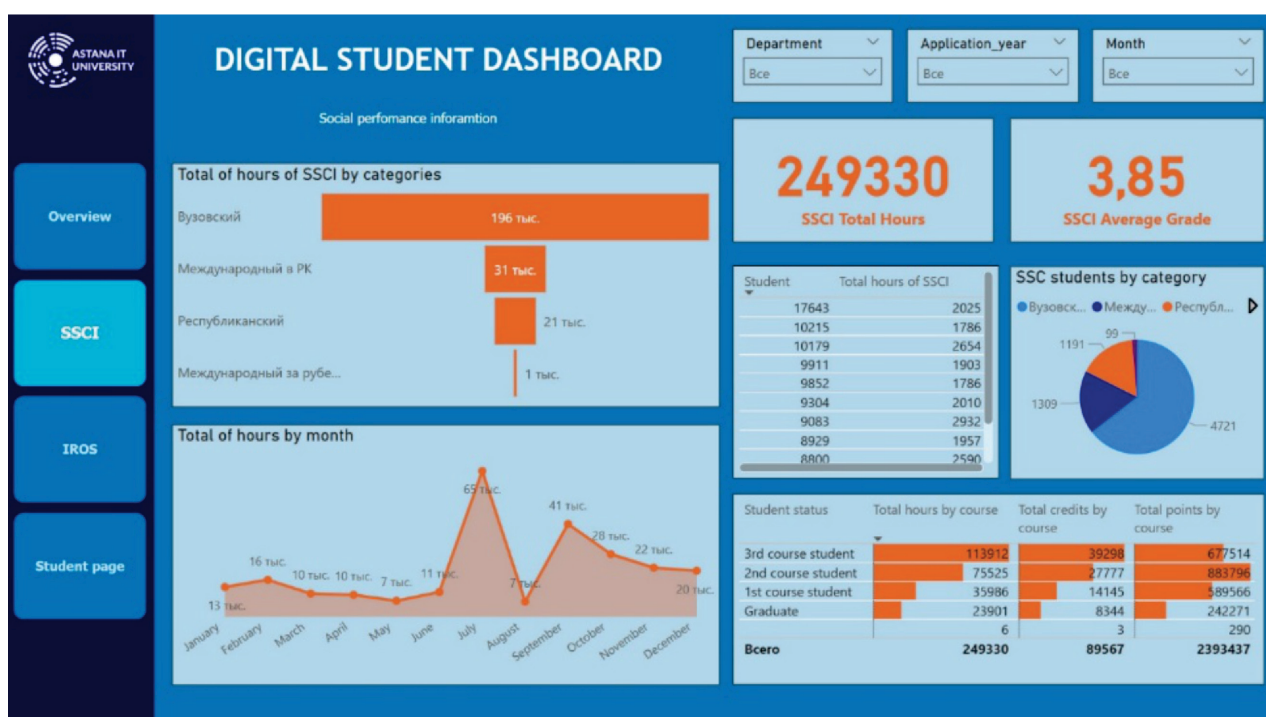


Figure 5. "SSCI" page.

A similar structure applies to the IROS indicator, for which a separate page is provided. This page enables in-depth analysis through visualizations and interactive filters, facilitating a more detailed exploration of students' research activity (Fig. 6).

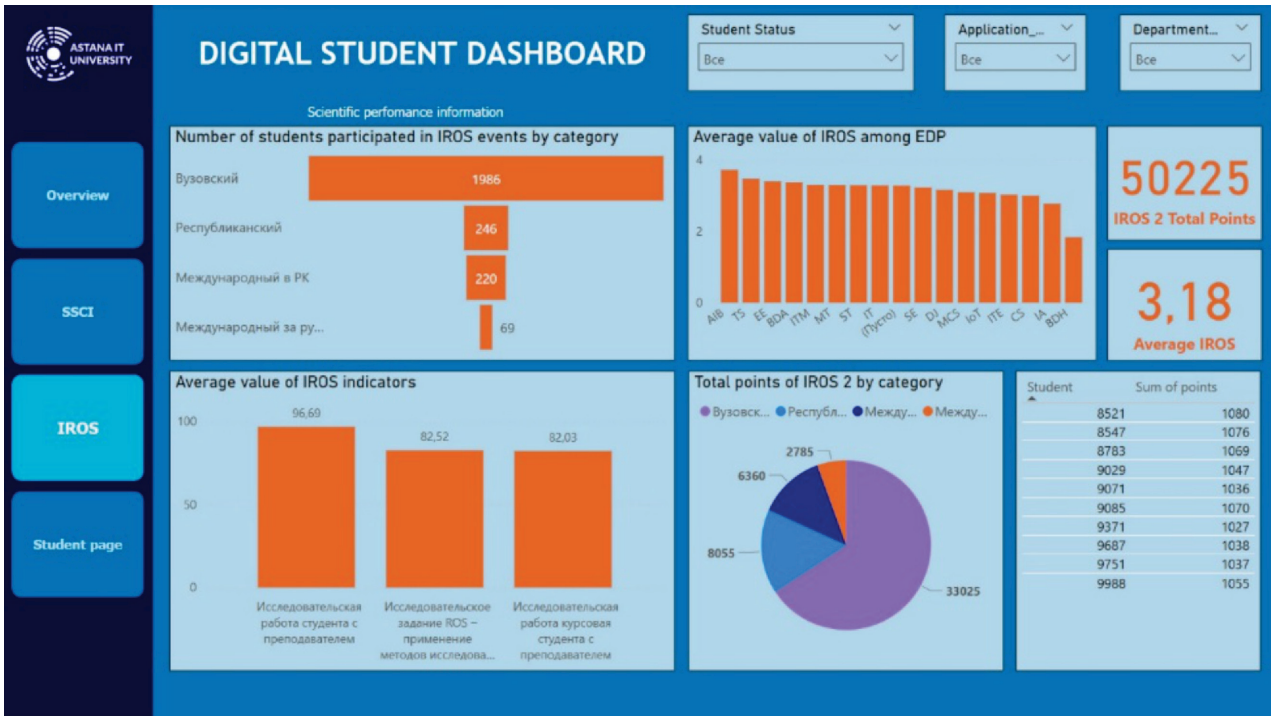


Figure 6. "IROS" page.

Following this is the Student Page, which presents a detailed profile of an individual student, including their participation in extracurricular events, research involvement, and academic achievements across various disciplines (Fig 7).

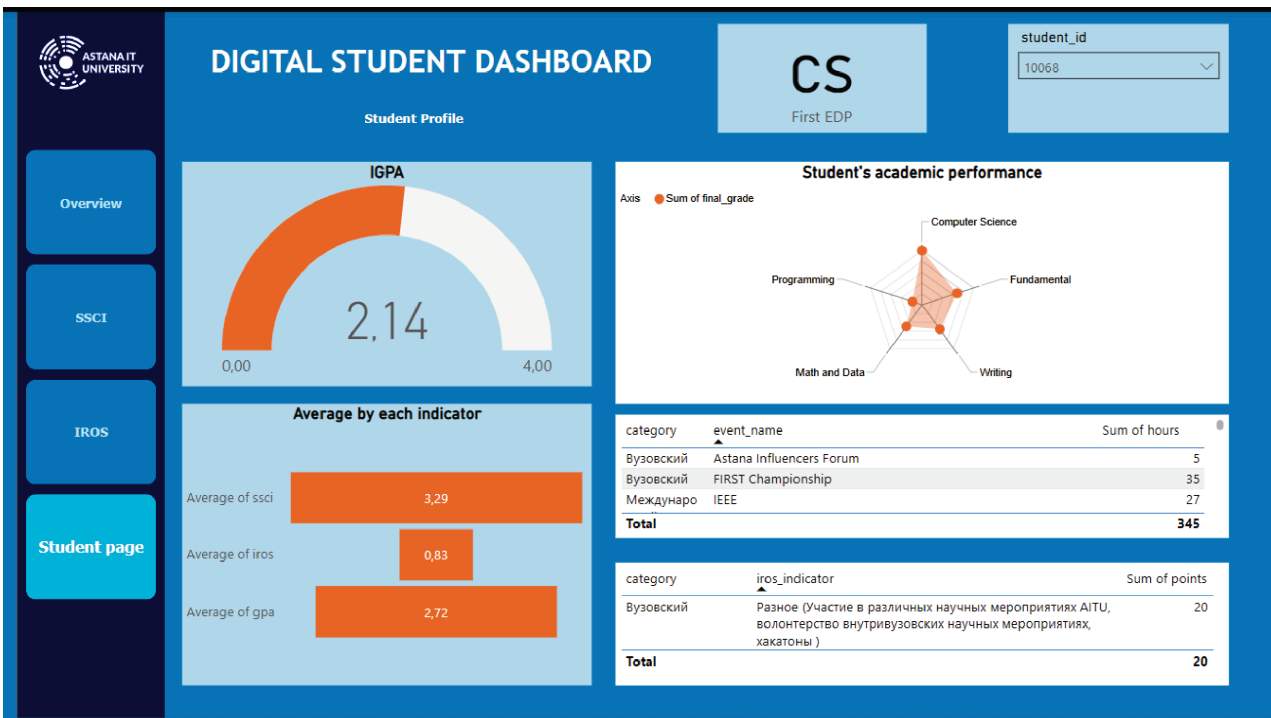


Figure 7. "Student page".

Correlation analysis revealed significant positive relationships between GPA and iROS, as well as between GPA and SSCI, indicating that academic performance is interconnected with other aspects of student activity. Based on these findings, clustering was performed using the KMeans algorithm, optimized via the elbow method. As a result, three stable clusters of students were identified, each with distinct profiles of engagement and performance. The Principal Component Analysis (PCA) method was used to visualize differences between clusters in a two-dimensional feature space. Interactive dashboards built in Power BI illustrate the distribution and dynamics of IGPA indicators, as well as comparative analysis across faculties and academic years. These results provide empirical grounds for the development of a digital graduate profile (Fig. 8).

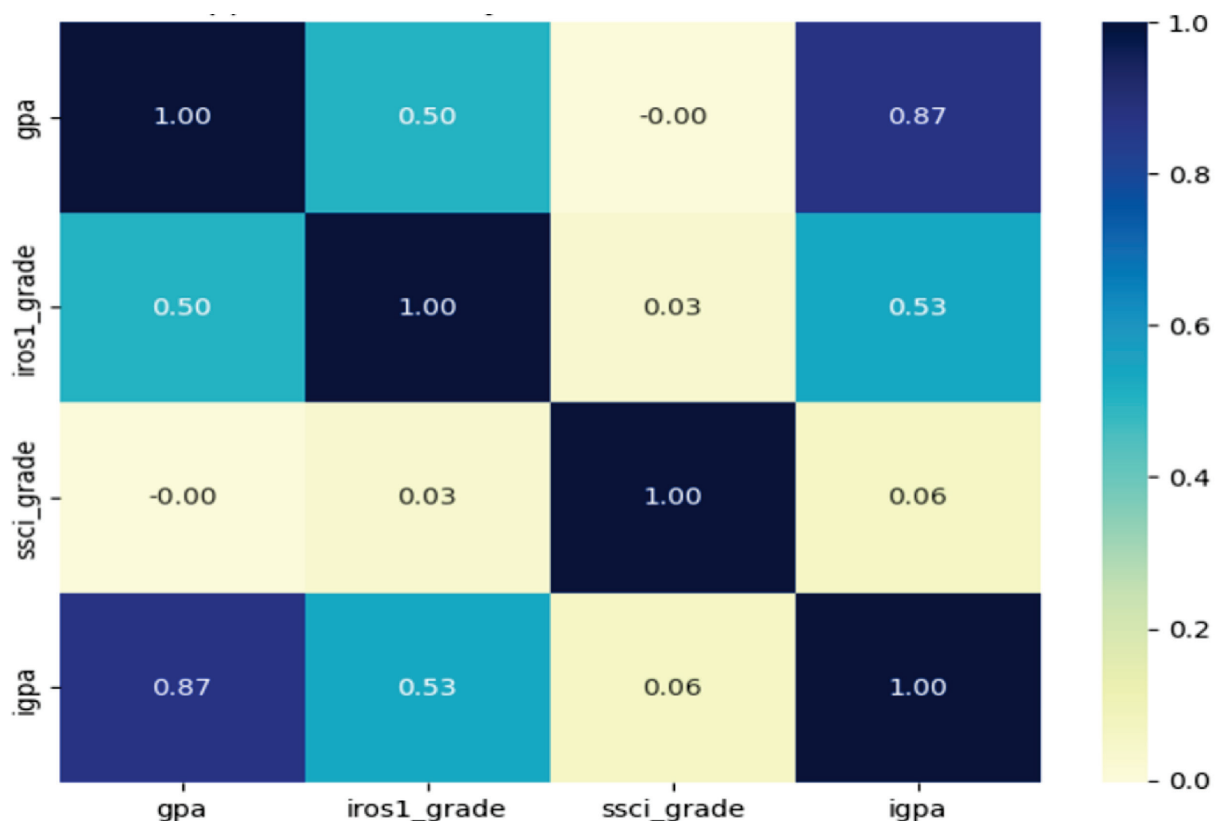


Figure 8. Correlation between IGP and components.

Discussion

The results obtained support the hypothesis on the feasibility of using the integrated IGPA score for constructing a digital graduate profile. Based on the collected data and visual analytical tools, stable patterns of student educational activity were identified, demonstrating the potential of learning analytics and BI tools in quality management and personalized academic support.

Analysis in the context of existing studies shows that the proposed approach offers several advantages over traditional academic performance metrics, as it covers a broader range of student characteristics. Unlike studies focused solely on predictive model accuracy [5], [9], this research emphasizes the interpretation of results within educational practice and the formation of a structured feedback system.

However, some limitations must be acknowledged. The data used were limited to internal sources from AITU, excluding external factors influencing academic performance (e.g., socio-economic status, learning motivation, and psycho-emotional well-being). In addition,

despite the application of machine learning techniques, the sample size and analytical depth can be expanded in future stages.

A promising future direction involves supplementing the model with behavioral data from LMS and introducing real-time dynamic IGPA assessment, which would enable the creation of an adaptive student support ecosystem.

Conclusion

The study demonstrated the importance and applicability of a comprehensive approach to educational data analysis in constructing a digital graduate profile for higher education institutions. The developed functional model, based on the integrated IGPA assessment, provides a more complete view of students' academic performance, research engagement, and social activity. This enables the implementation of targeted academic support and facilitates informed professional self-determination for students.

The results of the study emphasize the importance of digitalizing educational process management and the need to shift from fragmented analysis to systemic analytics at the institutional level. The developed visualization tools and machine learning methods can be scaled to other educational institutions and contribute to the integration of learning analytics practices into the national education system.

Thus, the research not only addresses the initial research questions but also lays the groundwork for further scientific and applied developments in the field of digital academic pathway support.

Acknowledgment

The work was carried out with the support of grant funding from the intra-university project of the Astana IT University on the topic "Development of a data analytics and visualization system for AITU graduate's digital profile (according to iGPA, SCI, ROS indicators from AITU Digital University).

References

- [1] Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, 12(8), 569. <https://doi.org/10.3390/educsci12080569>
- [2] Wu, C., & Carroll, J. M. (2024). Self-presentation and social networking online: The professional identity of PhD students in HCI. *The Internet and Higher Education*, 62, 100951. <https://doi.org/10.1016/j.iheduc.2024.100951>
- [3] Stojanov, A., & Daniel, B. K. (2024). A decade of research into the application of big data and analytics in higher education: A systematic review of the literature. *Education and Information Technologies*, 29, 5807–5831. <https://doi.org/10.1007/s10639-023-12033-8>
- [4] Zhang, M. (2024). Integrating deep learning into educational big data analytics for enhanced intelligent learning platforms. *Information Technology and Control*, 53(4). <https://doi.org/10.5755/j01.itc.53.4.36968>
- [5] Romero, C., & Ventura, S. (2024). Educational data mining and learning analytics: An updated survey. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2402.07956>
- [6] Stojanov, A., & Daniel, B. K. (2024). A decade of research into application of big data and analytics in higher education: A systematic review. *Education and Information Technologies*, 29, 5807–5831. <https://doi.org/10.1007/s10639-023-12033-8>
- [7] Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. *Education and Information Technologies*, 29, 22125–22145. <https://doi.org/10.1007/s10639-024-12704-0>

- [8] Batuchina, A., & Melnikova, J. (2023). Application of learning analytics in European general education schools: Theoretical review. *Quaderni di Comunità*, 2, 119–131. <https://doi.org/10.61007/QdC.2023.2.119>
- [9] Wang, C., Chen, J., Xie, Z., & Zou, J. (2024). Research on education big data for students academic performance analysis based on machine learning. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2407.16907>
- [10] Ed-Daoudi, R., Azhari, M., Ettaki, B., & Zerouaoui, J. (2024). Academic performance prediction in virtual environments using big data and machine learning. *Journal of Electrical Systems*, 20(3). <https://www.scimagojr.com/journalsearch.php?q=21100386444&tip=sid>
- [11] Gómez-Pulido, J. A., Park, Y., Soto, R., & Lanza-Gutiérrez, J. M. (2023). Data analytics and machine learning in education. *Applied Sciences*, 13(3), 1418. <https://doi.org/10.3390/app13031418>
- [12] Zhang, J., Liu, H., & Wu, P. (2025). Machine learning based big data analytics for education in curriculum reforms. *Applied Mathematics and Nonlinear Sciences*, 10(1). <https://doi.org/10.2478/amns-2025-0135>
- [13] Sorour, A., & Atkins, A. S. (2024). Big data challenge for monitoring quality in higher education institutions using business intelligence dashboards. *Journal of Electronic Science and Technology*, 22(1), 100233. <https://doi.org/10.1016/j.jnlest.2024.100233>
- [14] Luo, X. (2024). Learning analytics based on big data: Student behavior prediction and personalized educational strategy formulation. *Applied and Computational Engineering*, 116, 7–13. <https://doi.org/10.54254/2755-2721/116/20251732>
- [15] Khan, F., Al-Shammari, M., & Alghamdi, A. (2022). Artificial intelligence and big data: The advent of new pedagogy in the adaptive e-learning system in the higher educational institutions of Saudi Arabia. *Education Research International*, 2022, 1263555. <https://doi.org/10.1155/2022/1263555>
- [16] Shaulska, L., Yurchishena, L., & Popovskiy, Y. (2021). Using MS Power BI tools in the university management system to deepen the value proposition. *Proceedings of ACIT*, 294–298. <https://doi.org/10.1109/ACIT52158.2021.9548447>
- [17] Gonçalves, C., Gonçalves, M. A., & Campante, M. (2023). Developing integrated performance dashboards visualisations using Power BI as a platform. *Information*, 14, 614. <https://doi.org/10.3390/info14110614>
- [18] Sequeira, R., Reis, A., Alves, P., & Branco, F. (2024). Roadmap for implementing business intelligence systems in higher education institutions: Systematic literature review. *Information*, 15(4), 208. <https://doi.org/10.3390/info15040208>
- [19] Azevedo, A., Azevedo, J. M., & Hayakawa, M. (2021). Designing and implementing a dashboard with key performance indicators for a higher education institution. *Proceedings of CSEDU*, 165–172. <https://doi.org/10.5220/0010539501650172>
- [20] Badhe, R., & Khond, R. (2024). Student performance monitoring system using Power BI. *ResearchGate*. <https://doi.org/10.13140/RG.2.2.32532.10884>