POLITECNICO DI TORINO
Repository ISTITUZIONALE

Artificial Intelligence Bringing Improvements to Adaptive Learning in Education: A Case Study

Original
Artificial Intelligence Bringing Improvements to Adaptive Learning in Education: A Case Study / Demartini, C., Sciascia, L., Bosso, A., Manuri, F.. - In: SUSTAINABILITY. - ISSN 2071-1050. - STAMPA. - 16:3(2024), pp. 1-25. [10.3390/su16031347]

Availability:
This version is available at: 11583/2994103 since: 2024-11-02T19:23:30Z

Publisher:
MDPI

Published
DOI:10.3390/su16031347

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

18 November 2026

sustainability

MDPI

Article

Artificial Intelligence Bringing Improvements to Adaptive Learning in Education: A Case Study

Claudio Giovanni Demartini [1,*], Luciano Sciascia [2], Andrea Bosso [3] and Federico Manuri [1]

[1] Department of Control and Computer Engineering, Politecnico di Torino, Corso Duca Degli Abruzzi, 24, 10129 Torino, Italy; federico.manuri@polito.it
[2] Fondazione per la Scuola della Compagnia di San Paolo, Piazza Bernini, 5, 10138 Torino, Italy; luciano.sciascia@fondazionescuola.it
[3] Links Foundation, Via Pier Carlo Boggio, 61, 10138 Torino, Italy; andrea.bosso@linksfoundation.com
* Correspondence: claudio.demartini@polito.it

Abstract: Despite promising outcomes in higher education, the widespread adoption of learning analytics remains elusive in various educational settings, with primary and secondary schools displaying considerable reluctance to embrace these tools. This hesitancy poses a significant obstacle, particularly given the prevalence of educational technology and the abundance of data generated in these environments. In contrast to higher education institutions that readily integrate learning analytics tools into their educational governance, high schools often harbor skepticism regarding the tools' impact and returns. To overcome these challenges, this work aims to harness learning analytics to address critical areas, such as school dropout rates, the need to foster student collaboration, improving argumentation and writing skills, and the need to enhance computational thinking across all age groups. The goal is to empower teachers and decision makers with learning analytics tools that will equip them to identify learners in vulnerable or exceptional situations, enabling educational authorities to take suitable actions that are aligned with students' needs; this could potentially involve adapting learning processes and organizational structures to meet the needs of students. This work also seeks to evaluate the impact of such analytics tools on education within a multi-dimensional and scalable domain, ranging from individual learners to teachers and principals, and extending to broader governing bodies. The primary objective is articulated through the development of a user-friendly AI-based dashboard for learning. This prototype aims to provide robust support for teachers and principals who are dedicated to enhancing the education they provide within the intricate and multifaceted social domain of the school.

Keywords: education; learning analytics; artificial intelligence; adaptive learning; machine learning; dashboard

check for updates

Citation: Demartini, C.G.; Sciascia, L.; Bosso, A.; Manuri, F. Artificial Intelligence Bringing Improvements to Adaptive Learning in Education: A Case Study. *Sustainability* **2024**, *16*, 1347. https://doi.org/10.3390/su16031347

Academic Editors: Matteo Sacchet, Marina Marchisio and Alice Barana

Received: 26 December 2023
Revised: 25 January 2024
Accepted: 31 January 2024
Published: 5 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Artificial intelligence (AI) is poised to revolutionize higher education (academy), ushering in advancements in learning across a spectrum of applications. By evaluating individual learning styles, preferences, and strengths, AI facilitates the creation of personalized learning experiences. This involves adaptive learning platforms that dynamically adjust content difficulty and pace based on individual performance [1].

In the same framework, intelligent tutoring systems, driven by AI, provide real-time assistance and feedback to students, guiding them through coursework, elucidating concepts, and enhancing comprehension and retention.

The grading process for assignments, quizzes, and exams is streamlined through AI, enabling educators to focus on delivering personalized feedback and engaging with students. AI's analytical process extends to learning analytics, where it delves into extensive datasets of student performance to unveil trends and patterns. Educators can leverage

these insights to refine curriculum design, identify struggling students, and implement timely interventions.

In addition, natural language processing (NLP) tools empower students to hone their writing and communication skills. NLP-powered tools analyze essays, offering suggestions for grammar, style, and content improvements.

On the one hand, virtual labs and simulations that are fueled by AI provide a secure environment in which students can conduct experiments and explore intricate scenarios; these applications are particularly beneficial in science, engineering, and medicine studies. On the other hand, language learning experiences can be transformed by AI-powered apps that deliver personalized instruction, pronunciation feedback, and virtual conversation practice.

AI algorithms are crucial for suggesting additional learning resources—articles, videos, and books—that are tailored to a given student's interests and learning trajectory. Still, in this context, assistive technology is a hallmark of AI, offering features like speech-to-text and text-to-speech capabilities alongside customized learning materials to support students with disabilities. AI-driven chatbots and virtual assistants are able to swiftly address common queries and promote student engagement and satisfaction.

In the realm of academia, AI automates tasks like data analysis, literature reviews, and hypothesis generation, expediting the research process for scholars. Here, a robust scenario is given by predictive analytics; this scenario—driven by AI—identifies students who are at risk of academic difficulties or dropout, allowing institutions to provide timely support and interventions.

However, it is important to note that successfully integrating AI into education requires careful planning, ethical considerations, data privacy protections, and ongoing evaluation of AI systems. Additionally, AI should complement, rather than replace, human educators—the human touch and mentorship are essential components of effective learning [2]. This topic is being explored in detail within the Data2Learn@Edu project, started in April 2023, with a strong partnership sustained by the "Compagnia di San Paolo" Foundation.

This work is one of the first outputs of the project, which is still in its preliminary phase. Project partners include Politecnico di Torino (PoliTo); The National Institute for the Evaluation of the Education and Training System (INVALSI); Politecnico di Milano (PoliMi); Foundation for the School (FpS); LINKS Foundation; Ufficio Scolastico Regionale per il Piemonte (USRPi); Ufficio Scolastico Regionale per la Sicilia (USRSi); Research Institute for the Evaluation of Public Policies (IRVAPP); Schools from Piemonte (SCPi); and Schools from Sicily (SCSi). The PoliTo research environment, together with its own experimental testbed, has been assumed as the site for the preliminary investigation carried out in this work.

This paper is organized into six sections. Section 1 deals with the state of the art, referring to the principles, problems, and solutions that have already been proposed in the literature, and mainly focuses on learning analytics and profiling. Section 2 briefly describes the Data2Learn@Edu project, its general architecture, and the deepening methods, concepts, and tools that were applied in its development; the focus is the specific experience of the first phase of the implementation cycle at Politecnico di Torino. Data analysis, the analytics platform utilized, the clustering algorithms utilized, and the RapidMiner process are also covered in Section 2. Section 3 addresses the case study and its associated dataset analysis, having been the case considered as an outcome of the activity carried out for this work. Section 4 presents a discussion and information about the clustering process and cluster characterization, enabling the interpretation of the extracted knowledge. Section 5 reports on the research questions and corresponding answers, while suggesting sustainable paths for future work.

All relevant acronyms used in this work are summarized in the Nomenclature section.

According to the sustainability issues outlined in the United Nations' (UN) 2030 Agenda, this project meets Sustainable Development Goal 4 (SDG 4): ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all.

1.1. Learning Analytics

The focus of the Data2Learn@Ed project is learning analytics (LA). "Learning" is broadly defined across a range of contexts, including informal learning on the internet, formal educational study in institutions (primary/secondary/tertiary), and workplace learning. "Learning analytics" is, according to the *Journal of Learning Analytics*, defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs, as defined back in 2011 for the first LAK—Learning Analytics & Knowledge—this general definition still holds true even as the field has grown" [3].

The educational landscape is continuously expanding, driven by advancements in network infrastructure, storage technology, and data processing systems. This surge in educational data, originating from both individual learners and educational institutions, presents an unprecedented opportunity, as highlighted in the OECD Digital Education Outlook 2021 [2]. Within this context, learning analytics (LA) assumes a central role. It involves the analysis and assessment of static and dynamic information, aiming to model learning environments in real time. These environments capture and categorize learners' behaviors and personal attributes. Detailed descriptions, predictions, and optimizations of educational domains become feasible with the use of analytics. These insights inform decision making at various levels of authority within educational organizations, from individual learners to classes, institutions, and even ministerial bodies. Learning analytics encompasses a range of techniques, including statistical and predictive modeling, data mining, and artificial intelligence (AI) methods [4]. It enables the analysis, evaluation, and synthesis of organizational, educational, personal, and behavioral data, rendering them valuable for decision-support systems across all layers of educational organizations. While the *Journal of Learning Analytics* defines LA as research aimed at enhancing learning, other researchers [5] emphasize the roles of LA and machine learning within the broader context of the education domain, such as prediction, profiling, and data visualization.

LA can be applied in various educational settings, including schools, universities, online courses, and corporate training programs. This subsection presents key aspects of the domain, which can be used to provide suggestions for development.

As stated above, LA is a multifaceted field with several integral components. At its core is the meticulous process of data collection, drawing from diverse sources such as student performance metrics, engagement indicators, and demographic details. This wealth of information is gleaned from platforms like learning management systems, online quizzes, surveys, and other educational tools.

Once these data are amassed, the subsequent phase involves sophisticated analysis. Techniques encompassing statistical analysis and the application of machine learning algorithms come into play, aiming to distill valuable insights that can guide decision making and elevate the quality of learning experiences.

Visual representation is a crucial aspect of this process, as data are often translated into graphs and charts. This graphic format is a communicative tool that ensures that the data are accessible to educators and administrators, allowing them to seamlessly discern trends and patterns.

Looking forward, learning analytics can extend its impact through its predictive capabilities. Educators can intervene proactively, fostering improved outcomes and academic success, foreseeing future student performance, and identifying those at risk [6].

Furthermore, the personalization of educational content emerges as a distinctive feature. Analyzing individual students' performances and preferences allows learning materials to be tailored to meet the unique needs and learning styles of each student.

The role of learning analytics is not limited to assessment and feedback at the individual level; it extends to institutional considerations. It informs decisions at an institutional level, influencing curriculum design, resource allocation, and policy changes, holistically contributing to the enhancement of the quality of education.

Despite these advancements, ethical considerations remain paramount. The ethical dimensions of learning analytics emphasize the importance of data privacy, transparency, and fairness in collecting and utilizing data about students and their educational experiences.

Overall, learning analytics has the potential to revolutionize education by providing educators and institutions with valuable insights to improve teaching and learning processes. It enables data-informed decision making, enhances student engagement, and improves educational outcomes [7]. However, it is essential to use learning analytics responsibly and ethically to protect the privacy and interests of students.

In higher education, numerous studies and benchmarks have explored the adoption of learning analytics, while its implementation in schools still needs to be more widespread [2]. Higher educational institutions and schools often collect student data related to exam results or educational backgrounds, which are not fully leveraged. These data usually reside in separate repositories without effective mechanisms to link them together. Maximizing the use of this information could significantly enhance the quality of the services provided by universities and schools, thereby improving student success rates [7].

1.2. Student Profile

As our understanding of student profiles deepens, making it possible to develop true digital twins of the learners, the ability to predict potential academic challenges improves. This has far-reaching implications for stakeholders within the educational community, impacting key domains and their respective interested parties:

1. Student behavior and performance—affecting students and educators.
2. Course management—involving teachers, assistants, and tutors.
3. Decision support in educational institutions—engaging administrative directors and principals.
4. Retention and dropout prevention—primarily influencing students and indirectly affecting local and national educational bodies.

The first two stages in the list deal with teachers and educators, who oversee the assessment of learning processes and shape a continuous adaptation of course content and organization to follow the learning dynamics that emerge in classrooms. The other two stages concern policies that are established at the institutional and governmental levels.

Due to the implications mentioned above, the current state of education institutions—conceptualized and identified as Learning Organizations—demands immediate attention and necessitates further involvement; this is in line with emerging innovation-driven perspectives [2]. According to most university governing bodies and school principals, education leaders must take charge of the data analytics environments they establish. These environments should remain in the public domain, with transparent governance to ensure the healthy functioning of educational communities.

The utilization of data analytics and artificial intelligence (AI) in educational spheres is still in its early stages and is undergoing rapid development. Consequently, it should be straightforward to prevent the complete control of these activities from falling into the hands of commercial vendors whose primary objective is profit maximization for their owners and shareholders.

The original concept, independently conceived by several projects' partners in education, aimed to create statistical models based on historical data concerning students and teachers, spanning multiple semesters and years. Initially, this involved working with isolated and limited data repositories; this is because achieving a broader institutional perspective proved challenging. Consequently, an ad hoc and localized data collection approach was pursued, focusing on the data that deans, principals, and teachers usually required. Data have been used to construct models capable of adapting to students' profiles and recommending organizational adjustments.

The model architecture chosen in this work underwent validation through pilot trials, utilizing information from the past five years within the educational context [8]. These findings gave educators and administrators insight into the effectiveness of systemic

decisions regarding their Learning Organizations, based solely on data collected from and available within their information system repositories [9].

Following the model's necessary political and technical validation, the academic governing body, at Politecnico, affirmed that every student has the right to access the reports on their academic performance to facilitate their self-assessment process. Similarly, professors/teachers should have access to these results to evaluate their learning processes and make ongoing adjustments to their course content and organization to align with emerging classroom dynamics. With the support of the entire teaching and learning community, the program coordinator and the institution's dean can make informed decisions to enhance the overall educational landscape based on a shared data-processing framework.

Analytical techniques empower educational institutions to gather, measure, and analyze information about each student, including their contextual connections, and to determine if and how these connections influence a given student's learning journey. Furthermore, analysis outcomes can provide insights into students' strengths and weaknesses, improving their productivity, performance quality, and overall effectiveness. This same framework can assist with modeling individual and collective profiles, involving intermediate and higher governing bodies.

In light of the growing needs of new generations of students and educators concerning innovation, advanced solutions are essential for enhancing the educational environment. Various projects are addressing these needs globally, such as the "Riconnessioni" Project in Italy [10], which pursued several actions addressing technological improvements in schools, enforcing institutions' network access, and widening teachers' training on the internet and computational thinking.

As these demands continue to rise, implementing and utilizing this analytical framework will enable data collection to be performed by both learners and educators. Combined with the models, managing these data will facilitate the discovery of valuable insights to shape scalable responses to the requirements of both the education and labor sectors.

2. Materials and Methods

2.1. The Data2Learn@Edu Project

Figure 1 depicts the project organization, where the main actor, Leading_PoliTo, relies upon the two pilot projects, Pilot_INVALSI and Pilot_PoliMi; these bring the experimental side of the project to life. This is carried out by working directly with—and within—the schools that are chosen as a testbed. They are in two different, nonhomogeneous geographical sites to draw comparisons between concepts, procedures, and organizations; such an arrangement can help develop an appropriate dashboard tool suited to meet the needs of teachers, principals, and governing bodies. Hence, the perspective is scaling from a broad scale to a smaller one, where the descriptions of problems and objectives reduce their impact to a narrower scenario.

The pilot projects, integral to the overarching initiative, execute the experimental phase of the applied research. They progressively develop the dashboard as an evolving prototype within the learning cycles implemented in the schools selected by the local partners, USR-Piedmont and USR-Sicily. A comprehensive two-year training program is also anticipated to equip teachers and principals with the necessary skills for utilizing the dashboard effectively. This training involves direct collaboration with third-sector partners, namely Fondazione per la Scuola (FpS) (Turin, Italy) and Fondazione LINKS (Turin, Italy).

To complete the scene, an academic testbed has also been included to support the specific analysis carried out on benchmark tools; these are able to sustain the requirements of primary users exposed to the dashboard in the first phase of the project. The academic testbed has been rooted in the Politecnico di Torino data and learning environment, which is the focus of the investigation carried out in this work.

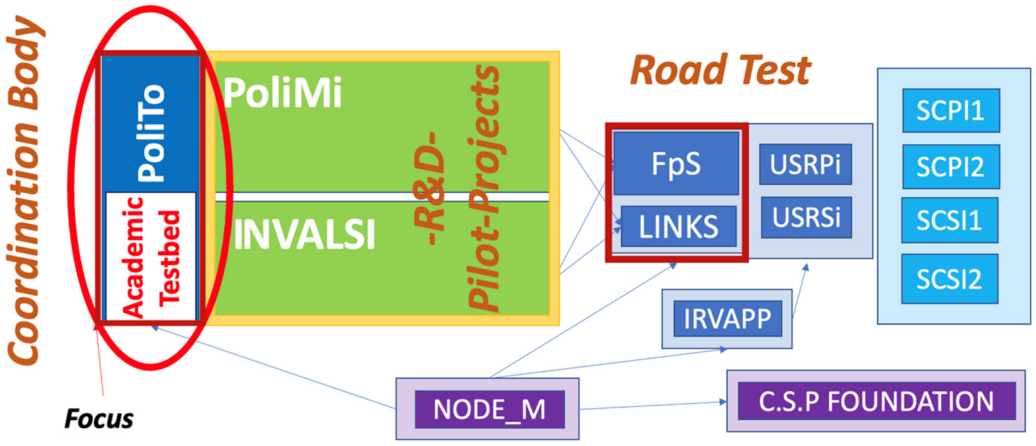


Figure 1. Data2Learn@Edu. Project partners' identification and roles: Politecnico di Torino (PoliTo); The National Institute for the Evaluation of the Education and Training System (INVALSI); Politecnico di Milano (PoliMi); Foundation for the School (FpS); LINKS Foundation; Ufficio Scolastico Regionale per il Piemonte (USRPi); Ufficio Scolastico Regionale per la Sicilia (USRSi); Research Institute for the Evaluation of Public Policies (IRVAPP); Schools from Piemonte (SCPi); Schools from Sicily (SCSi). The PoliTo research environment, together with its own experimental testbed, has been assumed as the focus for the investigation carried out in this work.

The project Data2Learn@Edu is illustrated in Figure 2, outlining a comprehensive Plan, Do, Check, (Act) system model that aims to provide a holistic perspective on the educational domain. The Deming cycle, together with the closed-loop control model that inspired the project, primarily focuses on creating an adaptive teaching and learning framework that is rooted in a scalable data-driven approach. Integrating data mining and machine learning techniques establishes a continuous closed-loop regulatory environment that is equipped with an adaptable and intelligent toolkit. The latter makes a pivotal contribution to enhancing the learning process, with an emphasis on customization and contextualization. Teachers play a crucial role in assessing the learning processes and continuously adapting the course content and organization to align with the evolving dynamics in the classroom. The project envisions that both the head of the study program and the principal of the institution, with the support of the entire teaching and learning community, can make informed decisions based on a common data-processing framework. In particular, the adaptive updating of learners' profiles allows for a deeper understanding of the underlying dynamics in educational and teaching activities, as discussed by Brancaccio et al. [11]. Content updating and teaching delivery heavily rely on data-driven methodologies and AI algorithms to enhance the personalization and contextualization of both learning outcomes and learners' profiles, as depicted in the figure. According to this framework, learning outcomes—encompassing competencies, skills, and knowledge to be achieved—are mapped onto the offered profile of students (item 5 in Figure 2). This profile is developed through the learning process (item 4), which operates on a specific timing schedule. It is constructed based on the collective description of students' learning outcomes, measured through their performance data, and collected and processed in the sensor (item 6). These data are also linked to personal information and are further integrated with historical and real-time data. Additionally, the offered profile of students may also be enriched by incorporating other data sources, such as social and emotional skills, which are increasingly relevant in education [2], or data sampled in the open educational resource's domain [12].

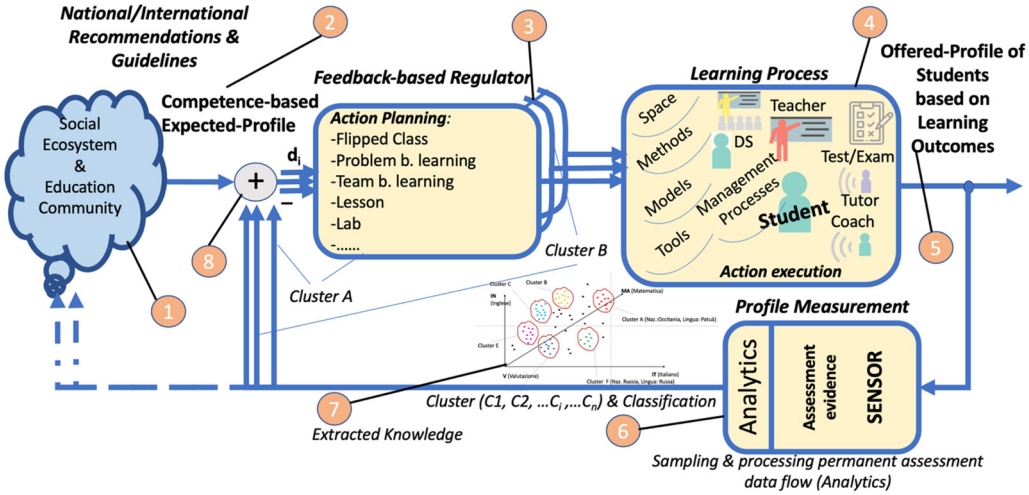


Figure 2. Data2Learn@Edu: A data-driven adaptive learning model within a scalable learning organization. (1) The social ecosystem and the education community able to define the expected profile (2) in terms of skills, attitudes, knowledge, and competence. (3) The regulator, able to plan future actions carried out in box (4), which depicts the learning process in charge of executing those actions, including student assessment. Its output is (5), the offered profile, which is then sampled through the sensor (6), which is able to gather data (previously collected through tests and exams) and compute and define the classified students' clusters (7); here, each cluster has its own collective profile. (8) Compares the expected profile with the one each cluster exposes. This comparison gives a difference (error), d_i, which is used by (3) to plan a further action to be executed again in (4). The whole system conceptualization derives from the closed-loop control model, although the learning process, shown in (4), almost always relies on persons and not on mechanisms.

The sensor ((6) in Figure 2) operates by examining the actual data collected during test and exam sessions conducted as part of the learning process ((4) in Figure 2). That sensor houses smart algorithms that analyze the extensive data cloud generated throughout the learning process. Each individual learner is represented as a drop in this cloud, and the algorithms can group these drops into clusters based on density measurements. Consequently, the extracted knowledge ((7) in Figure 2), the output of the sensor ((6) in Figure 2), consists of the measured offered profiles. Specifically, the sensor output ((7) in Figure 2) produces a set of clusters, which are individually compared ((8) in Figure 2) to the competence-based expected profile ((2) in Figure 2), serving as an institutional reference for any offered profile.

The expected profile is constructed based on the learning outcomes specified by the national standards, such as the National Recommendations ((2) in Figure 2); these are the National Guidelines (National Recommendations and Guidelines for primary and lower secondary education and the second-level secondary school—D.M. n. 254 del 16/11/2012, 2012) for education, or a Technical Standard in Systems Engineering (ISO/IEC/IEEE 15288/2023) [13] for higher education. The comparison ((8) in Figure 2) reveals the difference or error (d_i) between any cluster-related profile and the reference one. This computed error (d_i) drives the feedback-based regulator actions ((3) in Figure 2), which are designed to address and rectify this error [8]. Each cluster ((7) in Figure 2) the sensor generates is associated with specific properties determined through the analysis the algorithms run ((6) in Figure 2). This information may enable a decision-support system (DSS) to identify the appropriate regulator instance responsible for devising a plan for corrective action. When implemented in the learning process ((4) in Figure 2), this plan aims to resolve any emerging errors (d_i).

Addressing this difference (d_i) may involve various actions, such as implementing flipped classrooms within living labs, adopting problem-based learning, enhancing lectures, or executing specific knowledge and teaching strategies in suitable environments. The

regulator ((3) in Figure 2) focuses on devising a tailored plan for corrective action, and the instance chosen depends on the offered profile associated with the cluster.

The regulator instances, a vital element of this investigation, are depicted as bold blue overlapping boxes, representing a multilayered device. The choice of a specific instance depends on the cluster ((7) in Figure 2) selected from those computed by the sensor's analytic engine component.

Considering the scenario outlined above, the research questions (RQs) central to the project's development include the following:

1. How can student data generated through education be effectively managed within a Learning Organization (Figure 2)?
2. What methods and tools are employed to collect and process data, including measuring student performance ((6) in Figure 2)?
3. What are the basic components of the competence-based expected profile ((2) in Figure 2), built around social ecosystem requirements?
4. How is the learner offered profile ((5) in Figure 2) described based on learning outcomes carried out in the learning process ((4) in Figure 2)?
5. What actions are required to regulate the learning process ((3) in Figure 2)? How are these actions identified, selected, planned, and implemented? What level of authority oversees them?

2.2. Data Analysis

Data analysis is the process of inspecting, cleaning, transforming, and modeling data to discover useful information, draw conclusions, and support decision making. It is a critical component of various fields, including business, science, finance, healthcare, and more.

Tools and software commonly used for data analysis include programming languages like Python and R, as well as specialized software like Excel v2401, Tableau, Power BI v2.122.746.0, RapidMiner v10.3, and various statistical packages.

The specific techniques and approaches used in data analysis can vary widely depending on the goals of the study and the type of data being analyzed. Data analysis is crucial in decision making, problem solving, and gaining insights across various domains.

2.3. Analytics Platform

To conduct our analysis, different tools, briefly described in this paragraph, were chosen:

1. Microsoft Excel was employed for initial dataset management and the processing of data integration.
2. PowerBI, which is an interactive data visualization software, was developed by Microsoft with a particular focus on business intelligence. In this study case, the data processed with RapidMiner were loaded into this tool and used to create an interactive dashboard to explore and analyze findings. Most of the charts shown in this paper have been extracted from a proper PowerBI dashboard.
3. RapidMiner was utilized instead for conducting data mining activities, including the execution of clustering algorithms. This tool is well recognized within the field of educational data mining. RapidMiner stands out as a visual tool that enables the development of data mining analyses and the creation of models without requiring proficiency in a specific coding language. One of its notable advantages is its accessibility and user-friendly interface, although it may not offer the same level of flexibility as dedicated coding languages like Python or R.

These tools are well recognized within the field of educational data mining [14].

2.4. Clustering Algorithms

k-Means is a simple unsupervised machine learning algorithm, conceived in 1967 by James MacQueen [15]; it is well known in the educational data mining area [16]. This algorithm aims to assign each dataset point to a cluster, trying to minimize the distance between each point and the respective cluster centroid to which it has been assigned.

The other algorithm tested was k-medoids, which is similar to k-means but makes some substantial changes. Centroids are always chosen from the dataset points, while in k-means, they can also be points not belonging to the dataset.

Two metrics were used to evaluate the performance of these algorithms and to choose the one that fitted the dataset best:

- Average within-centroid distance: This measures the distance between the cluster points and their respective centroids. It is an intra-cluster measure, and the lower the value, the better the cluster compactness. This measure is used to assess the best parameters for the algorithms.
- Davies–Bouldin index (DBI) [17]: This measures the ratio of intra-cluster distances to inter-cluster distances. It compares the performances of various algorithms. The lower the value, the higher the overall quality of the clustering process.

2.5. RapidMiner Process

In Figure 3, the RapidMiner process is represented; this is used for the clustering operations. The process starts with some preliminary steps:

- Exclusion of students who have not yet attempted the exam.
- Selection and normalization of the clustering attributes.

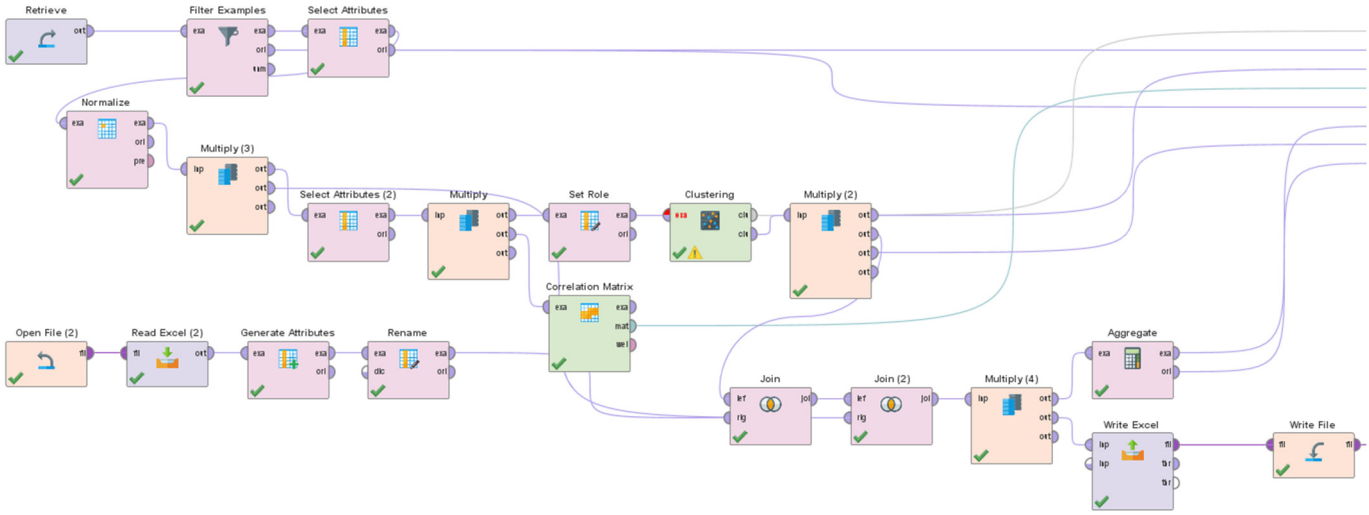


Figure 3. Data2Learn@Edu: A data-driven adaptive learning model for school/academia; a scalable learning organization focusing on the analytics and the decision-support system roles.

The next steps involved the creation of a correlation matrix and then the clustering itself.

In the last steps, after the clustering, the information removed before the clustering was added back using an ID to characterize our clusters in the analysis presented in Section 4.2.

The clustering output was then saved in an Excel spreadsheet; this was used to create a PowerBI dashboard that visualized and analyzed the data in more depth.

3. Results

3.1. Case Study: Innovation Management and Product Development Course at PoliTo

The current situation within education demands immediate attention and a comprehensive exploration of emerging perspectives. As per the directives issued by the Board of Directors at Politecnico, an academic institution must assert control over its data analytics infrastructure, ensuring accessibility for academic purposes. Transparent governance mechanisms must be established to foster the well-being and functionality of the academic community.

While the traditional scholarly publishing infrastructure is deeply entrenched and resistant to swift change, the integration of data analytics and artificial intelligence into academia remains in its infancy, subject to ongoing evolution. Consequently, it is crucial to prevent the complete

relinquishment of control over these activities to profit-driven commercial entities, which, understandably, prioritize the maximizing of returns for their stakeholders.

The proposed concept suggests a solution that is capable of generating statistical insights from historical data about students and educators, spanning multiple academic periods. This approach seeks to collect information that can aid directors, educators, and even students in making informed decisions regarding their organizational and career trajectories. The shaped data and process model architecture underwent validation using data collected from students and educators over recent years. The resulting insights empowered educators and academic leaders to make informed systemic decisions concerning the utilization of data available in information system repositories across educational institutions.

Upon completion of the requisite political and technical validation processes, at Politecnico di Torino, it was established that students have the right to access their reports on their academic performance, facilitating their self-assessment endeavors. Similarly, educators can leverage these results to evaluate their learning processes and adapt their course content in accordance with emerging classroom dynamics. Furthermore, study program administrators, committed to supporting the broader teaching community, can base their decisions on the insights derived from the same data processing procedures.

Therefore, applying analytical techniques enables educational institutions to gather, assess, and process data concerning each student in connection with their specific context, revealing how this connection influences the learning experiences of the same student. Furthermore, it facilitates the description of students' strengths and weaknesses, thereby shaping the quality and effectiveness of their output and performance and building individual and group profiles. Given the evolving landscape of students and educators needing innovative solutions to enhance education, the adoption of this methodology enables the collection of data generated by both learners and educators. The utilization of these data, coupled with analytical models, aids in uncovering valuable insights to craft optimal responses to the educational community's demands.

The ongoing project, Data2Learn@Edu, introduces an "adaptive learning system", as illustrated in Figure 2. That system, working by construction as a closed-loop control/regulator, has—as a primary objective—enhanced learning process comprehension within the field of teaching and education through the utilization of data-driven methods. By integrating data mining and machine learning techniques, this platform transforms itself into an adaptive and intelligent tool that is capable of significantly influencing learning processes. It achieves this by reinforcing and personalizing educational experiences.

Notably, the automatic generation and allocation of learner profiles provide deeper insights into the dynamics underlying current educational and teaching practices. The adaptation of content and support for teaching activities heavily relies on data-driven methodologies and AI algorithms. This approach boosts the personalization and contextualization of learning materials and learner profiles, as depicted in Figure 4. Consequently, learning outcomes can be aligned with the learner's profile in terms of competencies, skills, and acquired knowledge; in the case of the academy testbed, readers could also refer to the work presented in [15].

The extracted knowledge is crucial for enhancing tools and platforms that are centered around data and focused on individuals. The primary objective is to impact the educational community's processes, which are typically formalized within the organization but are also informally represented in its perceived image.

Particular attention is dedicated to the learners' profile difference detector, as depicted in Figure 4. This detector, or comparator, works on the students' offered learning outcome profiles, through the identification of the set of clusters' specific profiles; these are output by the sensor, which can also access personal data repositories in conjunction with historical and current students' performance data (PoliTo Data Lake). These offered profiles are then compared to the reference job profile, or expected profile, which is built around learning outcomes recommended by industry standards and market assessments (e.g., ISO 15288: Technical Standard in Systems Engineering [18]). The difference computed by the detector,

or error, determines the regulatory/control actions, as outlined in Figure 4, where an intelligent decision-support system (IDSS) [19] can provide appropriate recovery proposals for teachers, deans, or other responsible agents. These actions include, as an example, the establishment and subsequent execution of flipped classes, whether in physical or virtual living labs, further access to appropriate learning materials (available online), or further discussions with specific tutors on critical subjects.

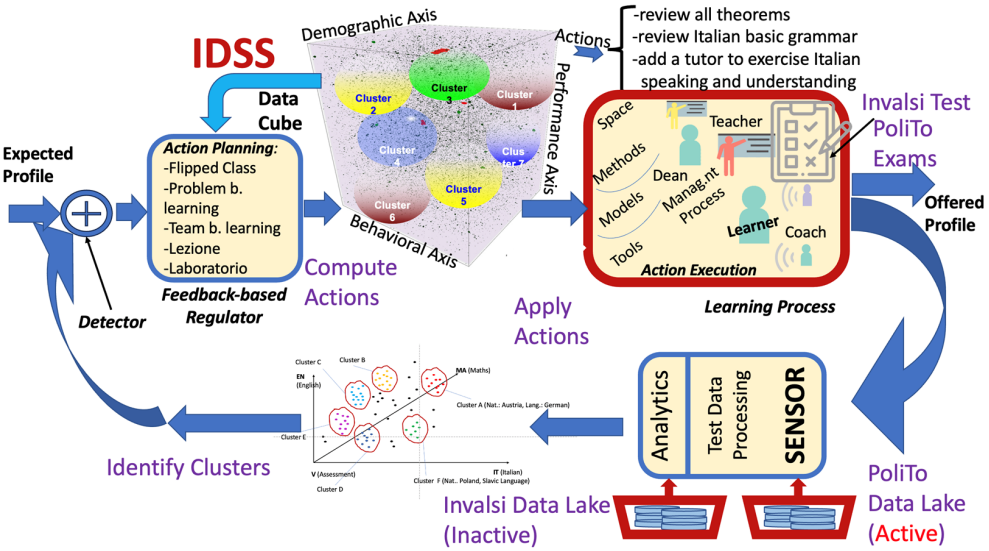


Figure 4. Data2Learn@Edu: A data-driven adaptive learning model in school/academia; a scalable learning organization focusing on the analytics and the decision-support system roles. Polito Data Lake is the active repository supplying the system for this experience.

Problem-based learning, lectures, and various specific learning methods or environments are under consideration. The primary focus of the regulatory approach is outlined as a recovery action plan, which will be devised based on data collected by the sensor. Within this context, relevant components of this investigation are marked by the bold lines on the image.

Consequently, the main research inquiries primarily revolve around two aspects: firstly, understanding the data-driven adaptive learning model, reported in Figure 4, its components and operations, and the tools it employs, as clearly stated by RQ1; secondly, exploring the processing that needs to be conceived within the sensor to identify, represent, and classify the various offered profiles, which are subsequently compared with the given expected profile, so as to rectify relevant deviations from the chosen and stated reference, as traced by RQ2.

In the course under investigation, the learning process adheres to a six-stage life cycle [20], as depicted in Figure 5. In the initial stage, the problem stems from the context of a company or organization that participates in the course. This often entails analyzing images that are perceived through the lens of emerging innovation trends, prompting companies to reassess their value chain based on the evolving flow, shaped by technological advancements and competitive practices. The problem is submitted to the student's team, who has to study and develop a sustainable solution to be proposed to the company.

The second stage involves problem-posing, where a selection of methods and tools specific to the domain is employed to comprehensively describe the issue at hand. This phase aims to fully grasp the problem and assess its impact on the systems and the environment.

The third stage focuses on developing effective strategies to address the problem, leveraging algorithmic approaches based on insights and perspectives gathered in the previous stage, which are then suitably formalized using established practices and standards.

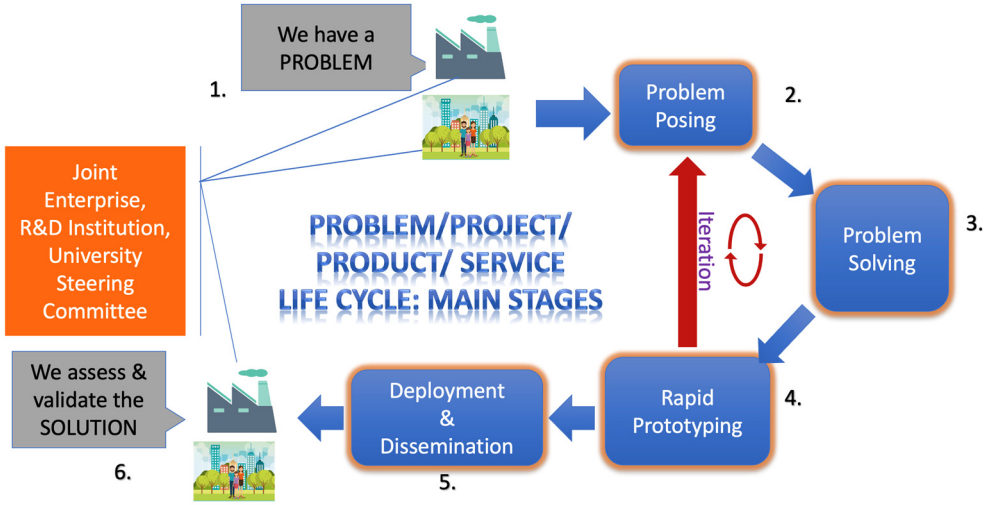


Figure 5. Data2Learn@Edu: Problem/project life cycle at the root of the problem-posing/-solving-based learning approach.

Moving on to the fourth stage, a minimum viable product (MVP) begins to take form. Here, hardware platforms, software components, and programming languages are combined. Rapid developing principles are applied, emphasizing the "reuse" of previously developed components to create a sustainable performance level, ultimately enhancing cost-efficiency and reducing the time taken to bring the product to the market.

The subsequent stage, encompassing deployment and dissemination, explores how marketing and communication strategies are employed through the appropriate channels to engage various stakeholders and secure funding sources.

In conclusion, from an educational perspective, the assessment is based on predefined learning objectives and outcomes for both the specific course and the Master of Science program. The perspective of the enterprise or organization plays a pivotal role in the assessment process, and self-assessment is also encouraged to compel team members to account for their contributions and estimate the costs incurred throughout the entire life cycle development process.

3.1.1. Context and Data Framework

This case study is conducted within the framework of the Innovation Management and Product Development course (GISP: PoliTo Data Lake), which currently attracts over 100 new students, making it a popular course among all offerings in Politecnico's Master of Science programs. Students simultaneously engage in various other subjects alongside the GISP course, including project management, object-oriented programming, business planning, quality management, and data-driven application development.

3.1.2. Classroom

The constructivist classroom in the 2023 academic year accommodates a diverse population of over 100 students and is organized weekly into two theory lessons and two teamwork-based lab sessions. In this classroom, instructors, teachers, and trainers have the role of creating a collaborative environment where learners are actively involved in their own learning. Each group, consisting of five individuals, autonomously manages their working process, enforcing internal collaboration as they tackle intricate challenges associated with specific projects. The latter are often shaped based on brainstorming sessions, sometimes with the assistance of external enterprise actors. Within each group, cooperation is pursued through various collaborative tools such as Dropbox and Google Drive for data storage, Skype, Teams, or Zoom for synchronous communication, and appropriate application and system development specification tools, such as Visio or StarUML.

The classroom layout includes multiple zones where students can convene and sit in circles, in stark contrast to the conventional teacher-centered arrangement where students sit in rows while receiving a continuous stream of lectures.

3.1.3. Course Delivery

The active learning methodology presented in [20] represents a unique fusion of traditional and constructivist approaches within a dynamic learning framework, as described in Figure 4. In this approach, the course structure functions as a living laboratory, mirroring the project's life cycle in accordance with the project work syllabus. The weekly schedule is divided, with 50% of the time dedicated to project development and the remaining 50% devoted to conventional lecture-style teaching. This amalgamation addresses a dual challenge.

On the one hand, it aligns with a university's corporate-style organization, where time is systematically regulated based on labor coordination and passive interactions. On the other hand, it accommodates the demands of creativity-driven processes, primarily rooted in stimulating student engagement. An intriguing aspect of this methodology is that students actively participate in the course's organization. They kick off the course by engaging in a meeting with the Joint Steering Committee, which was formed specifically for this purpose. This meeting serves as a platform for addressing fundamental questions that unveil various dimensions of the proposed problem.

The course spans a duration of 13 calendar weeks. The initial week focuses on introductory activities, including an overview of the course schedule and organization. The second week delves into kickoff discussions regarding the challenges that companies, or other organizations, aim to tackle. Students can also build teams during this period, bringing together complementary skills, knowledge, and experiences. The team composition is finalized after considering the introductory insights regarding the issues raised by the companies and the problem-specific needs.

Students immerse themselves in the problem-posing phase between the second and fourth weeks. Here, projects begin to take shape through a top-down deductive approach. At this stage of project life cycle management, the existing framework is recognized and serves as a foundation for developing new proposals. Questions play a pivotal role within the "problem-posing" domain, enabling a comprehensive exploration of the problem's context. This exploration is facilitated through Lean Model Canvas (LMC), logical framework analysis (LFA), and quality functional deployment (QFD).

Moving from the fifth to the seventh week, the focus shifts to problem solving, emphasizing formal and informal specification development, often involving algorithmic techniques. Building upon the earlier problem analysis and process planning, students become proficient in using integrated computer-aided manufacturing definition for function modeling (IDEF0) and unified modeling language (UML) notation for specification processing. Their goal is to create a "to be" model, which can be compared to existing benchmarks—the "as is" state of the art.

Weeks 8 through 10 are dedicated to building a sustainable prototype that aligns with the goals and constraints established by the Joint Steering Committee.

Students engage in deployment and dissemination activities during the final three weeks (11–13th weeks). They test the prototype on an appropriate testbed and plan comprehensive communication strategies for the closing exposition, which is presented to the Joint Steering Committee. This presentation includes videos, reports, and a complete technical demonstration for the final discussion. Intermediate release dates are strategically placed to ensure the timely delivery of the LFA, QFD, and UML specifications, as well as a preliminary prototype implementation. Additionally, a well-structured timetable is established to align individual skill development.

3.1.4. Assessment

In the conducted study, the assessment process for students was structured into four distinct steps, as outlined in Table 1. The most impactful component of that assessment was the project work, which received collective feedback from the Joint Steering Committee. This feedback was generated following group discussions on project development. The concluding discussion was documented comprehensively and included a summary slide sequence. This documentation coped with the project work syllabus framework: it was accompanied by various elements, including a concise technical video illustrating the prototype's functional behavior, a brief (three minutes) emotion-based Kickstarter-like video, the coding software used for the prototype, its testing, and the toolkit for its management and development.

The project work syllabus serves as the foundational reference point for planning and is a crucial aspect of the regulation unit (as depicted in Figure 6). This syllabus provides a detailed job profile interface and corresponding descriptions, specifying primary activities within the enterprise/organization, and establishes a link to the learning outcome profile. The learning outcome profile encompasses the skills, attitudes, competencies, and knowledge elements.

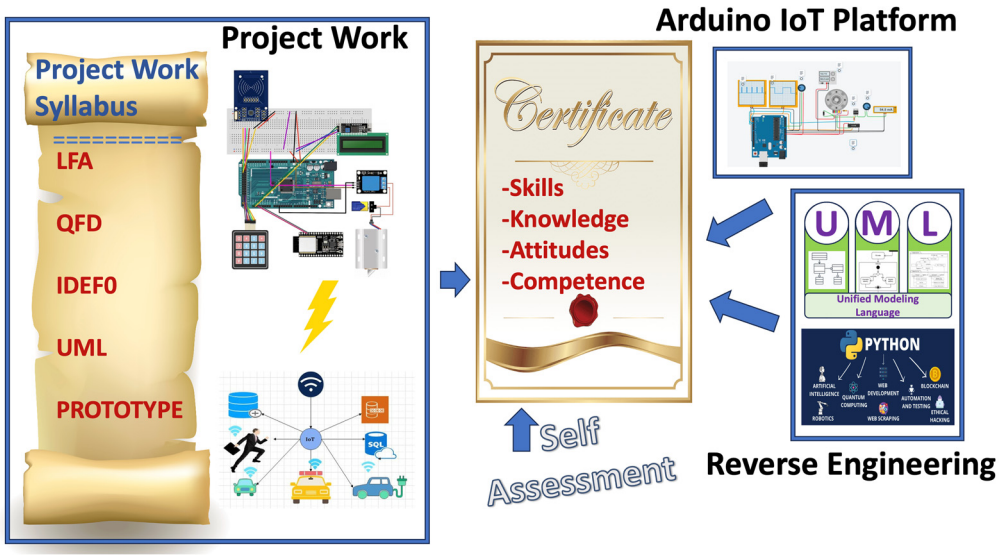


Figure 6. Data2Learn@EDU: Collective and individual student assessment map.

The second component listed in the assessment table pertains to a test bank. This test bank includes a course reference text, "UML 2.0", an Open Educational Resource (OER) that comprehensively covers software engineering. Students are required to complete a test based on this topic. For this purpose, a reverse-engineering section is introduced; here, students work with provided Python code segments to be processed and understood to create functional and system diagram interpretations. They also work on IoT devices using Arduino platforms to collect field data to be sent to the cloud using specific Wi-Fi devices to accomplish the task. Figure 6 also illustrates the interconnected relationship between the project syllabus, UML, the reverse-engineering process, and the corresponding assessment tools.

Furthermore, an individualized self-assessment mechanism is also implemented. This mechanism allows for the differentiation of project work assessments based on each group member's abilities and level of participation. In practice, it involves allocating a specific number of credits to each team member, who then distributes these credits among their peers based on their assessment of each colleague's practical contributions to the prototype development.

3.1.5. Assessment Management

Managing a course while simultaneously enhancing the learning process can be a complex and non-intuitive task for educators, professionals, and their support staff. The challenges related to improving teaching and learning can be identified using questionnaires to detect sustainable evidence in practice. Table 1 illustrates the assessment schema adopted in the GISP course.

Table 1. Students' assessment distribution.

Section	Description	Weight
A.	Project Work	75%
B.	Arduino platform	7.5%
C.	Reverse Engineering (UML)	7.5%
D.	Behavior	10%

As stated, Figure 6 depicts an assessment scenario in the background, including the project work syllabus, course material, specific reverse-engineering activities, and a knowledge-, skill-, attitude-, and competence-based map. Corresponding assessment tools and targets for both teams and individuals complement this tool.

To gain insights into the challenges and improvements needed for course delivery, both teachers and students can express their perceptions and assess various aspects of the process using semi-quantitative scales. Educational data mining and learning analytics (EDM/LA) play a vital role in extracting hidden knowledge from educational data. These datasets often comprise data collected during course delivery periods from the university's information system and digital learning platforms.

Educators can use tools to evaluate the course content's structure and its effectiveness in facilitating the learning process. These tools can classify students based on their feedback and monitoring perspectives. In some instances, they can even identify regular and atypical patterns in students' behavior, helping to pinpoint their most common mistakes and develop more effective teaching activities.

Beyond the broader domain of course management, it is essential to consider the individual student's perspective. Both perspectives benefit from the knowledge generated through the methods described above, as teaching improvements also contribute to students' success. In the realm of EDM/LA applications, which primarily focus on modeling behavior and evaluating students' learning performance, various documents in the literature discuss theoretical concepts and practical implementations. These systems generate valuable feedback for both educators and students; in fact, they can detect learning behaviors and proactively flag potential issues. They follow a student-oriented approach, recommending relevant activities, resources, curriculum adjustments, or links to help foster and enhance the learning experience.

3.2. Dataset Analysis

3.2.1. Course Data

The analysis in question utilized a dataset, as outlined in Table 2; this was primarily sourced from two distinct data repositories:

1. Course performance records: These recordings contain the examination assessments of each student, referring to the 2022–2023 academic year. These records establish a connection between the student's identification and assessments undertaken in various examination sections and sessions. This includes details on the final examination results and scores (i.e., TopQ, MedQ, LowQ, and VLowQ).
2. Students' Personal Data: These data were obtained from Politecnico's information system. They encompass information regarding the universities where students previously received their undergraduate degrees, the type of enrollment, the type of high school, and other potentially relevant details for constructing student profiles.

Table 2. Dataset description.

Attribute	Description	Column
Student ID number	The student identification code	Student ID
Sex	Student gender: F for female and M for male	GEND
Project work grade	Exam assessment: project work scoring, range [0, 25]	PJEV
Project work group identifier	Identifies the group to which the student belongs	PJGR
Open question section	Exam assessment: Arduino platform scoring, range [0, 20]	Arduino
Python/UML section	Exam assessment: reverse-engineering scoring, range [0, 30]	Python
Engineering cultural area	Student's engineering competence area [Industrial (L-9), Information (L-8)]	TYP
Bachelor's degree University	University where each student attained their bachelor's degree. Three values allowed: POLI (for Politecnico), FOREIGN (for students who earned their degree abroad), and ITALIAN NOT POLI (for students who graduated in Italy, but outside Politecnico)	BD-UNI
Bachelor's degree mark	The student's bachelor's degree mark, range [0, 110]	BACH-MK
High school	Type of high school attended	TY-HS
High school exam mark	High school final exam score	HS-SCOR
Master's degree mark average	The student master's degree average score	MS-MK
Italy Zone	Student's origin, four values: NORTH, CENTER, SOUTH/ISLANDS, and ABROAD; the first three options feature the Italian student's birth region	ZONE
Residence in the country	Specifies the town in which the student has residence	RES-TOWN
Number of seminars and conferences	Seminars and conferences students attended	NS&C
Final grade	Weighted sum of the three partial assessments, range [0, 30]	F.SCOR

The student assessments encompass three sections, each defined by an attribute specified within the dataset. The primary objective of this analysis is to categorize students based on their examination scores. Consequently, the clustering process revolves around three key assessment sections.

1. Python/UML-REng: This measures students' proficiency in Python coding and UML modeling using reverse-engineering techniques. This evaluates theoretical knowledge acquired during class lectures, labs, and from reading the corresponding notes and textbook chapters.
2. Project Work: This section is based on the autonomous development of project work by student teams. These projects address real-world issues that are typically proposed by business entities that stand to benefit from participating in the process and capitalizing on potential outcomes derived from prototyped solutions.
3. Student Behavior: This measures the student's involvement in different activities, such as seminars and conferences.

Students are identified by an anonymized ID, which ensures that their career performances are tracked. The Politecnico di Torino DPO has the authority to associate the real identity of the students to the anonymized ID used in this research.

The following subsections delve into data examination, employing a systematic, step-by-step approach and assuming related activities as the primary phases:

1. Dataset exploration.
2. Clustering process.
3. Cluster characterization.

3.2.2. Dataset Exploration

The first step in the process involved dataset exploration to validate the data and clean them if needed.

The dataset gathers information that is organized in 22 columns from about 102 students and brings out interesting findings and suggestions. To start with, five students were excluded since they had not yet submitted their project work at the time of writing, as outliers could affect the clustering process. All the remaining 97 students were Italian, mainly residing in Piedmont, the region site of this university. However, a relevant share of the students also came from the southern regions, such as Puglia and Sicily, as depicted in Figure 7.

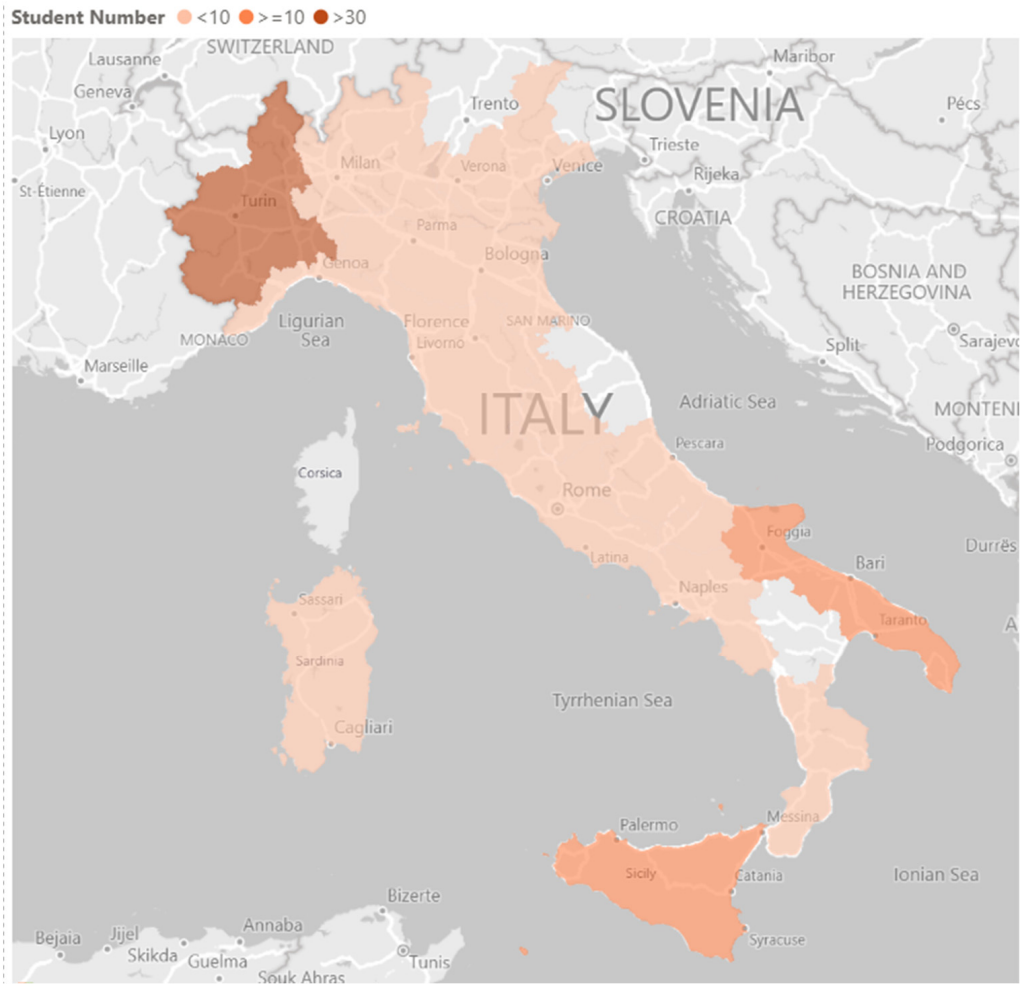


Figure 7. Students' residence per region.

These 97 students completed their bachelor's at the case study university; only 27 of them were educated in other Italian universities.

Then, the attributes relevant to the clustering process were appropriately chosen. The focus was centered on variables that could identify the different skill sets that the course aimed to develop:

- STUDENT ID CODE: used just as a label and not as a core clustering attribute.
- ARDUINO: referring to the IoT platform unit used to sample data from the field.
- PYTHON: a general-purpose language potentially in use for application development.
- PJEV: project work evaluation.

Since these attributes were represented on different scales, they were normalized in a 0–1 range before applying various clustering techniques to identify the attribute that was best suited to the available dataset.

The following section describes the methodology used to select the most appropriate clustering algorithm. Furthermore, once the clusters have been identified, the associa-

tion rules adopted to link them to other attributes in the dataset are also represented to characterize relevant properties, which are useful in shaping the emerging picture of the whole class.

4. Discussion

4.1. Clustering Process

The k-means and k-medoids algorithms require the cluster number K as an input to identify the best K value. For this purpose, an "elbow graph" was used, where the metric "average within centroid distance" was represented as the number of clusters by which K varies. K was set to 5 in both algorithms, being the value where the curve generated an "elbow". It is also the point that, beyond which, an increase in the number of clusters k does not lead to a significant reduction in the intra-cluster distance and consequently in the algorithm's performance.

To conclude this phase, the choice of the algorithm most suitable for the clustering process was determined by comparing the performance of the k-means and the k-medoids using the Davies–Bouldin index. As highlighted in Table 3 for the selected k-value of 5, the Davies–Bouldin index was close; however, the results suggested that the k-Medoids algorithm could perform better than k-means. Thus, this approach was chosen and tested.

Table 3. Davies–Bouldin index distribution.

Number of Clusters	k-Means	k-Medoids
4	−1.106	−1.059
5	−1.036	−0.976
6	−1.035	−1.266
7	−0.904	−1.379

Analyzing the clusters built using k-medoids, as shown in Table 4, it appeared that close to 90% of the points (students under investigation) were assigned to two dominant collections, leaving others with a residual relevance.

Table 4. k = 5 k-medoids cluster numerosity.

Cluster	Numerosity %
Cluster_1	7%
Cluster_2	33%
Cluster_3	4%
Cluster_4	53%
Cluster_5	3%

On the contrary, the k-means algorithm showed much more interesting figures, even keeping an appropriate Davies–Bouldin index, providing clusters with a better information quality. Hence, the k-means algorithm was finally chosen for this purpose.

The k-means algorithm identified the clusters that are shown in Table 5.

Figure 8 shows a 3D scatter plot of the clusters; this visualization is useful in comprehending their distribution among the attributes used in the clustering process (reported on the axis of the plot).

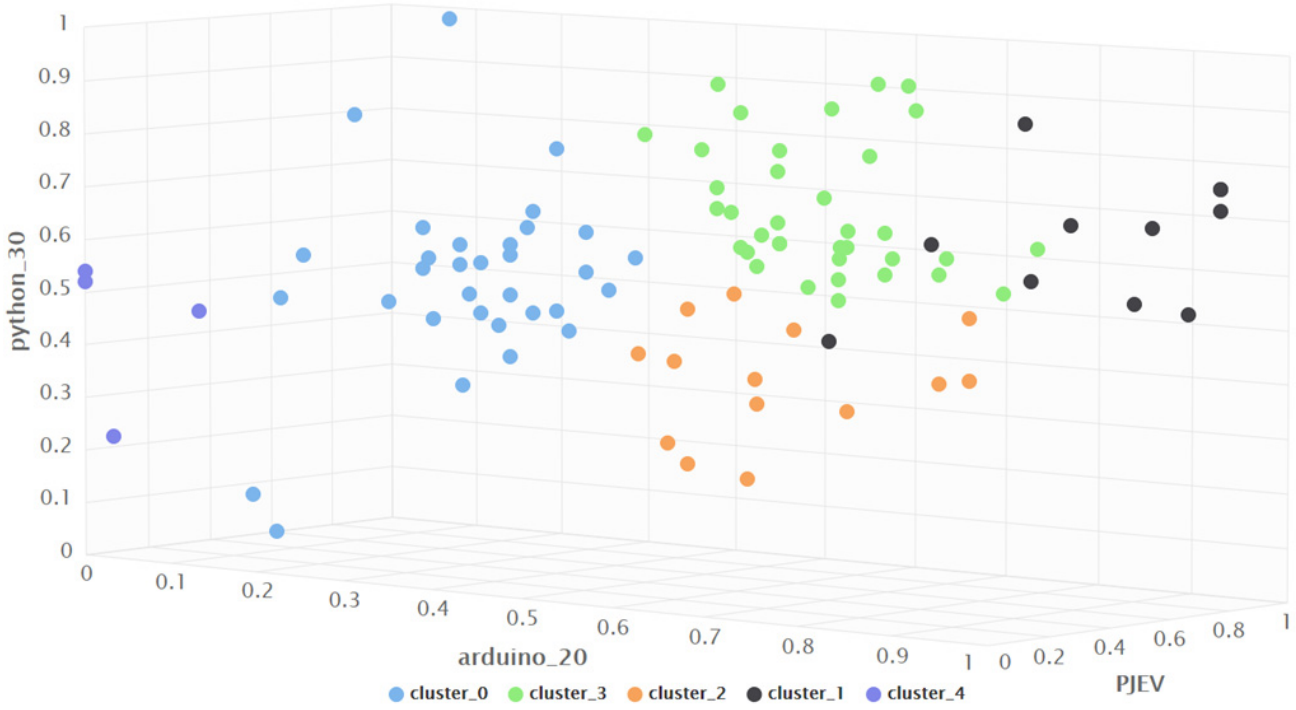


Figure 8. Cluster visualization.

The five clusters identified using the k-means algorithm are shown in Table 5:

Table 5. k = 5 k-means clustering results.

Category	Cluster	Numerosity %	Description
High	Cluster_1	11.4%	Average PJEV, good Python, excellent Arduino
	Cluster_3	37.1%	Good PJEV, good Python, good Arduino
Average	Cluster_2	14.4%	Good PJEV, low Python, average Arduino
	Cluster_0	33%	Low PJEV, average Python, low Arduino
Low	Cluster_4	4.1%	Very low PJEV, low Python, very low Arduino

Hence, the computed clusters show the following properties:

- Cluster 4: four students have a low performance in their project work and a very low score in the Arduino section.
- Cluster 2: fourteen students have good project work, an average Arduino section score, and a low score in the Python section.
- Cluster 0: thirty-two students have a low performance in their project work and a low score in the Arduino section, while the Python section is comparable with other, better performance clusters.
- Cluster 1: eleven students have an excellent performance in the Arduino section, a good performance in the Python section, and an average project work evaluation.
- Cluster 3: thirty-six students show a good performance in all three analyzed sections.

Based on this analysis, clusters 3 and 1 emerge to be the student profiles associated with the best performance, clusters 2 and 0 are the student profiles with average performance, and cluster 4 encompasses the student profiles with the lowest performance.

To further validate those student profiles, the exam score distribution among the clusters is shown in Figure 9. Clusters 1 and 3 expose the highest share of students with very high scores.

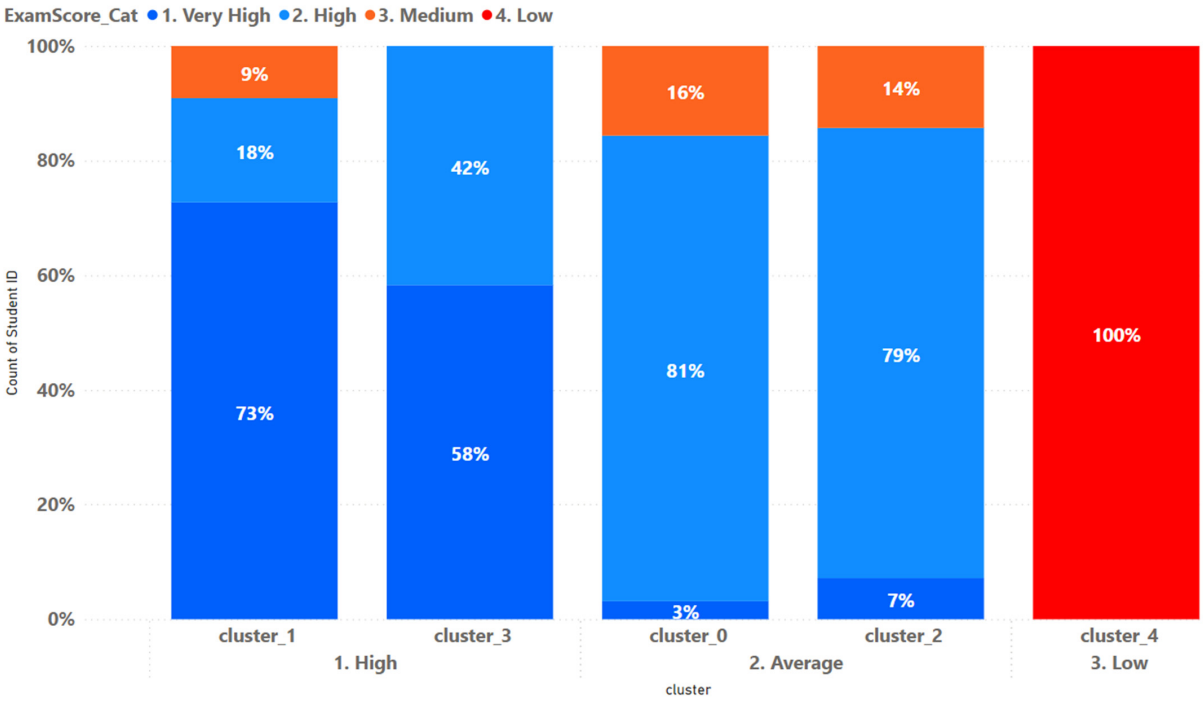


Figure 9. Clusters by exam score distribution.

4.2. Cluster Characterization

To better characterize the five clusters, information about the students' careers is considered to show how those data are distributed among the clusters to extract some further useful insights; this is in accordance with the impact that students' previous careers may have had on their current performance.

For this purpose, the bachelor's degree area is computed in Figure 10, where the higher-performance clusters are associated with a higher share of students who have an information technology (IT) background (L-8). The lower-performance clusters show a higher incidence of students with an industrial background (L-9).

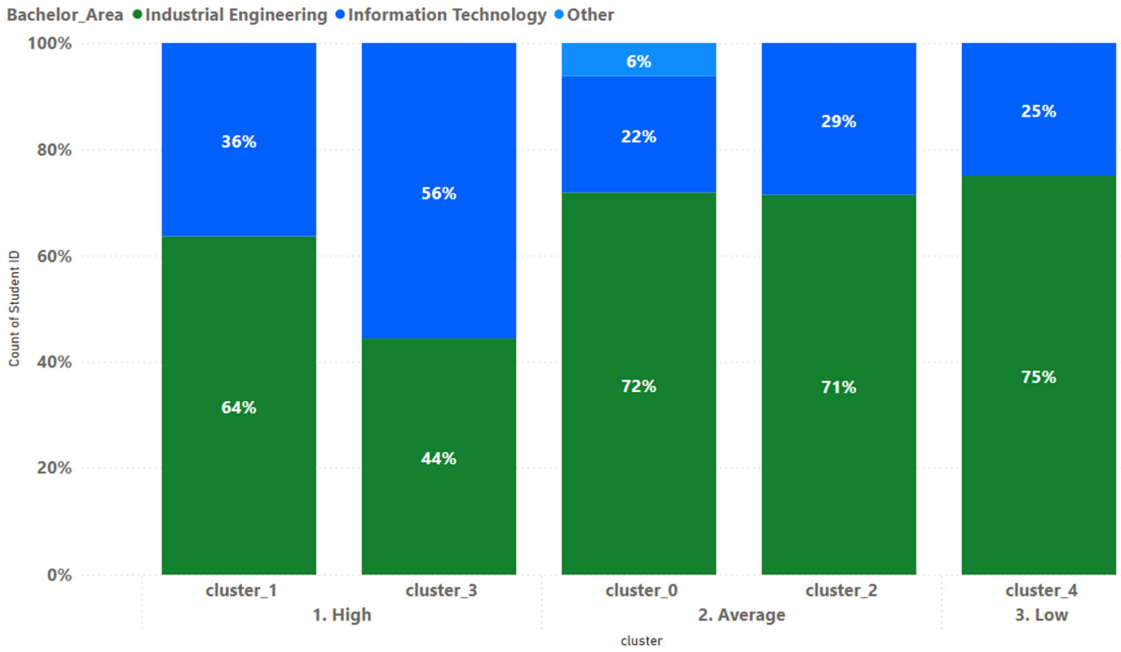


Figure 10. Clusters by bachelor's degree area.

Another interesting finding concerns the final scores that students obtained in their secondary school final exams; these are related to their final bachelor's degree scores and their current average scores in their master's degree. To compare students' performances, their scores are described in Table 6 in four categories.

Table 6. Score mapping.

Category	Secondary School	Bachelor's Degree	Master's [1]
Low	[0; 70]	[0; 80]	[0; 22]
Medium	(70; 80]	(80; 90]	(22; 25]
High	(80: 90]	(90; 100]	(25; 28]
Very High	(90; 100]	(100; 110]	(28; 30]

[1] Note that the students are currently enrolled in their master's degree and this score evaluates the average mark in their exams.

Figure 11 reports students' school exam score distribution per cluster, clearly showing that students achieving the highest final score are also classified within the high-performance clusters (cluster_1 and cluster_3).

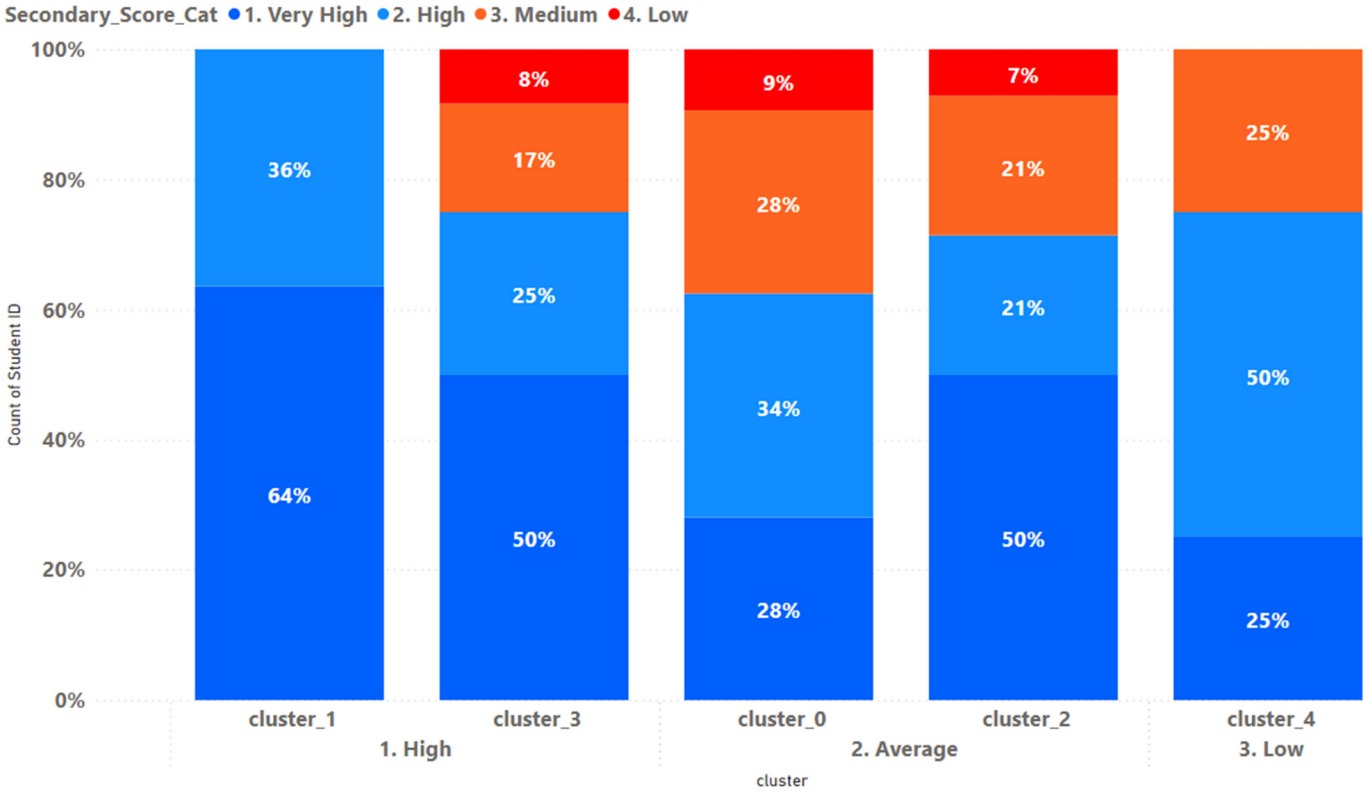


Figure 11. Clusters by final secondary school score.

Analogously, the same trend as before emerges in the distribution of bachelor's degree scores. In this case, it is even clearer that the high-performance clusters gather students with good or excellent performance in their career path, as shown in Figure 12.

This analysis ends by considering the students' current performances in the master's degree where they are currently enrolled; even here, the same figures as before are confirmed.

In summary, students classified in high-performance clusters typically had good or excellent performance in their previous career path; on the contrary, students who struggled during their previous education path can mostly be found in the lower clusters, as shown in Figure 13.

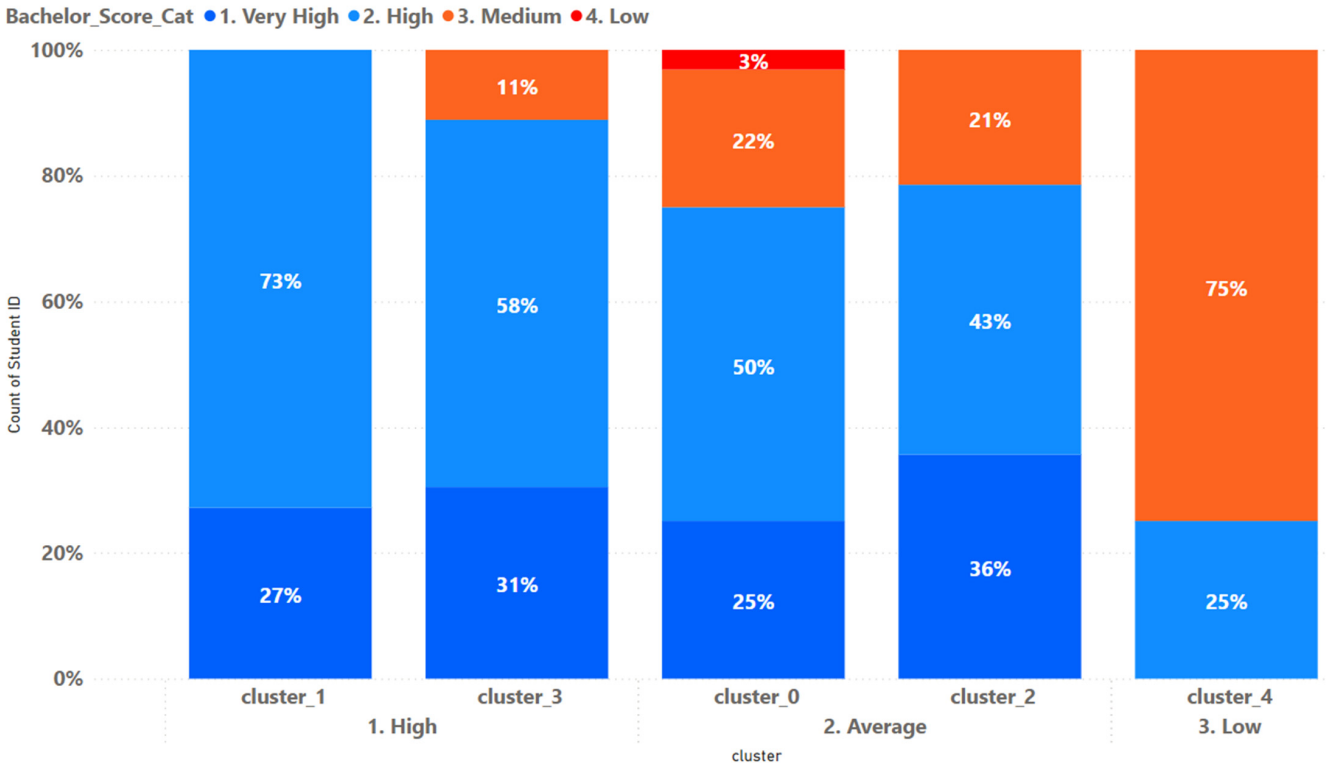


Figure 12. Clusters by bachelor's degree score.

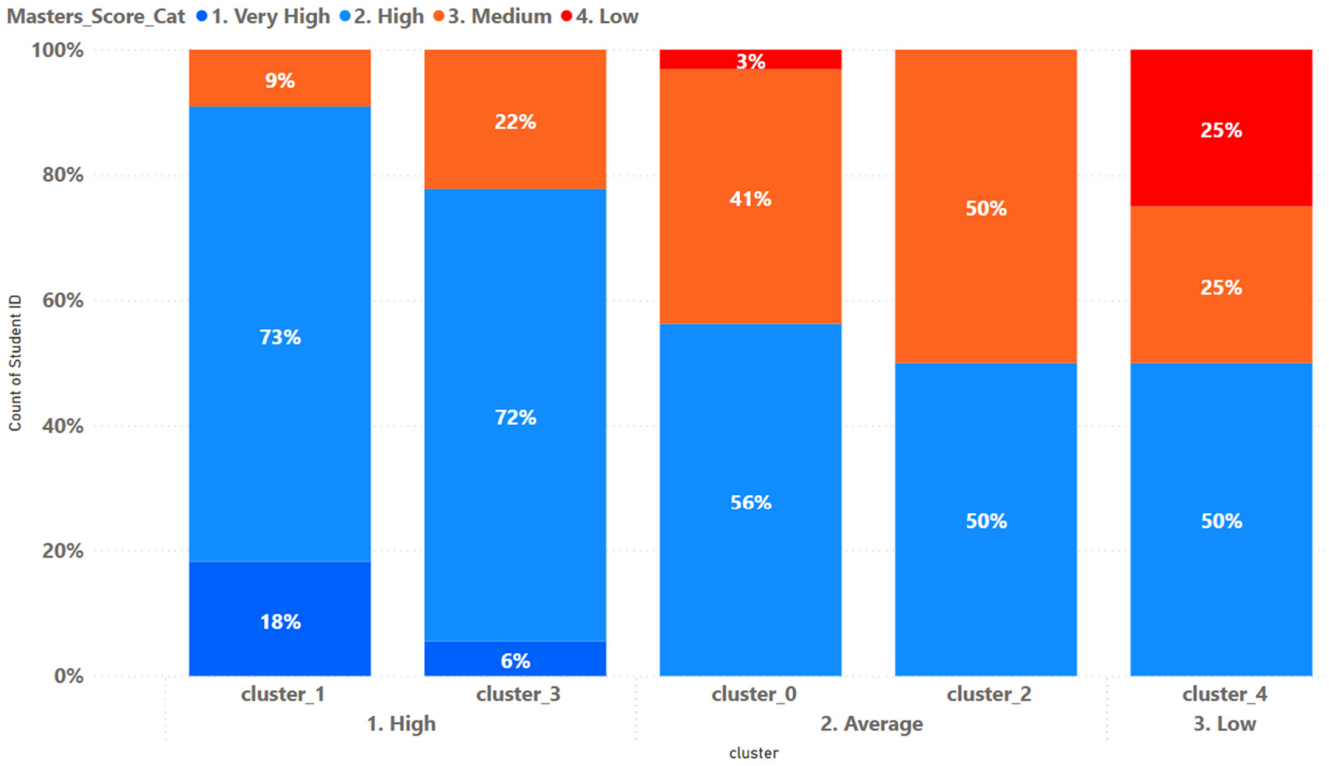


Figure 13. Clusters by current average master's degree mark.

The limits of this investigation mainly concern the nature and features of the students involved in the experience. Only 100 students were considered; this remains the case, even in the event that a further class is included for the implementation of this work, according to plans that are already underway for the 2023/24 academic year. The intention of this is

to widen the impact and effectiveness of this analysis. Furthermore, the students enrolled in the course show some common features: all of them are Italian and come from similar backgrounds in their bachelor's degrees, despite having different specializations.

In the literature, analogous investigations have been conducted that have worked with clusters concerning the study approach that different kinds of students have been exposed to [21]. Other relevant studies have also been carried out in the school [22], showing interesting figures that will be helpful in shaping and promoting further analyses on the interlacing paths between forms of higher education (academia).

5. Conclusions

Considering the scenario outlined in the previous paragraph concerning the results, the research questions (RQs) that are central to the project's development—stated since the beginning—were held as a guideline for this investigation.

The primary objective of this study was to explore (RQ1) the potential of educational data mining in supporting teaching at the faculty level and enhancing learning from the students' standpoint. The investigation was centered on the performance of about 100 students pursuing a master's degree in engineering and management at Politecnico di Torino, with a specific focus on "innovation management and product development" (in a course of the same name).

This objective has been presented with the intention of demonstrating the application of specific toolkits (RQ2), such as Excel, PowerBI, and RapidMiner—these are three effective environments in which data can be processed according to different perspectives.

Preprocessing was conducted by working with Excel to obtain the preliminary data representations; these were then extended visually, taking advantage of PowerBI. To supply data to specific clustering algorithms, RapidMiner was used, choosing specific association rules to characterize each identified cluster.

Concerning the expected student profile dealt with in this work, an engineering and management standard (RQ3) was taken as a reference to specify the skills, attitudes, knowledge, and competencies that were to be achieved by the students attending the course—ISO 15288, Technical Standard in Systems Engineering.

Key performance indicators (RQ4) for assessing students have been shaped according to the assessment schema that has been assumed to sample the relevant skills, attitudes, and competencies shown by the same students engaged in specific tests and exams. That schema focused on project work, a reverse-engineering test related to a python code segment, and a hardware platform development that is able to sense and actuate specific physical variables.

Regulatory/control actions (RQ5) have been chosen and proposed by an intelligent decision-support system (IDSS), which can provide appropriate competence recovery proposals to teachers, deans, or other responsible agents. These actions include, as an example, the establishment and subsequent execution of flipped classes, in either physical or virtual living labs; other examples are the provision of further access to appropriate learning materials available online or further discussions with specific tutors on critical subjects.

Key recommendations for future work include the following:

- To perform analogous analyses on data sampled in the four schools selected in the Data2Learn@Edu project, according to already-stated plans.
- Expand data collection to encompass additional student attributes, such as class attendance, knowledge, skills, competencies, and behaviors acquired during their training path; thus, the dataset will be enriched and enlarged; it will potentially cope with the new wider privacy policies that will be implemented during adoption.
- Emphasize data quality to fully leverage mining approaches, striving for high-quality data at various granularities.

This research represents an initial step in a series of planned trials that are aimed at a more targeted exploration of novel information and unforeseen behavioral dynamics in primary, middle, and high school and university pathways. The current work adopted

an explorative approach by applying educational data mining techniques to a university course.

The analysis revealed that inadequate coding skills significantly contribute to low performance within the "Innovation management and product development" course. Addressing the widespread issue of limited digital and computing competencies requires a systemic approach, advocating for the early introduction of these skills, even at the primary school level. The Data2Learn@Edu Project, active in Turin since March 2023, strives for innovative teaching approaches based on continuous adaptive learning.

By concentrating on the subset of schools involved in the Data2Learn@Edu project, there is an opportunity to integrate the data provided by those same schools and the Ministry of Education with other information gathered during the project. This collaborative approach could offer valuable insights into the intersection between educational initiatives and data-driven analyses.

According to the sustainability issues outlined in the 2030 Agenda, this project meets Sustainable Development Goal 4 (SDG 4), ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all.

Author Contributions: Supervision, C.G.D.; Software, A.B.; Writing—original draft preparation, C.G.D. and A.B.; Writing—review and editing, L.S. and F.M. All authors have read and agreed to the published version of the manuscript.

Funding: The project Data2Learn@ was partially funded by Fondazione Compagnia di San Paolo, grant n. 71171.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to Politecnico di Torino DPO direct responsibility.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets presented in this article are not readily available due to privacy reasons. Requests to access the datasets should be directed to claudio.demartini@polito.it.

Acknowledgments: LINKS Foundation, Fondazione per la Scuola, Politecnico di Milano, INVALSI.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Nomenclature

Acronym	Definition
EDA	Exploratory Data Analysis
EDM	Educational Data Mining
LA	Learning Analytics
DPO	Data Protection Officer
AI	Artificial Intelligence
DBI	Davies–Bouldin Index
GISP	Innovation Management and Product Development
ID	Identifier
IDEF	Integrated Definition
IDSS	Intelligent Decision-support system
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
LFA	Logical Framework Approach
LMC	Lean Model Canvas
MVP	Minimum Viable Product
NLP	Natural Language Processing
QFD	Quality Function Deployment
RQ	Research Question

SDG	Sustainable Development Goals
UML	Unified Modeling Language
OER	Open Educational Resource

References

1. Hooda, M.; Rana, C.; Dahiya, O.; Rizwan, A.; Hossain, M.S. Artificial Intelligence for Assessment and Feedback to Enhance Student Success in Higher Education. *Math. Probl. Eng.* **2022**, *2022*, 5215722. [CrossRef]
2. OECD Digital Education Outlook. Available online: https://www.oecd-ilibrary.org/fr/education/oecd-digital-education-outlook_7fbfff45-en (accessed on 6 December 2023).
3. What Is Learning Analytics? Society for Learning Analytics Research (SoLAR). Available online: https://www.solaresearch.org/about/what-is-learning-analytics/ (accessed on 12 January 2024).
4. Salas-Pilco, S.Z.; Xiao, K.; Hu, X. Artificial Intelligence and Learning Analytics in Teacher Education: A Systematic Review. *Educ. Sci.* **2022**, *12*, 569. [CrossRef]
5. Herodotou, C.; Naydenova, G.; Boroowa, A.; Gilmour, A.; Rienties, B. How can predictive learning analytics and motivational interventions increase student retention and enhance administrative support in distance education? *J. Learn. Anal.* **2020**, *7*, 72–83. [CrossRef]
6. Ouyang, F.; Wu, M.; Zheng, L.; Zhang, L.; Jiao, P. Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 4. [CrossRef] [PubMed]
7. Rienties, B.; Køhler Simonsen, H.; Herodotou, C. Learning Analytics: A Need for Coherence. *Front. Educ.* **2020**, *5*, 128. [CrossRef]
8. Demartini, C.G.; Bosso, A.; Ciccarelli, G.; Benussi, L.; Renga, F. Adaptive Learning Profiles in the Education Domain. In *Artificial Intelligence in STEM Education*; CRC Press: Boca Raton, FL, USA, 2022.
9. Jones, K.M.; McCoy, C. Reconsidering data in learning analytics: Opportunities for critical research using a documentation studies framework. In *The Datafication of Education*; Routledge: London, UK, 2020; pp. 69–80. Available online: https://www.taylorfrancis.com/chapters/edit/10.4324/9780429341359-6/reconsidering-data-learning-analytics-kyle-jones-chase-mccoy (accessed on 6 December 2023).
10. Riconnessioni Project Website. Available online: https://www.riconnessioni.it/ (accessed on 9 December 2023).
11. Brancaccio, A.; Marchisio, M.; Palumbo, C.; Pardini, C.; Patrucco, A.; Zich, R. Problem Posing and Solving: Strategic Italian Key Action to Enhance Teaching and Learning Mathematics and Informatics in the High School. In Proceedings of the 2015 IEEE 39th Annual Computer Software and Applications Conference, Taichung, Taiwan, 1–5 July 2015; pp. 845–850. [CrossRef]
12. Tang, H. Implementing open educational resources in digital education. *Educ. Technol. Res. Dev.* **2021**, *69*, 389–392. [CrossRef] [PubMed]
13. IEEE Standards Association. Available online: https://standards.ieee.org (accessed on 12 December 2023).
14. Dol, S.M.; Jawandhiya, P.M. Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining—A survey. *Eng. Appl. Artif. Intell.* **2023**, *122*, 106071. [CrossRef]
15. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–298. Available online: https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992 (accessed on 6 December 2023).
16. Vankayalapati, R.; Ghutugade, K.B.; Vannapuram, R.; Prasanna, B.P.S. K-means algorithm for clustering of learners performance levels using machine learning techniques. *Rev. D'intell. Artif.* **2021**, *35*, 99–104. [CrossRef]
17. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *1*, 224–227. [CrossRef] [PubMed]
18. *ISO 15288*; International Standard—Systems and Software Engineering—System Life Cycle Processes. ISO: Geneva, Switzerland, 2015.
19. Intelligent Decision Support in Healthcare | Analytics Magazine. Available online: https://pubsonline.informs.org/do/10.1287/LYTX.2012.01.05/full/ (accessed on 15 January 2024).
20. Paravati, G.; Lamberti, F.; Gatteschi, V. Joint Traditional and Company-Based Organization of Information Systems and Product Development Courses. In Proceedings of the 2015 IEEE 39th Annual Computer Software and Applications Conference, Taichung, Taiwan, 1–5 July 2015; pp. 858–867. [CrossRef]
21. Asikainen, H.; Salmela-Aro, K.; Parpala, A.; Katajavuori, N. Learning profiles and their relation to study-related burnout and academic achievement among university students. *Learn. Individ. Differ.* **2020**, *78*, 101781. [CrossRef]
22. Mergoni, A.; Soncin, M.; Agasisti, T. The effect of ICT on schools' efficiency: Empirical evidence on 23 European countries. *Omega* **2023**, *119*, 102891. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.