




Integrated learning pathways in higher education: A framework enhanced with machine learning and semantics

Omiros Iatrellis¹  · Ilias K. Savvas¹ · Achilles Kameas² · Panos Fitsilis¹

Received: 1 November 2019 / Accepted: 9 January 2020 / Published online: 24 January 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

The present research work proposes the development of an integrated framework for the personalization and parameterization of learning pathways, aiming at optimizing the quality of the offered services by the Higher Educational Institutions (HEI). In order to achieve this goal, in addition to the educational part, the EDUC8 framework encloses the set of parameters that cover both the technical and the financial dimensions of a learning pathway, thus providing a complete tool for the optimization and calculation of the offered services by the HEIs in combination with the minimization of respective costs. Moreover, the proposed framework incorporates simulation modeling along with machine learning for the purpose of designing learning pathways and evaluating quality assurance indicators and the return on investment of implementation. The study presents a case study in relation to tertiary education in Greece, with a particular focus on Computer Science programs. Data clustering is specifically applied to learn potential insights pertaining to student characteristics, education factors and outcomes. Generally, the framework is conceived to provide a systematic approach for developing tertiary policies that help optimize the quality and cost of education.

Keywords Semantic meta-modeling · Machine learning · Learning pathways · Modeling and simulation · Academic advising

1 Introduction

A core challenge faced by the Higher Education sector is at the same time to improve the educational services while reducing the organizational and financial costs of

✉ Omiros Iatrellis
iatrellis@hotmail.com

¹ University of Thessaly, Larissa, Greece

² Hellenic Open University, Patras, Greece

education. A possible solution to this challenge is the utilization of personalized academic procedures alongside with their modeling while covering all the possible subdomains of higher education. However, as numbers of students grow the capacity for supporting and guiding each one diminishes, thus more innovative approaches are required to serve the provision of personalized education plans. Important information such as interests, prior academic performance, aspirations and motivations, personal barriers to study and other complex aspects of each individual that frame and influence educational experience over a period of time, is often reduced to rote enrolment information filled out at admission. Yet it is within this crowded, complex and often under-resourced higher education environment that taking into account who the students are at admission, what they want, how they change as they progress, and what they may become beyond their studies, has become more important than ever before in order to achieve the aforementioned challenges.

Cost minimization will allow HEIs to remain competitive and survive in the tough and highly demanding environment of the tertiary sector. One of the key factors embraced by many industries that can make cost reductions possible is standardization. The provision of standardized, and as much as possible, personalized higher education services can yield tremendous benefits from the economies of scale that can be achieved both in terms of tangible as well as of intangible resources of an institution.

The literature is replete with studies that applied modeling and simulation to a multitude of issues pertaining the education context. This paper discusses the design and evaluation of learning pathways in Higher Education. We present a framework to guide the design, development and evaluation of approaches for personalized and pathways-based coordination of higher education. The focus is placed on the design and evaluation of architectures that treat students as agents interacting with systems or services coordinated using higher education pathways and electronic student records. Success in this direction can contribute to the universal goal of optimizing HEIs in terms of curbing costs while improving or maintaining education quality. The framework builds upon previous work developed by (Iatrellis et al. 2019a). Based on the specific work, EDUC8 learning pathways constitutes a technical and standardized tool that allows design, modeling, recommendation and execution of complex, self-evolving and personalized education plans, which incorporate academic knowledge as well as financial and operational knowledge. The basic contribution of this paper is to include machine learning as an assistive artifact to leverage the development of simulation models. Machine learning consists in developing algorithms that learn to recognize complex patterns within valuable and massive data. When applied to higher education, the objectives are often summarized in four main areas (Kučák et al. 2018):

- a) Grading students - machine learning can grade students by removing human biases.
- b) Improving student retention - by identifying “at risk” students early, HEIs can detect and contact those students and help them to be more successful.
- c) Predicting student performance - the technology can identify weaknesses and suggests ways to improve, such as additional practice tests.
- d) Testing students - provides constant feedback to teachers, students and parents about how the students learn, the support they need and the progress they are making towards their learning goals

This study initially started with the aim of grouping students from a data-driven perspective. In this regard, unsupervised clustering using the K-Means algorithm was applied. More specifically, we present the use of unsupervised machine learning techniques to gain insights into identifying academically at-risk students and predicting student's academic career. As an example, this paper demonstrates a case study in relation to Computer Science program of study in Greece.

2 Background and methodology

This section provides a basic background for the remainder of the paper. Initially, the proposed concept of learning pathways is overviewed. Subsequently a framework for designing learning pathways that extends the earlier work by (Iatrellis et al. 2019a) as elaborated in the ensuing discussion.

2.1 The necessity of modeling of learning pathways

The main objectives of a HEI in addition to research are to assure the quality of learning and training, to facilitate students in achieving their academic goals, and to improve the efficiency and effectiveness of the provided services. In this competitive environment of Higher Education, a learning pathway becomes the main concept of educational process (Pavlenko et al. 2017), (Iatrellis et al. 2019b). The specific concept nearly always implies an expansion of educational options beyond the course sequences historically offered to students so as to include multi-facet educational experiences, that occur outside of traditional classroom settings or HEI buildings, such as summer schools, independent research projects, online classes, internships, student exchange programs, or dual-enrollment experiences. While a learning pathway may encompass a wide variety of educational experiences in diverse settings, these experiences impose challenges to HEIs so as to provide support and guidance to students at key decision points or exceptional situations that require appropriate modifications/reconfigurations of the academic plans of a student, thus increasing the flexibility of the learning processes. Any delay in degree completion represents a waste of resources both for the learner and the HEI, thus affecting the returns to investment in higher education. From HEI's point of view, especially in countries where tertiary education is highly publicly subsidized, students who postpone graduation contribute to the misallocation of such resources (Casalone and Orientale 2011).

(Jenkins and Cho 2013) and (Bailey et al. 2015) concluded that a significant number of students experienced academic failure as a result of poor academic procedures (confusing array of academic choices, inefficient provided services and counselling) and that the time to degree rate is extended due to poor decision making about the courses they take and the programs they enter. (Scott-Clayton 2011) explored ideas derived from the fields of behavioral economics and psychology, including choice architecture, to consider how the above can influence students' choices negatively and act as barriers to their success. Two of the most cited studies on this subject were published by (Grunschel et al. 2013) and (Schouwenburg et al. 2006). The imperative need for efficient academic advising and educational provision services with a student-centered focus is a major challenge that is being taken seriously by the global HEI

leaders (Buizza et al. 2019). However, this goal is yet to be fully achieved (Kumar 2016), (Prebble et al. 2004).

Although learning pathway changes and deviations during a program can be avoided to a large extent, they are not unusual for tertiary students. Also, despite the fact that it is quite difficult to identify a correlation between extension to time to degree and poor academic choices, the personalization and standardization of learning pathways for a specific group of students with a predictable academic course can become effective tools in reducing the abovementioned risks and lowering the probability of academic failure (Quinn et al. 2009) (Fernando et al. 2010). Personalization and standardization of intra-organizational processes are two strategic objectives that other industries have long ago embraced. In this context, several initiatives and reforms in education such as the Guided Pathways (Jenkins and Cho 2013) and ASAP (Oliveira 2016) models, aim at re-engineering processes in HEIs by sharing an emphasis on acceleration, greater transparency of paths to completion for students, and more effective advising from day one through completion. The provision of as much as possible standardized Higher Education services and academic guidance towards personalized learning can provide tremendous benefits from the economies of scale that can be achieved both in tangible and intangible resources of a HEI.

A model represents a structural design of the system that we want to build and a set of decomposed decisions in building the system, in which a multi-level abstraction is necessary for communicating the design of a complex system clearly (Wu 2011). (Booch 1999) made this point clear on the necessity of modeling in building systems where teamwork is required. This is the case in development of modern HEI management systems, and this is the reason why modeling of learning pathways is important in the field of tertiary education.

2.2 Learning pathways are introduced as a concept, model, process and product

In personalized e-learning environments, a learning pathway is described and studied as the chosen route, taken by a learner through a range of educational activities, which allows them to build knowledge progressively (Janssen et al. 2011). Based on this definition, the "conventional" learning pathway planning can be perceived as finding a path in an educational activities network that leads to one or more learning goal(s), which is analogous to the destination node in network routing (Wong and Looi 2009).

However, in formal education, the interaction between a student and an institution, from application, to enrolment, to course completion and finally to the award of a degree, is, in essence, a set of time-dependent processes. These processes which are related to both the way education is being provided and the administration of HEIs have shifted in recent years to the adoption of business process management principles that were historically the domain of manufacturing companies and, more recently, the service industry. HEIs were slow to adopt such practices because it meant approaching the institution with a business mentality rather than through the lens of an education service provider. Today we know that each HEI can be both an institution that cares about its students and an organization making business decisions that improve efficiency and save money.

In the following paragraphs, the way in which learning pathways can be transformed from an abstract concept into a model, process and final product for a HEI is described with the learner being in the center of this attempt (Fig. 1).

An important assumption of the proposed approach is the perception of the learning pathway as a sequence of processes instead of educational activities. As it is presented in Fig. 2, a learning pathway consists of Higher Education “business” processes some of which may require decision-making while others correspond to pure executable parts of learning pathways. The EDUC8 learning pathway is totally in line with the notion of business process management as it applies in Higher Education, by providing a holistic conceptualization of processes that occur in HEIs. At the same time, EDUC8 provides the conceptual foundations for leveraging academic advising decision support services which now become part of learning pathway management enhancement. Thus, this attempt goes in the direction of merging two research areas, namely the business process management and the academic advising in Higher Education, which are usually studied unconnectedly.

Different models of learning pathways exist, and can provide a mechanism to support the quality and efficiency improvement process. For modeling learning pathways, two important issues, can be identified: a) the confidence level of the academic advisors and b) the level of predictability of the academic course. Based on these two issues, three different models of pathways can be described, which define a network topology of linked nodes represented as rhombi in Fig. 3: a) bus models, b) tree models, and c) complex models.

Bus models can be used for high predictable learning processes with a high level of confidence from the academic advisors and predictability level. This mechanism can be used for example in cases in which there are no alternative sub-pathways for the student to select or specialize in. For these processes, pathways can be used as time–task matrixes. The sequence of the timing will mostly be term-by-term. For some shorter learning programs, half terms or month-by-month timings will be used to describe the timing of the process.

Tree models are used for less predictable processes in which students face dilemma while choosing their academic direction. This model can be used for learning processes, which contain critical decision nodes to represent specific milestones such as specialization field declarations. However, the tree model maintains linearity concerning the start and end points.

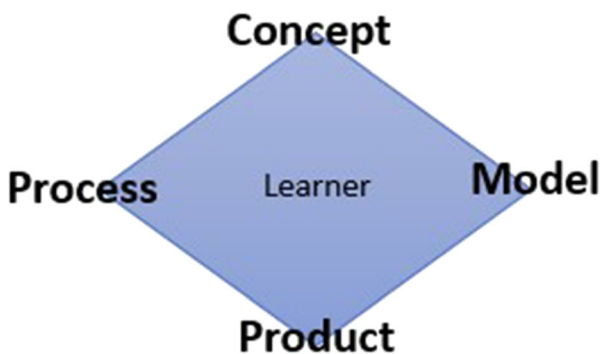


Fig. 1 Learning pathways are introduced as a concept, model, process and product

Complex models can be used for unpredictable learning processes with low confidence from the academic advisor and low predictability level in which it is necessary to have frequent academic advising meetings to be able to organize and structure the process. In such models the “time”–task matrix can be changed into a “goal”–task matrix. Examples can be PhD programs in which research is in most cases uncertain.

Learning pathways are also introduced as a process on their own to develop and implement well-organized intra-organizational processes and to improve quality and efficiency. Thus, in order to improve the organization, it will be necessary to involve a multidisciplinary team consisting of academic advisors of various expertise as well as administrative and managerial personnel. Active ingredients of this educational reform will be the feedback on the actual organization on the educational process, the availability of academic advising recommendations and outcome indicators and the continuous quality and efficiency improvement process that will take place within the multidisciplinary team. Over time, the team will improve the quality and efficiency of the learning process by analyzing the actual organization and performance of the learning pathways. Based on the bottlenecks, data insights and feedback, the team can improve the process by using the plan-do-study-act cycle for continuous improvement with respect to student interests, requirements and aspirations.

Next to considering a learning pathway as a concept, model and process leading to quality and efficiency improvement, it is important to consider it as a product; however, from HEI’s perspective, without the learning pathway concept, process and method, the product is incomplete. The learning pathway product is the final output, which constitutes the sequence of educational activities being studied by the research community.

2.3 Needed domain knowledge streams for modeling the learning pathways

The first step towards modeling learning pathways is knowledge representation. It involves the abstraction of domain-specific knowledge concepts that encapsulate the domain knowledge, problem-solving behavior and operational processes. The design and implementation of the proposed approach was performed so as to meet a set of specific goals concerning the modeling of learning pathways and their dynamic composition. These goals (Iatrellis et al. 2019b) are presented as follows:

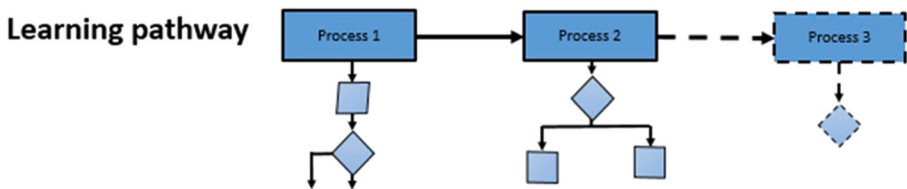


Fig. 2 The learning pathways concept

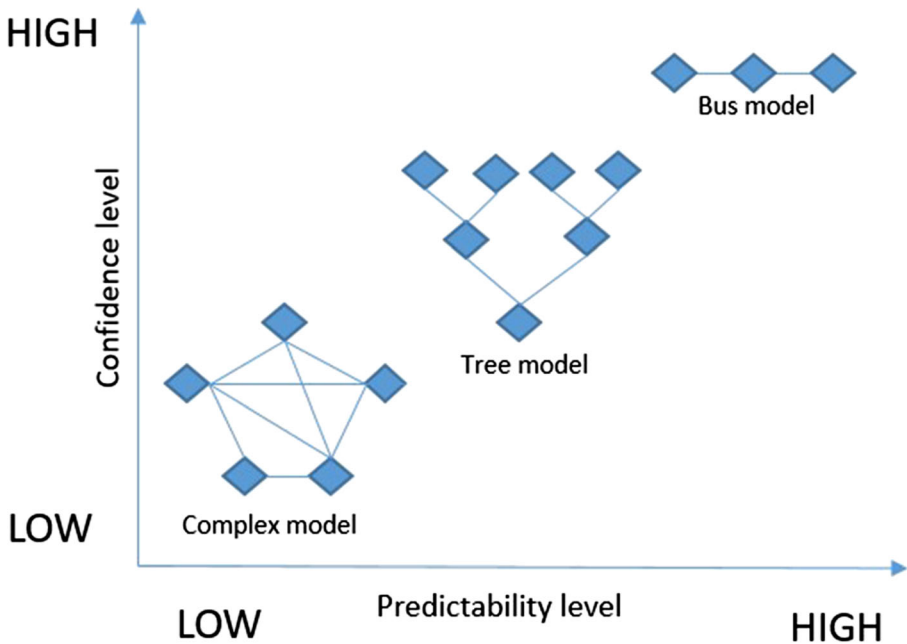


Fig. 3 Different models of learning pathways

- The proposed approach is intended to cover the modeling of the academic advising part of learning pathways for a HEI. Thus, the experience and knowledge of the academic advisor is important to be included.
- A core concept of the proposed approach is to attempt to cover the business part of a learning pathway by modeling both the financial and organizational aspects. The organizational part corresponds to the intangible and tangible resources of a HEI, while the financial part comprises the corresponding financial flows of entities related to a learning pathway.
- The proposed conceptualization aims at modeling the education domain of a HEI as well as the environment of operation (in terms of collaborating companies through internships and cooperative study programs, MOOC providers, partners and other organizations or individuals such as adjunct faculty). Therefore, it requires a holistic conceptualization of the participating entities in the execution of a learning pathway.
- Finally, a core concern is to model the set of concepts that represent important learner parameters, which influence their decision-making. For this reason, a suitable learner model should be implemented and included.

Based on the abovementioned goals, the authors implemented the EDUC8 Ontology (Iatrellis et al. 2019b), which models the needed domain knowledge streams for learning pathways in four (4) main domains (Fig. 4): a) the learner domain, b) the learning pathway, c) the organizational domain, and d) the quality assurance domain.

The first one concerns the learner model, and stores information about the learner that represent important parameters affecting decision-making process. More specifically, learner's interests, personality (based on Holland Occupational Themes/RIASEC

(Holland 1997)), requirements and learning state have been included as concepts of the implemented ontology. The second module contains resources that describe the building blocks of the learning pathways. It is utilized by the software prototype for the design of the education workflow. The third module attempts to cover the business and financial dimensions of a learning pathway. The specific module utilizes the REA business ontology introduced by (McCarthy 1982) and conceptualizes the various transactions that occur at multiple levels of the HEI during the learning pathway execution and affect its evolution. Finally, the fourth one models the "quality assurance" (QA) procedure in the Higher Education domain by implementing the EFQM quality assurance model ("The EFQM Excellence Model," 1988) and a set of QA indicators derived from U-multirank ("U-Multirank | Universities compared. Your way,," 2011).

The four parts provide the ontological infrastructure of EDUC8 environment for the modeling of learning pathways, in terms of structure and content. Moreover, they leverage the semantic modeling of the rules that cover the academic advising part during the execution of a learning pathway. These rules are the cornerstone of the dynamic personalization and adaptation performed in the learning pathways.

2.4 The EDUC8 framework

The main goal of the EDUC8 framework is the specification and implementation of an approach and integrated IT system that leverages the provision of highly personalized academic services from the tertiary institutions. EDUC8 approach is based on a meta-model learning pathway establishment (Iatrellis et al. 2019a). Each executable learning pathway is a meta-model of a set of child and parent sub-processes. The child processes are executable parts of the higher education business process forwarded to the execution engine. The parent processes are sub-workflows, which contain child processes and a set of decisions leading to a unified result. The child and the parent processes are

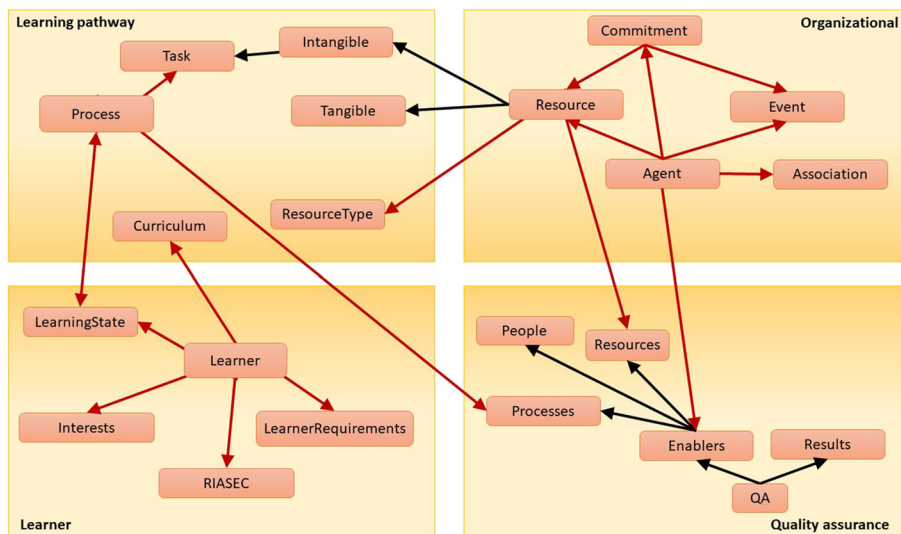


Fig. 4 EDUC8 ontology abstract diagram

interconnected in the meta-model level. Their connections are based on SWRL rules. Once a child or parent process is executed, the rule-base is triggered. The knowledge stored inside the ontology, the student parameters and the rule-set are interoperating in order to select the next executable part of the learning pathway. Thus, the personalization occurs during each decision node of the pathway execution.

Figure 5 depicts in a graphical way the EDUC8 system, which is based on semantic meta-modeling:

1. A learning pathway consists of a set of parent processes, child sub-processes and a set of decisions. The connections between parent processes are based on SWRL rules.
2. In order for the system to recommend the next step of the learning pathway, the rule engine is triggered.
3. The knowledge that is modeled as part of the EDUC8 system and more specifically the student's current learning state, academic background, interests, learning requirements and personality type is used as input to the rule engine.
4. The rule-set is executed by combining the input from step 3 and the process that was completed. The result of the specific interoperation is the next appropriate step of the learning pathway to be proposed.
5. The proposed step is sent to the Learning Pathway Manager component.
6. The specific component selects, and
7. Executes the corresponding sub-workflow in order to proceed to the next step of the education plan.

During each cycle of execution, the triggering of the rule-base may result to new knowledge creation that will be utilized in next steps during the execution. This fact ensures the constant update of academic advising, operational, organizational and quality related knowledge stored inside the EDUC8 ontology and consequently to the rule-base. In order to facilitate the continuous maintenance of the rule base in an integrated way by the domain knowledge experts, a graphical semantic rule generator interface was implemented. The rules implemented with the specific tool are stored as an SWRL repository.

In order to perform the evaluation of the proposed framework, the necessary technological infrastructure has been implemented the core of which is based on the EDUC8 approach and on a set of semantic web technologies. The architecture of the software infrastructure that has been designed supports two distinct modes: the design mode and the execution mode. During the design mode, the respective actors maintain the accumulated knowledge inside the system and implement the appropriate semantic rule set, which models the domain knowledge for the learning pathway composition. During the execution mode, the EDUC8 software platform executes all the required processes so as to facilitate the totally personalized and self-evolving execution of the learning pathway for each student. The technical architecture and the respective business logic that refer to both modes of the lifecycle of learning pathways were presented in (Iatrellis et al. 2019a) and (Iatrellis et al. 2019b). EDUC8 integrated software environment is available online at: <http://www.cs.teilar.gr/EDUC8/>.

The work presented in this paper extends the EDUC8 framework by including a component for incremental learning from data during the design mode. Higher Education services are delivered in data-rich environments where sheer amounts of data are created in HEI’s databases. Machine learning can therefore present a key enabling factor to avail of such data repositories to provide insights for educational improvement, cost cutting or identifying students who are academically at risk.

In this regard, our viewpoint is that unsupervised machine learning suitably serves the purpose of knowledge elicitation at early stages of problem formulation so as to help academic staff to become more proactive. Thus, unsupervised algorithms are employed during conceptual modeling as an assistive artifact to help conceptualize the structure of behavior of the student body. More specifically, data clustering is utilized to discover significant structures or patterns, which represent an abstraction

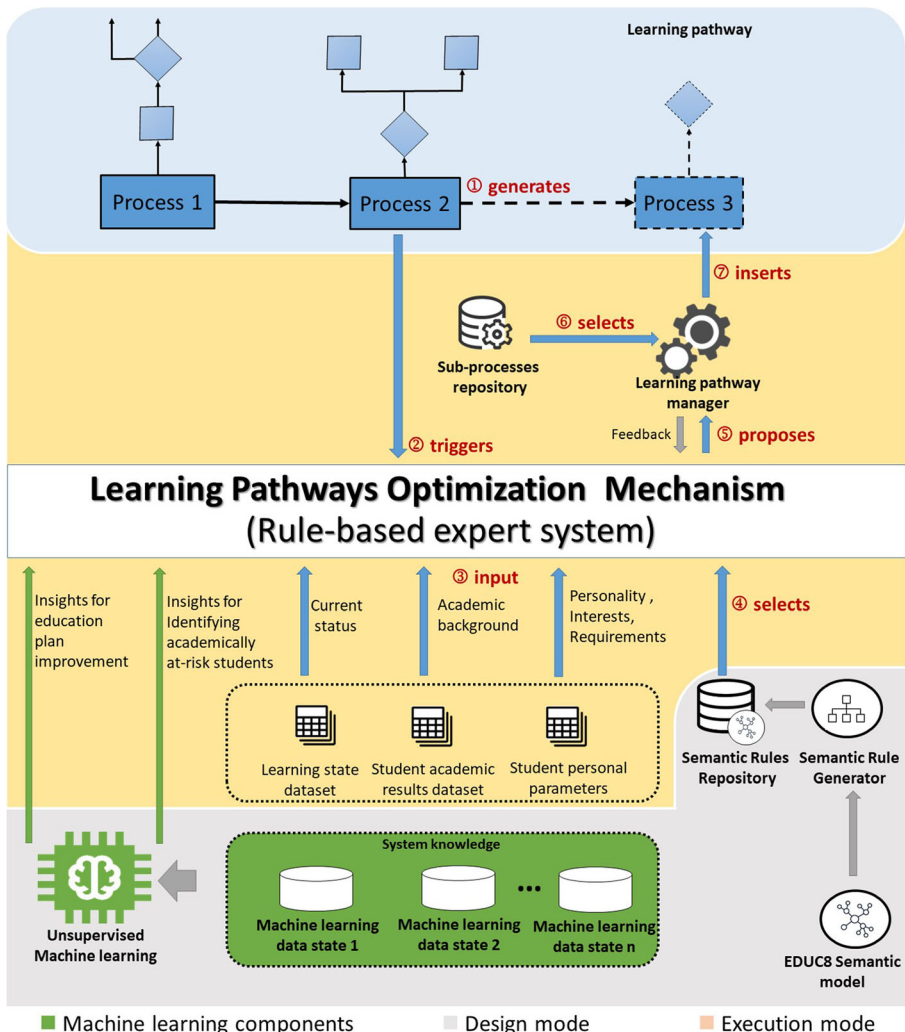


Fig. 5 EDUC8 framework for integrating the learning pathways scheme with machine learning components

of the actual student body under study. This in turn reflects on the simulation model used to design the architecture of a learning pathway. Figure 5 illustrates the framework along with the machine learning components (represented by green color icons).

An important attribute to the extended framework is the idea of incremental learning. The extended framework aims at training and fitting machine learning models on an accumulative and evolving basis, rather than as a single process. The role of incremental learning is predicated on the premise that new parameters (such as student preferences and HEI resources among others) are being repeatedly captured in timely snapshots of data and stored in a repository representing the accumulated system knowledge. In this way, machine learning models can be iteratively trained to learn about possible updates in the EDUC8 system.

3 Case study: Computer science program of study in Greece

In order to verify the effectiveness and completeness of the proposed framework, a specific experimental scenario has been chosen and implemented. The following segments elaborate the case setting, and the development of simulation and machine learning models. The main goal here was to provide a practical scenario where simulation models can be designed or guided in concert with knowledge learned from machine learning as proposed by the EDUC8 framework.

3.1 Overview of the higher education system in Greece

A structured approach to the planning of simulation studies involves identifying aims, data-generating mechanisms, methods, estimands and performance measures (Morris et al. 2019). Likewise, in the machine learning context the perception of business and data is the first important step in the process of data mining (Song et al. 2018). In this sense, this section delivers a basic background of the higher education system in Greece, and its underpinning components.

Starting with an outlook on the entrance options to HEIs, admission to Greek universities takes place through centrally organized, nation-wide exams for candidates who have obtained a High School graduation certificate. There are the two main types of High School existing today which can both lead to a HEI: General High Schools (General Lyceum, acronym in Greek: GEN) and Vocational High Schools (Vocational Lyceum, acronym in Greek: EPAL). Apart from general education courses, Vocational High Schools also provide technical and vocational courses as well as laboratory sessions.

Over the past decades, the existence of performance metrics for higher education can inform stakeholders about HEI outputs, resource allocation and learning outcomes. Surprisingly, the use of completion rates as a measure of success in higher education is a fairly recent development. In Greece, the availability of good national numbers started as recently as the early 2000s. Thus, according to a report from Hellenic Quality Assurance and Accreditation Agency (HQA), many students at Greek HEIs do not graduate on time (HQA 2017). In response to the foreseeable challenges, the education system in Greece has been undergoing substantial reform efforts in recent years aiming at optimizing the quality of the offered services by the HEIs in conjunction with the minimization of the respective costs of education. However, this goal is yet to be achieved.

3.2 Our focus: Computer science program in Greece

The design, implementation and execution of the case study were carried out in the Computer Science and Engineering department of TEI of Thessaly, Greece. TEI of Thessaly is a Higher Educational Institution offering degrees in the traditional, face-to-face way. The curriculum of the specific department comprises the basic cycle (semesters 1-4) and 3 parallel specializations (semesters 5-8). The basic cycle is common to all students and provides the necessary background knowledge in mathematics, physics, informatics, electronic & computer engineering and telecommunications. After the basic cycle, students enroll in one of three 3 distinct areas of specialization: Software Engineering, Network Engineering, or Computer Engineering, where they attend more applied and technology-oriented courses in Computer Science and Engineering, most of which are related to the area of specialization they have chosen. Every year, the Computer Science and Engineering department admits a student intake of about 250 students (including student transfers) with about 10% EPAL students.

The grading system in Greece is between the scale of 0-10 (0% – 100%) and the pass (module) is 5 (50%). The grading system of the HEIs is linear and can be well described as: Excellent (8.50 to 10), Very Good (6.50 to 8.50), Good (5 to 6.49) and Fail (0 to 4.90)

3.3 Data description

The HEI's database for the particular department archives abundant information about the students' studies. Specifically, a typical student record includes various data fields about a variety of information collected from application, to enrolment, to course completion and finally to the award of a degree. We acquired a dataset that covered three academic years from 2011 to 2013, which contained 200 records excluding students who graduated on time (4 years). All the records were fully anonymized for the purpose of privacy.

3.4 Unsupervised machine learning: Discovering student clusters

Clustering is defined as the grouping of similar objects or the process of finding a natural association among some specific objects or data (Ghiasi et al. 2002). As such, the student database records presented a good opportunity to realize the grouping of students from a data-driven viewpoint. The attempt was to group students based on the similarity of characteristics, education-related factors, and outcomes. The clusters could help explore the similarities and differences between participating entities (i.e. students) that are related with the learning pathway under study. The following sections briefly summaries the data pre-processing procedures, and clustering experiments. A detailed analysis of those aspects would go beyond the scope of this study.

3.4.1 Selected features

Initially the clustering model included Time To Degree (TTD) and Time To Complete Basic Cycle (TTCBC) as features. More features could be extracted from the dataset, which are related to quality standards as described in relevant research works (Kappe and Van Der Flier 2012) (McKenzie and Schweitzer 2001). In this respect, two quality measures could be captured from the student database:

- Higher grades in High school and entrance exams can be associated with better university results and higher grades
- Performance in first year core subjects is one of the most commonly found metrics for student success

Our intuition was those factors could have an influence on education outcomes, and they can therefore serve as candidate in the clustering model. In this way, two more features were added named as Grade In Entrance Exams (GIEE) and First Year Performance In Core Subjects (FYPICS). However, the GIEE was subsequently excluded as it contained a significant amount of missing values. Instead, the type of student's high school (GEN or EPAL) was used as a factor that could have potential influence on education outcomes.

3.4.2 Clustering experiments

One of the most significant clustering techniques is k-means algorithm (MacQueen and others 1967). The algorithm requires as input the number of clusters (k), and starts by choosing k initial centroids. Then, k-means assigns the data points to their closest centroid by calculating their distance. In this study Euclidean distance between points was used. After that, the centroids are recalculated and the algorithm terminates when the centroids do not change position (or when some convergence criteria are fulfilled). In this work we implemented k-means for k varying from 2 to 5 and in addition we applied the Principal Component Analysis (PCA) (Pearson 1901) in order to reduce the dimensionality of data and increase its visualization. The quality produced by PCA was 96.95% by reducing the dimension from 3D to 2D, thus the results obtained are extremely accurate considering their quality. To implement the above experiments one computational node was used (AMD@ FX(tm)-4100 Quad-Core Processorx4, with UBUNTU 16.04 LTS operating system, gcc 4.84, and Python 3.7).

In Fig. 6, with $k=2$, the output indicated a promising tendency of clusters, where the data is obviously separated into two big clusters. Likewise, when $k=3$ the clusters are still well-separated. However, when k is greater than 3, the produced clusters only separate well defined groups so it is clear that this is the most significant division of the data set. The general characteristics of the three clusters are briefly presented in Table 1. The table compares in terms of specific education-related factors.

Table 1 Summary of cluster characteristics

	~ % of dataset	High school		Avg. FYPICS	Avg TTCBC (Years)	Avg. TTD (Years)
		% GEN	% EPAL			
Cluster 1	18,50%	15,40%	84,60%	8.0	2.6	5.9
Cluster 2	36,00%	19,60%	80,40%	7.0	3.3	5.9
Cluster 3	45,50%	18,30%	81,70%	6.1	4.7	7.5

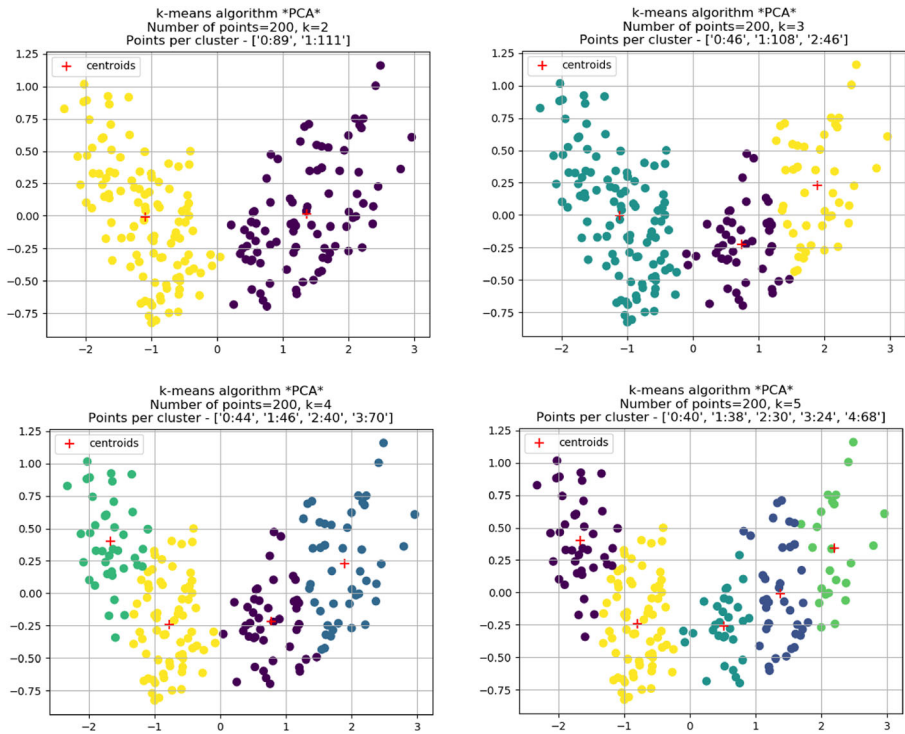


Fig. 6 Visualization of student clusters with K ranging from 2 to 5

3.4.3 Learning insights from clusters

Initially, the clusters were inspected with respect to the potential impact of TTCBC on TTD. The student TTD is of significant concern in Higher Education as a measure of education outcomes. From an operational standpoint, the TTD was considered as a valid proxy to measure the consumption of HEI resources. Further, it was also reported that the TTD largely accounts for the cost of student education.

Exploring clusters revealed that the students in cluster 3 tended to have a relatively wider dispersion of TTCBC compared to Cluster1 and Cluster2. Interestingly, Cluster 3 students also experienced longer TTD compared to Cluster 1 and Cluster2, which shared a very similar distribution of the TTCBC and TTD variables. Figure 7a plots the TTCBC variable in the three clusters, while Fig 7b plots the TTD variable.

Subsequently, the FYPICS structure of clusters was explored. Figure 8 plots the variations of FYPICS with respect to the three clusters. On one hand, it can be observed that Cluster1 and Cluster 2 included mostly the students with very good and excellent grades, while Cluster3 contained mostly students with good average performance in first year core subjects.

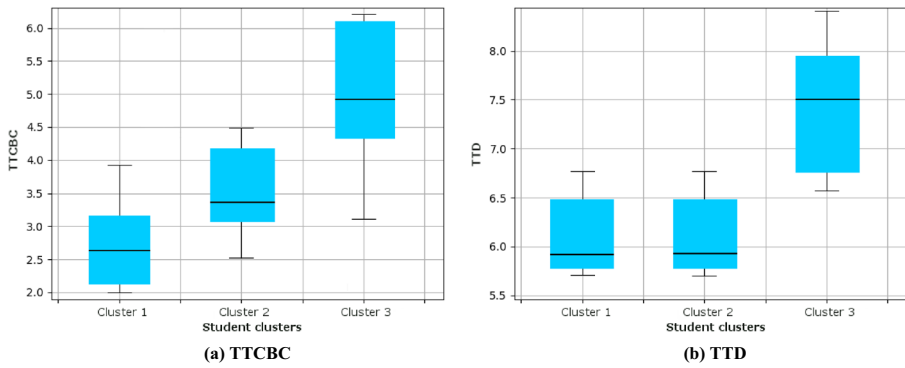


Fig. 7 The variation of the TTD and TTCBC variables within the three student clusters. **a** TTCBC. **b** TTD

3.5 Learning pathway modeling

3.5.1 Population-level model: Integrating learning pathways simulation with machine learning insights

We now demonstrate how the knowledge from the machine learning experiments are utilized within modeling the education process. The data-driven knowledge was used to reflect on the structure and behavior of the learning pathway model in different respects as follows.

Initially, a system dynamics model was built representing the three clusters of students in order to assess the dynamic complexity of the behavior that arises over time. In particular, the model was disaggregated into three stocks corresponding to the clusters of students. Furthermore, the auxiliary variables were decided based on the cluster analysis conducted in the previous section. For instance, the first and second clusters were set to undergo the same time to degree, while the third cluster was assigned a different time.

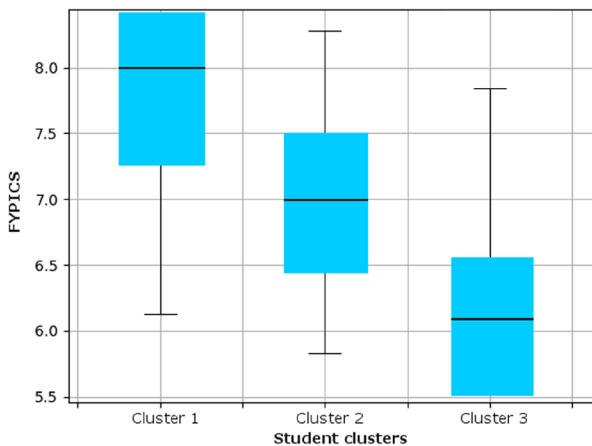


Fig. 8 The variation of the FYPICS variable within the three student clusters

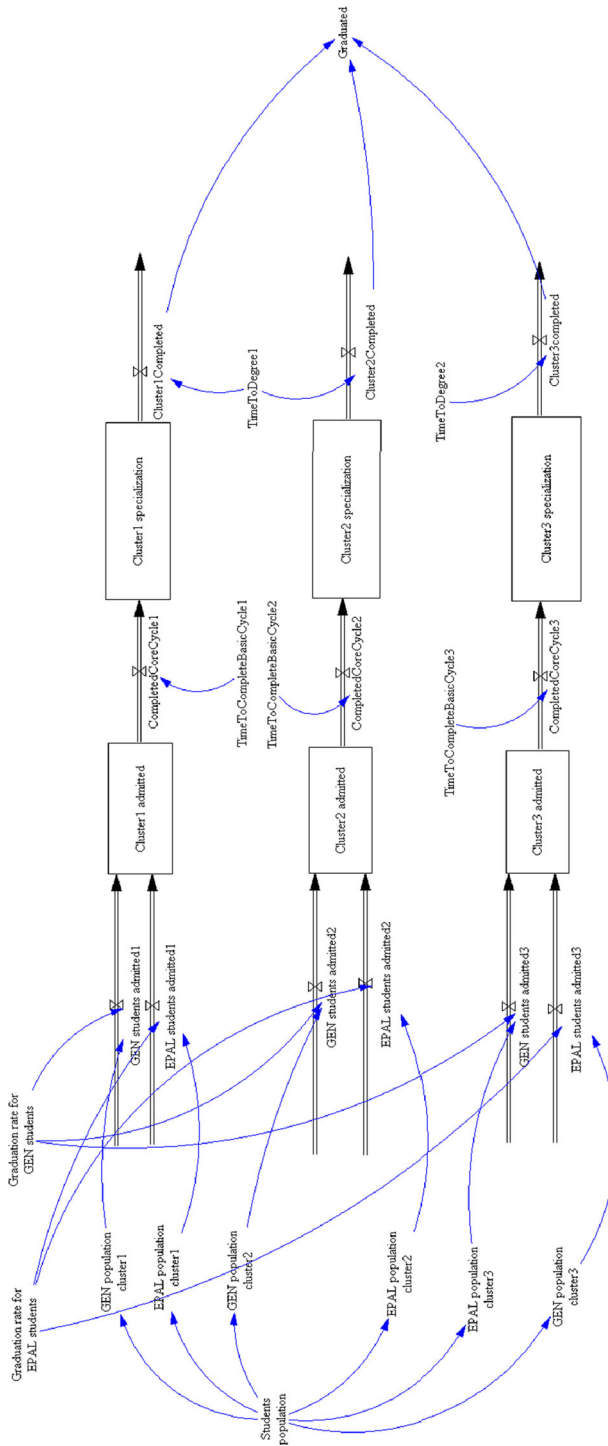


Fig. 9 Cluster-based system dynamics model

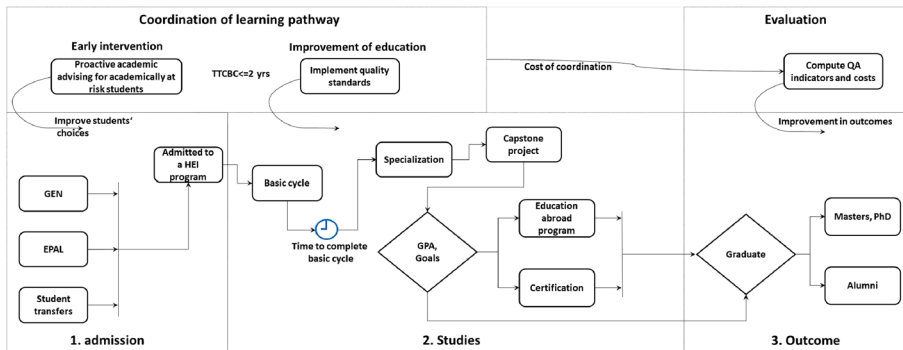


Fig. 10 The student-level model

Similarly, the inflow of students was structured based on the FYPICS variation within clusters. Both of the first and second student clusters were modeled to include students with very good and excellent performance in first year core courses, while the third cluster was associated with students with lower FYPICS average. This reflected the students within the clusters. In general, the system dynamics model can be used to provide projections of student graduation with a focus on different education-related factors and outcomes. Figure 9 illustrates the cluster-based system dynamics model.

3.5.2 Modeling learning coordination at student-level

Subsequently, we model a finer-grained perspective of the student's learning pathway while considering students as entities, rather than aggregate populations so as to achieve effective personalization. In this manner, each student could be treated individually in terms of various education-related factors. The discrete-event modeling is more suitable in this regard. The model can also be used to produce a realistic sequence of events in time within the learning pathway as sketched in Figure 10.

Along with the student's learning pathway, points of education coordination are simulated where a change of the process can take effect. Specifically, Fig. 10 highlights two components as (coordination of learning pathway) and (Evaluation). The coordination component initially aims at academic intervention including identifying population at dropout risk by implementing proactive academic advising or personality assessment tests (RIASEC) for improving student choices. For student cases, the coordination also focuses on the implementation of education standards towards reducing the TTCBC to be within 2 years from admission. The cost of coordination and the improvement in outcomes are reported to the Evaluation component where the EFQM quality assurance indicators and costs are eventually computed.

4 EDUC8 realization methodology

The proposed approach of the present research work requires a specific realization methodology to make the EDUC8 framework (approach and integrated IT system) operational during the execution of an education plan. The following paragraphs

describe all the actions to be performed in each of the four (4) stages of the realization methodology of the EDUC8 framework:

1. **Select and review of field of study.** EDUC8 approach recognizes that a HEI supports different fields of study and contains faculty mentors, academic advisors, managerial and administrative personnel with various specialties and expertise, thus comprising a multidisciplinary team responsible for the learning pathway maintenance. Initially, the field of study is selected for which the entire procedure will be performed. Choosing the field of study is the primary step, as it will also guide the choice of the corresponding educational options. The multidisciplinary team then explores contemporary educational trends and practices and selects the optimal available educational options. At the same time, a thorough review of academic advising recommendations is performed.
2. **Sub-Workflow Design.** The next step of the EDUC8 framework concerns the design and modeling of the sub-workflows that will be available for selection during the execution of the education plan. Initially, the academic advising guidelines for the specific field of study are represented as a decision tree, since it comprises a simple way to understand and interpret. Any decision node of a tree is associated with a one-to-many relationship with a rule, thus alleviating the well-known combinatorial rule explosion problem. Subsequently, the decision tree is degraded to distinct sub-workflows, which do not include any decision, but only one or more academic actions to be forwarded to the execution engine. The set of sub-workflows that are identified are modeled and stored in a repository, which will contain all the available learning sub-pathways so as to select the most appropriate for each student.
3. **Knowledge representation.** During the third step, the initial version of the semantic model, which comprises the learning pathway, learner, organizational and quality assurance parts is created and imported into the system. After the import, the domain experts (academic staff) interact with EDUC8 platform so as to represent and store their knowledge over the selected field of study. In order to facilitate the continuous maintenance of the rule base in an integrated way by the domain knowledge experts, a graphical rule generator interface was implemented. The rules implemented with the specific tool are stored as an SWRL repository.
4. **Learning pathway execution.** The last stage of the realization methodology is handled by the implemented EDUC8 software platform and concerns the recommendation and execution of the education plan as well as the exploitation of the potential insights generated by the machine learning components. First, an initial student orientation and definition of student parameters is performed, which allows the system to propose an appropriate learning pathway. Once a learning pathway is instantiated, EDUC8 environment reasons over the stored knowledge at every decision point taking into account both the personal parameters of the learners and their prior and current learning steps that have been executed. At the same time, the system executes the semantic rules that represent the organizational knowledge in order to propose the optimal management and allocation of the corresponding tangible and intangible resources of the HEI. These last steps are repeated until the degree attainment. After each execution cycle, machine learning techniques together with modeling and simulation are utilized to conceptualize the

learning pathway organization and system's future behavior. For this purpose, EDUC8 builds up the knowledge base over time as an incremental learning system so as to benefit future learning and decision-making process.

5 Contribution of this research work

In this research, we have presented a framework that utilizes simulation modeling along with machine learning to enhance the design of learning pathways and then provided a case study to demonstrate the EDUC8 approach for developing tertiary policies that help optimize the quality and cost of education. The major works of this research are summarized as follows:

1. Developing an integrated framework that offers an ICT based solution regarding the standardization as well as the dynamic adjustment of the Higher Education procedures of learning pathways tightly integrated with machine learning and modeling and simulation. The provision of as much as possible standardized higher education services and academic guidance towards personalized learning for a specific group of students with a predictable academic course can provide tremendous benefits from the economies of scale that can be achieved both in tangible and intangible resources of a HEI. Moreover, by identifying “at risk” students early, HEIs can improve retention rates, which affects almost all segments of HEI metrics: reputation, financials, ranking. Thus, the goal of optimizing the quality of the services provided in combination with the optimal cost is achieved.
2. A unique contribution that this study has made is to model the learning pathway based on the results of an unsupervised clustering mechanism using the K-Means algorithm to deal with real student data. The clustering-aided approach iteratively revises the patterns of educational processes.
3. Using modeling and simulation during the EDUC8 design mode makes HEI managers easier to know the educational process transitions and to accumulate knowledge from the educational processes. At population or student level, simulation modeling is used to predict the impact of changes in student flow, to assess resource needs such as staffing or physical capacity, and to investigate the relationships among the different model variables.

The extended EDUC8 framework is currently in the process of pilot execution in a HEI so as to measure several quantitative and qualitative factors and conclude about the following issues:

1. the educational support quality offered by the EDUC8 framework,
2. the results produced by the proposed K-means algorithm compared to other machine learning algorithms,
3. the performance of the EDUC8 software in the case of several learning pathways execution,
4. the EDUC8 system comprehension, etc.

6 Conclusions

HEIs are in transition from a traditional university model to a demand driven organization. The key challenges facing HEIs in future years are perhaps more organizational and logistical than learning and scientific advances. The Higher Education arena can substantially be transformed by advances in our capabilities to understand and model Higher Education systems. Modeling of learning pathways presents as a pragmatic method for understanding the complex dynamics underlying higher education systems towards devising policies for education improvement. It is hoped that the framework presented in this paper to contribute in that direction within the context of developing learning pathways with the aim of optimizing educational services in conjunction with the minimization of costs for tertiary institutions.

Further, the paper presented an extended framework for integrating the learning pathways scheme with machine learning. The study utilized machine learning as an assistive artifact within the process of conceptual modeling. Applying a machine learning enhanced framework is a useful tool for coordination of learning pathways since it can provide HEI decision makers and researchers with alternatives to understand student parameters, learning trajectories and numbers in the age of “big data.” Through a case study, it was demonstrated how insights learned from data clustering could be used effectively to conceptualize the system’s structure or behavior. In a broader context (more programs, departments, students or even HEIs), we conceive that machine learning can play an important role in tandem with modeling and simulation of learning pathways to help address more complex set of analytical questions.

References

- Bailey, T., Jaggars, S. S., & Jenkins, D. (2015). What We Know About Guided Pathways, (March). Retrieved from <http://academiccommons.columbia.edu/catalog/ac:185800>
- Booch, G. (1999). *The unified modeling language user guide by object technology Series*. Addison-Wesley.
- Buizza, C., Ghilardi, A., Olivetti, E., Costa, A., Buizza, C., Ghilardi, A., ... Costa, A. (2019). Drop out from a university counselling service : A quantitative and qualitative study. *British Journal of Guidance & Counselling*, 0(0), 1–13. <https://doi.org/10.1080/03069885.2019.1566513>.
- Casalone, G., & Orientale, P. (2011). *Does time-to-degree matter ?* The effect of delayed graduation on employment: AlmaLaurea Inter-University Consortium.
- Fernando, D., Cabral, R. B., Souza, C. P., & Gam, C. (2010). Maturity model for process of academic management. In *2010 International Conference on Information Society*.
- Ghiasi, S., Srivastava, A., Yang, X., & Sarrafzadeh, M. (2002). Optimal energy aware clustering in sensor networks. *Sensors*, 2(7), 258–269. <https://doi.org/10.3390/s20700258>.
- Grunschel, C., Patrzek, J., & Fries, S. (2013). Exploring reasons and consequences of academic procrastination: an interview study. In *Exploring reasons and consequences of academic procrastination : An interview study*, 841–861. <https://doi.org/10.1007/s10212-012-0143-4>.
- Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities and work environments. *Psychological Assessment Resources*.
- HQA (2017). Έκθεση ποιότητας της ανώτατης εκπαίδευσης 2017. *HQA* (Vol. 1).
- Iatrellis, O., Kameas, A., & Fitsilis, P. (2019a). A novel integrated approach to the execution of personalized and self-evolving learning pathways. *Educ Inf Technol* (2019) 24:781–803, 24(ISSN 1360-2357). <https://doi.org/10.1007/s10639-018-9802-7>.
- Iatrellis, O., Kameas, A., & Fitsilis, P. (2019b). EDUC8 ontology: Semantic modeling of multi-facet learning pathways. *Education and Information Technologies*, 2019, 2371–2390. <https://doi.org/10.1007/s10639-019-09877-4>.

- Janssen, J., Berlanga, A. J., & Koper, R. (2011). Evaluation of the learning path specification. *Educational Technology & Society*, 14, 218–230.
- Jenkins, D., & Cho, S.-W. (2013). Get with the program ... and finish it: Building guided pathways to accelerate student completion. *New Directions for Community Colleges*, 2013(164), 27–35. <https://doi.org/10.1002/cc.20078>.
- Kappe, R., & Van Der Flier, H. (2012). Predicting academic success in higher education: What's more important than being smart? *European Journal of Psychology of Education*, 27(4), 605–619. <https://doi.org/10.1007/s10212-011-0099-9>.
- Kučak, D., Juričić, V., & Đambić, G. (2018). Machine learning in education - a survey of current research trends. *Annals of DAAAM and Proceedings of the International DAAAM Symposium*, 29(1), 0406–0410. <https://doi.org/10.2507/29th.daaam.proceedings.059>.
- Kumar, P. M. S. (2016). Student performance and learning outcomes in higher education, I(i), 674–684.
- MacQueen, J. and others. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281–297. Retrieved from http://books.google.de/books?hl=de&lr=&id=IC4Ku_7dBFUC&oi=fnd&pg=PA281&dq=MacQueen+some+methods+for+classification&ots=nNTcK1IdoQ&sig=fHzdVcbvMYJ-ITNHu1HncmOFokM#v=onepage&q=MacQueen+some+methods+for+classification&f=false.
- McCarthy, W. (1982). The REA accounting model: A generalized framework for accounting systems in a shared data environment. *Accounting Review*, 554–578.
- McKenzie, K., & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first year Australian university students. *Higher Education Research and Development*, 20(1), 21–33. <https://doi.org/10.1080/07924360120043621>.
- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074–2102. <https://doi.org/10.1002/sim.8086>.
- Oliveira, M. N. (2016). How DOES the ASAP model align with guided pathways implementation in community COLLEGES?, 1–3.
- Pavlenko, V., Prokhorov, A., Kuzminska, O., & Mazorchuk, M. (2017). Competence approach to modeling and control of students' learning pathways in the cloud service. *CEUR Workshop Proceedings*, 1844, 257–264.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572. <https://doi.org/10.1080/14786440109462720>.
- Prebble, T., Hargraves, H., Leach, L., Naidoo, K., Suddaby, G., & Zepke, N. (2004). *Impact of student support services and academic development Programmes on student outcomes in undergraduate tertiary study : A synthesis of the research*. Wellington: Ministry of Education.
- Quinn, A., Lemay, G., Larsen, P., & Johnson, D. M. (2009). Service quality in higher education. *Total Quality Management & Business Excellence*, 20(2), 139–152. <https://doi.org/10.1080/14783360802622805>.
- Schouwenburg, H., Lay, C., Pychyl, T., & Ferrari, J. (2006). Counseling the procrastinator in academic settings.
- Scott-Clayton, J. (2011). The shapeless river: Does a lack of structure inhibit students' Progress at community Colleges? *A working paper in the CCRC assessment of evidence series*, (January).
- Song, J., Li, F., Takemoto, K., Haffari, G., Akutsu, T., Chou, K. C., & Webb, G. I. (2018). PREvaIL, an integrative approach for inferring catalytic residues using sequence, structural, and network features in a machine-learning framework. *Journal of Theoretical Biology*, 443, 125–137. <https://doi.org/10.1016/j.jtbi.2018.01.023>.
- The EFQM Excellence Model. (1988). Retrieved January 30, 2017, from <http://www.efqm.org/the-efqm-excellence-model>
- U-Multirank | Universities compared. Your way. (2011). Retrieved August 14, 2018, from <http://www.umultirank.org>
- Wong, L. H., & Looi, C. K. (2009). Adaptable learning pathway generation with ant Colony optimization. *Educational Technology and Society*, 12(3), 309–326.
- Wu, B. H. (2011). Let's enforce a simple visualization rule in software architecture. In *2011 international conference on information science and technology. ICIST 2011* (pp. 427–433). <https://doi.org/10.1109/ICIST.2011.5765283>.