This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

Automatic tutoring system to support cross-disciplinary training in Big Data

Xavier Solé-Beteta · Joan Navarro · David Vernet · Agustín Zaballos ·
Ricardo Torres-Kompen · David Fonseca · Alan Briones

X. Solé-Beteta
Grup de Recerca en Internet Technologies & Storage
La Salle — Universitat Ramon Llull
Barcelona, Spain E-mail: xsole@salleurl.edu

Abstract. During the last decade, Big Data has emerged as a powerful alter-native to address latent challenges in scalable data management. The ever-growing amount and rapid evolution of tools, techniques, and technologies associated to Big Data require a broad skill set and deep knowledge of several domains—ranging from engineering to business, including computer science, networking or analytics among others—, which complicates the conception and deployment of academic programs and methodologies able to effectively train students in this discipline. The purpose of this paper is to propose a learning and teaching framework committed to train masters' students in Big Data by conceiving an intelligent tutoring system aimed to (1) automatically tracking students' progress, (2) effectively exploiting the diversity of their backgrounds, and (3) assisting the teaching staff on the course operation. Obtained results endorse the feasibility of this proposal and encourage practitioners to use this approach in other domains.

Keywords: Big Data training · Intelligent tutoring system · Master as a Service · Virtual Learning Environment

1 Introduction

The digitalization of modern societies together with the continuous improve-ment of Information and Communication Technologies (ICTs) have triggered the emergence of a wide plethora of new disciplines (e.g., Internet of Things, Social Networks, Industry 4.0, Cloud Computing, Big Data). Traditional un-dergraduate and graduate programs often struggle to keep up with the pace [39] of such fast evolution due to (1) the immaturity of these topics [33] (i.e., the technology that it is being taught today might become obsolete tomorrow),(2) the lack of consolidated technical experts available to train new students (i.e., it is hard to master a specific topic in this evolving context), and (3) the aversion of universities and schools to abruptly modify existing course syllabuses and curricula [7] that have remained stable for years. This situation might lead to a situation gap in which universities and teaching centers fail to effectively train students able to address modern challenges in the real-world. In contrast to traditional disciplines where a deep knowledge in a single vertical topic was enough to keep a consistent lifelong career, modern specialities typically require broad knowledge in several areas that are not necessarily related.

A good example of this situation is Big Data [51]. In the last ten years, Big Data has emerged as a modern discipline in which companies demand experts to address their latent problems related to data storage, processing, visualization, and analysis [36,30]. For instance, Big Data consultants are required to know about: Physical and/or virtual (i.e., cloud computing) datacenters—a topic which is mostly related to network engineering and telematics—, Massive data storage and processing technologies—a topic which is mostly related to computational studies—, Business intelligence—a topic which is is mostly related to business and management studies—, and Data analytics—a topic which is mostly related to mathematics and numeric studies.

Therefore, training Big Data professionals and providing them with this horizontal (from a technical point of view) skill set while also providing them with the soft skills associated to current society (e.g., leadership, communication, and teamwork) is challenging [51]—and unfeasible using classic educational strategies. On the one hand, the large number of different profiles and backgrounds that might, and usually, apply [51] for such studies (e.g., net-work engineering, computational studies, business and management, PhDs, and mathematics, numeric studies among others) would make it difficult to go deep into the Big Data contents (i.e., most students might struggle to follow those classes that are less related to their background due to their lack of knowledge). On the other hand, gaining such a horizontal knowledge set, even using modern active learning strategies (e.g., flipped classroom [9], project based learning [8], peer instruction [20]) might take a considerable amount of time, which is typically not available in a graduate program.

This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

Educational innovation [54] has positioned itself as an efficient, effective, and successful way to address teaching and learning challenges in modern edu-cation. In fact, there is a rising interest in universities to explore and conceive new strategies to effectively teach and shape this new type of horizontal pro-fessionals that are able to master loosely connected disciplines, contribute to their evolution, interact with radically different profiles, and, overall, add value to the industry.

Combining educational innovation with the advantages of having an inter-disciplinary set of students, led the authors of this work to propose the Master as a Service (MaaS) approach to train students in Big Data [44]. The MaaS paradigm aims to blend students from different graduate programs that are in some way related to the area of Big Data (e.g., eHealth, Digital Trans-formation, etc.) in order to solve real-world Big Data challenges while being mentored by industry professionals and, thus, provide them with a high quality yet continuously updated Big Data training. It is worth noting that this initiative is materialized in the context of the master's thesis project. In this way, every graduate program trains its enrolled students in a vertical way and, later during the master's thesis project development, this knowledge is horizontally spread among students from other programs by using educational innovation techniques.

As detailed in [44], the MaaS strategy strongly relies on multidisciplinary collaborative work (i.e., combining the best capabilities of each individual in order to exploit their skill set and reach a common goal), which is a very com-mon requirement in existing Big Data jobs. Additionally, collaborative work might be also seen as a powerful tool to stimulate some aptitudes related to socialization [4,10] such as conflict resolution, group motivation, role definition, argumentation, and feedback discussion [57].

Although the MaaS approach has been successfully implemented to train students coming from different master's degrees and undergraduate back-grounds (ranging from management studies to computer engineering, including architects, social and physical sciences) to address latent and future challenges in Big Data and High-Performance Computing technologies, its large scale implementation (i.e., exposing hundreds of students from dozens of master's programs to real-world challenges) poses the following issues:

- It is very time consuming to analyze every student profile in order to come up with an optimal working group configuration that balances the skill set of all the members [51].
- It is hard to accurately monitor the performance of every group (and its individual members) when the number of groups/students raises.
- Accordingly, it is unfeasible to deliver accurate feedback and guidance to students in order to enrich their learning experience.
- It is unfeasible to track students interactions on collaborative work as most of their interactions are made online instead of in-class.

Therefore, the purpose of this paper is to extend the work presented in the track of "Supercomputing education: Thinking in parallel" of the 7th Inter-national Conference on Technological Ecosystems for Enhancing Multiculturality [44], and propose an Intelligent Tutoring System (ITS) to support the MaaS deployment and address the aforementioned issues. More specifically, the proposed ITS uses machine learning and data mining techniques to (1) automatically track students' progress from their interaction with the Learning Management System (LMS), (2) effectively exploit the diversity of their backgrounds by automatically categorizing their profiles, and (3) assist the teaching staff on the course operation by providing them with precise and actionable insights. The fact that this ITS is conceived to be integrated inside the LMS used by all the stakeholders involved in the MaaS ecosystem enables the system itself to quantify the students activities associated to collaborative work.

Preliminary experimental evaluations have allowed to considerably reduce the amount of time spent on the MaaS management duties and considerably improve the overall students satisfaction.

The remainder of this paper is organized as follows: Section 2 reviews the related work on educational innovation and motivates the need of multidisciplinary strategies to address complex problems. Section 3 reviews the Master as a Service paradigm. Section 4 proposes a methodological framework to include an ITS in the LMS. Section 5 presents Sagittarius: the proposed ITS to assist on the MaaS development. Section 6 describes how Sagittarius has been used to train students in Big Data using the MaaS approach. Section 7 outlines the main conclusions of this work and discusses the obtained results. Finally, Section 8 draws some future work directions.

This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

2 Related work

The massive structural changes brought by the adaptation of existing programs to the so-called European Higher Education Area (EHEA) have reshaped university education in several dimensions [15]. According to this new perspective, the main goal of learning is to acquire knowledge while developing a series of competencies based on the academic profiles and professional outputs of each study [27]. In this regard, there are several works [6,21,14,13,29,2,16,18,31,5] that investigate the relationship between the acquired competences—usually specific—and the employability and/or professional skills related to their associated programs.

These works are typically segregated by university/location [6,21,14,13] or by educational/professional field [29,2,16,18,31]. As it can be seen, the evolution of the labor market, especially in the field of Big Data and Supercomputing, has been driven by the dynamics of technical change in the so-called knowledge society. As a result, organizations, working methodologies, technologies and, therefore, training requirements are continuously changing [21]. The increasing number of companies requiring broad and flexible profiles—in terms of leadership, communication skills, teamwork, organization, and ability to observe, learn, create, adapt, apply, identify problems, changes, opportunities, etc.—has established a multifaceted, yet strong, link between university training and industry demands [27,5].

Educational innovation can be described as the process of changing the teaching or learning activities in order to increase students' performance [54]. Additionally, the process of educational innovation should meet certain constraints [37] in terms of effectiveness, efficiency, sustainability over time, and replicability over time (i.e., it should produce transferable outcomes beyond the particular context in which it arose).

When including a new methodology in a specific teaching environment——especially in the case of multidisciplinary students——, the following four recommendations for avoiding student rejection must be considered [42]:

1. Promote professor-student relationships for a more effective and accurate feedback process.
2. Foster active learning among students, which is made possible by applying collaborative techniques.
3. Enhance task development by using heterogeneous learning methods, meeting high expectations.
4. Apply teaching/learning methods based on teaching innovation and new ICTs. Indeed, including ICTs [41] in innovative teaching environments, enables practitioners to (1) boost personal production (i.e., applications that allow both the professors and students to carry out tasks faster and more efficiently), (2) reduce contents staleness (i.e., using tools that allow a rapid and efficient modification of contents, such as video or multimedia resources, without changing the basic teaching method), and (3) shift the traditional teaching paradigm (i.e., the teacher reconfigures the teaching and learning activities to utilize the newly incorporated technologies).

There are several examples in the literature of educational methodologies that show the implementation of the first two recommendations, but examples that address the third one are much less common [26]. In fact, when addressing a highly multidisciplinary group of students, it is necessary to create environments that (1) are best fitted to the students' needs, (2) are open to the inclusion of all kinds of innovative technologies, methodologies and tools that they must use to learn, and (3) have the capacity to adapt to the students' profile from a technologically, social, and physical perspective. Furthermore, in these contexts, a critical change of focus must be considered: the tools will no longer be so much aimed at the specific course or institution, but at the student's own specific preferences. Therefore, it is necessary to customize the learning process [3], to give the student spaces and tools that he or she can configure according to his/her specific needs, not only within the institution, but at any given moment in his/her daily life [26].

Overall, training students to face and address complex problems or challenges in multidisciplinary teams using collaborative strategies encompasses a wide skill set that turns out to be very appealing for companies and industry professionals, which results of broad interest in the field of emerging and ever-evolving technologies such as Big Data and Supercomputing.

3 The Master as a Service (MaaS) approach

The MaaS approach was specifically conceived to train students to master technologies that do not yet exist and solve challenges that are still to be defined. That is, prepare them to be valuable professionals in an ever-changing environment in which new disciplines and their associated technologies rapidly appear (and disappear). It was found out that novel active learning strategies—that with no doubt give students a solid and valuable background—that were being used in current graduate programs, were not enough to provide them with the required skills to position them in the multidisciplinary environments that new disciplines currently demand. In this regard, we decided to design a strategy to expose them to a real multidisciplinary situation while they were enrolled in their graduate program, and, at the same time, train them in a new field such as Big Data.

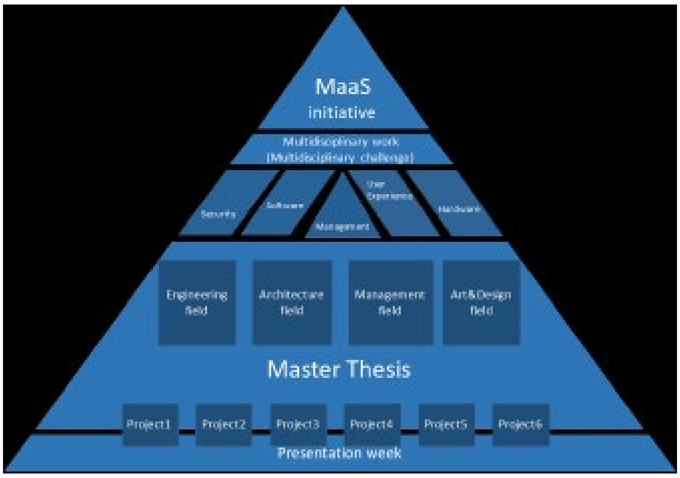

Fig. 1 Generic MaaS layout

As depicted in Fig. 1, the MaaS initiative consists of proposing a real-world multidisciplinary challenge—which involves both experts in education and experts in industry—that covers several vertical disciplines (e.g., cyber security, software engineering, management, user experience, and hardware design) to students from different graduate programs of different fields (e.g., Engineering, Architecture, Management, Art&Design). In this context, they have to solve the challenge in groups—typically not assigning more than one member from each graduate program involved in the MaaS initiative. To avoid rescheduling all the graduate programs that are involved in this initiative and ease the logistics associated to MaaS, the aforementioned challenge is proposed in the context of the Master's Thesis project.

The design of this challenge is twofold. On the one hand, the education experts in universities must design learning activities to foster the knowledge exchange between students from different programs. Note that these activities o beyond the peer instruction [20] since a rapid knowledge transfer and sharing is required to meet the challenge deadline. Therefore, reduced-scope mini challenges (i.e., mini challenge based learning) need to be set up in order to (1) ease the interactions between students that do not know them each other, (2) guide them on the key concepts and path that will enable them to gain the required new knowledge to solve the challenge, and (3) provide them with incremental deliverables that help them to quantify their progress.

On the other hand, industry experts must design feasible challenges that cover key topics from different graduate programs. While the former typically requires some feedback from the teaching staff (they have first-hand information and data regarding students' capabilities and available time), the latter is often straightforward when the challenge is related to emerging disciplines such as Internet of Things, Industry 4.0 or Big Data.

Finally, each group of students presents their solution proposal to the challenge in front of a multidisciplinary panel composed by a member of the company that proposed the challenge, a member from another company, and two members from the academic staff of the university hosting the MaaS initiative. In this presentation, all students are asked to accurately answer questions from all the disciplines that their challenge encompasses. That is, questions regarding topics from their graduate program and, also, questions regarding topics covered by the challenge that they have already solved but might be close to graduate programs that they have not been enrolled in. Additionally, the panel committee is asked not to consider the different backgrounds of each student in an attempt to avoid any bias in the questions part.

Although this methodology has been found to be very effective when training professionals in multidisciplinary domains such as Big Data [44], we have found that it is very difficult to scale the MaaS paradigm to a large number of students because:

1. It is difficult to come up with an optimal and balanced configuration of working groups.
2. The amount of data (deliverables, practical assignments, grading) gener-ated by each working group is overwhelming to process manually.
3. The number of teaching members required to develop the MaaS grows faster than the number of students enrolled in this initiative.
4. The complexity associated to manage a growing number of heterogeneous graduate programs grows exponentially.
5. It is very difficult to find a tutor with enough heterogeneous knowledge to effectively mentor a group of students coming from a, potentially large, number of different programs.

Therefore, we propose to take advantage of Virtual Learning Environments (VLEs) and Information and Communication Technologies (ICTs) to develop an automatic intelligent tutoring system able to assist the teaching staff on the implementation of the MaaS paradigm over a large number of students. The following section describes the methodological framework to include an automatic Intelligent Tutoring System in a Virtual Learning Environment.

4 Methodological framework for Virtual Learning Environments

The application of technology to the educational world has fostered the creation of new learning environments that have allowed the student to be positioned not only as a simple spectator of the educational process, but also as a player within the network that supports the experience. Technology evolution has made possible the redesign and refocusing of traditional training programs, resulting in models and tools that are strongly grounded in skill acquisition and the interactions between individuals.

Recently, with the overwhelming growth of the Internet, technology has taken an active role in several facets of our culture; from electronic press and information surveys, to personalized advertisements, including a permanent virtual connection with the rest of the world [28]. This situation entails a new way to conceive education, which encompasses the cognitive processes associated with learning and a set of cutting-edge technologies that assist the knowledge acquisition [38,40]. Therefore, universities and the educative community—aware of these new trends—have adjusted their methodologies and syllabus to fit this new paradigm [24,59,35]. Hence, the continuous evolution of both technology and society has driven to a new educative paradigm referred to as virtual education [23,45,50]. Web-based education and Virtual Learning Environments (VLEs) open up a whole range of possibilities when it comes to pedagogical strategies that have the potential to improve the learning process. It seems clear, then, that VLEs have great potential for learning process development, since these systems use technologies that enable both personalization of learning and socialization in learning.

In fact, education using VLEs shares several similarities with traditional in-class lectures in terms of learning goals and syllabus topics [12]. However, it opens new challenges in terms of student tracking since there is no physical link between all the members in a classroom: students and lecturers [34,22]. This lack of connection prevents some advances (e.g., heterogeneous collaborative work, which is the key of the MaaS approach) conducted in the face-to-face educative field from being applied to this new domain.

Accordingly, exploiting the collaborative work approach effectively—under-stood as a key strategy to enhance the learning experience [43]—using VLEs is still a hot research topic [53]. Indeed, when deploying the MaaS using on-line learning environments one may expect to find the following issues: (1) significantly distinct ages, experiences, motivations, and availability are found in each student, which may hamper initial interactions [32], (2) students may leave the course at will, which may break group dynamics [47], (3) virtual classrooms can potentially host a vast number of guests (also referred to as Massive Open Online Courses or MOOCs for short) [24,49], which complicates the act of keeping an accurate track of students' progress [17], and (4) the teaching staff is unable to physically and reliably perceive students' re-actions and feedback [25], which limits effective interactions between teachers and learners. So far, very few approaches have been proposed to successfully address these issues, which leads to a noteworthy quality gap between training using VLEs and traditional face-to-face courses.

This section sets out a learning scenario that seeks to combine the best features of VLEs with those of Intelligent Tutor Systems (ITS) to support the deployment of the MaaS approach for a large volume of students. The result is a model proposal that we have coined to as TICVA (Intelligent Tutoring of Virtual Learning Communities, in Catalan) that seeks to push the e-learning process towards the next generation of e-learning (e-learning 4.0). The next subsection shows the conceptualization and an overview of the

This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

TICVA system [57,55,58] based on the reference model of learning analytics and the architecture of the system developed from the structure of a classical ITS.

4.1 Conceptualization of the TICVA model

The transition from Computer-Assisted Instruction (CAI) systems to ITS led to a reformulation of the teaching methodology [58]. While CAI systems, as the name implies, were based on knowledge instruction, the integration of Artificial Intelligence led to an eminently constructive methodology, which is what ITS help implement.

ITS aims to guide the learner through a domain of knowledge by adapting activities to their specific needs. That is, an ITS has the key feature of helping personalize learning. However, for this very reason, ITSs are conceived as individual use tools where the ability to work in a group is non-existent. They do allow both the student and the teacher to track the learning process, but they are not intended to help in evaluating the performance of a group or virtual classroom.

Today, online learning is unthinkable without the concept of networking and group work. It is clear that personalization of learning is a distinguishing feature from generalist models of the past, but the conception of learning as a social process cannot be underestimated. From these premises, a system is proposed in which these fundamental factors converge and allow to create an environment in which it is possible to produce a process of personalized learning which is, at the same time, social.

As it has been said, the TICVA system seeks to combine the best of VLEs with the best of ITS: the first ones provide the capacity that they offer to carry out the learning process within a community, as a group experience; the latter offer the intelligent ability to predict and adapt to the needs of the individual. This results in a capable system, not only able to guide an individual in their learning process, but also allowing for this happen within a community.

Previous works [55,56] show how a learning system along these lines would look and operate—although none of them consider the heterogeneous and multidisciplinary topics covered in the Big Data field. The basic structure and the conceptualization of the TICVA system is shown in Fig. 2.

Modeling the TICVA system can be difficult if only the pedagogical dimension is taken into account, because if there is one relevant fact about VLEs is the immense amount of educational information that is generated within them. In essence the data flowing through a VLE (interactions, results, users, ...) represents multiple learning processes in an abstract, even confusing way. But if there was a way to somehow capture and process that information, then the extracted knowledge would provide a clear view of what is going on in the VLE.

As the TICVA system aims to emulate the behavior of an ITS but for more than one individual, it is required that it is also able to deal with the Big Learning Data that is generated in a VLE and, thus, may be able to provide a clear idea of what is happening in the environment.

As shown in Fig. 3 the overall structure of the TICVA architecture seeks to build on the traditional modules of an ITS. Unlike the traditional model, a new entity corresponding to the group module appears, which includes information corresponding to the learning process of a group of students who interact and collaborate with each other; this way, the classic student module is included within this higher module.

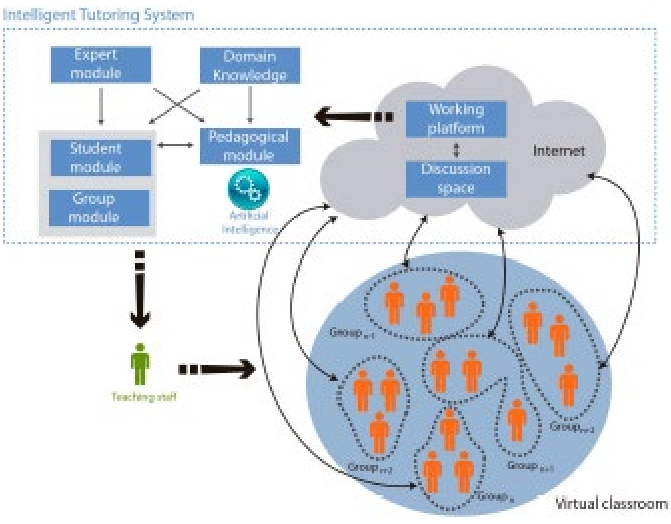

Fig. 2 Conceptualitzation of the TICVA architecture.

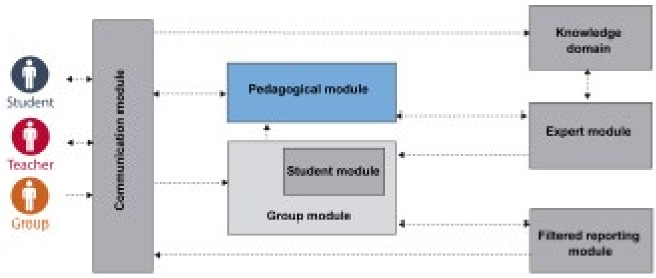

Fig. 3 Basic architecture of the TICVA system.

Fig. 4 shows an overview of the system's architecture, with the modules and the interconnections between them.

– Group module. It manages all the learning materials and activities associated to an individual student and/or group.

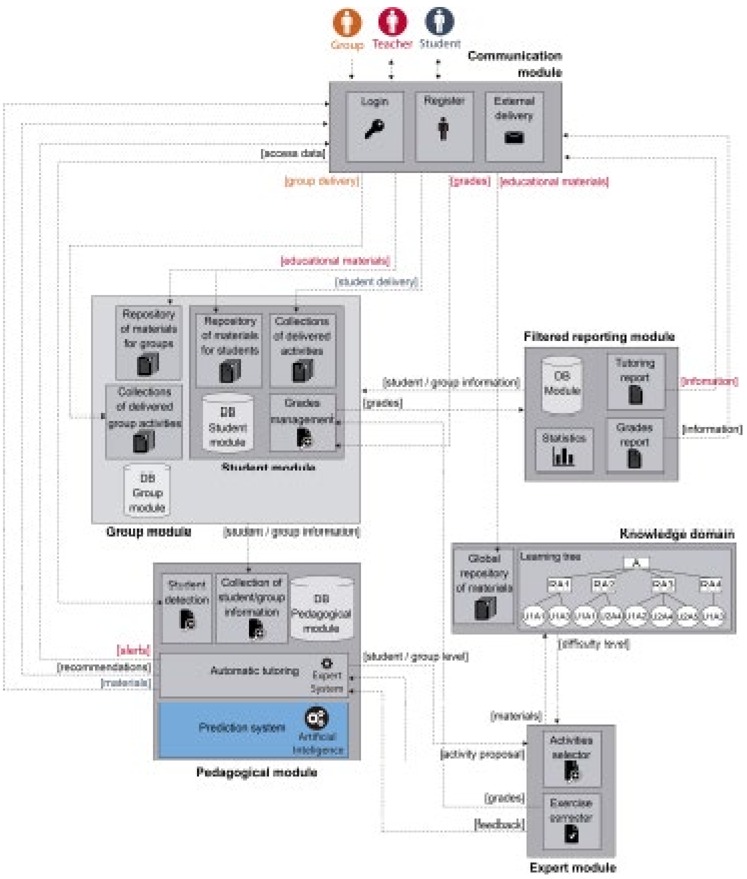

Fig. 4 Detailed architecture of the designed TICVA system.

– Communication module. It is the user interface. Students, teaching staff and groups login by means of this module. All the materials uploaded to the VLE are uploaded by means of this module.
– Filtered reporting module. It is committed to build integrated reports and compute statistics regarding the students/groups performance.
– Knowledge domain module. It stores all the learning materials associated to the (Big Data) course. The materials are organize in a hierarchical way (also referred to as Learning Tree) where Subjects derive on one ore more Learning Results (LRs), and each LR is linked to one ore more Topics (T) of the syllabus and their associated teaching Activities (A).

– Expert module. It is aimed to (1) grade the assignments that students submit to the system and (2) select the appropriate exercises to meet the learning requirements of the working group.

– Pedagogical module. It is responsible of tracking the student progress and performance along the course, thus, it must integrate the ITS to handle the big number of students enrolled in the MaaS.

The pedagogical module, which is the most relevant when implementing the MaaS, is explained in the following section.

4.2 Pedagogical module

The most important module of the system is the pedagogical module. Its function is to keep track of the students, evaluating at all times their evolution and progress and alerting, with recommendations or notifications, of potential problems, whether it be of a specific student or about the operation of a virtual group. It can also take care of the teaching strategy and how the materials are presented to the student in order to improve their learning process.

The pedagogical module is composed of several submodules, as depicted in Fig. 5 and are described below:

Student identification. This part of the module is responsible for identifying the student in the virtual class from the access data provided by the Communication module.

Information collection from student or (multidisciplinary) group.
Once the student has been identified, all the existing information of the student (or group, if applicable) must be retrieved in order to enable, in a later step, the prediction system with all the information recorded in the system. This information can range from the student's grades to their response time.

Prediction system. The prediction system is the intelligent part of the pedagogical module and it is responsible for assessing the status of a student or group in order to detect possible problems in the learning process. Note that the evaluation of groups (and individual students) strongly depends on the specific data mining algorithms selected to implement this TICVA module (see Section 5). For instance, a trivial way to evaluate the status of a group could be to sum the individual evaluations of all of its members. More sophisticated strategies could be used as well [19].

Automatic tutoring. Automatic recommendations for the student or alerts for the teaching staff are generated through this submodule. In short, this is a small expert system capable of interpreting the answer given by the prediction system and launching outgoing messages to the teaching staff. The final design of the specific interactions that take place can be key to the proper behavior of the tutoring system [11].

A complete representation of this module can be seen in Fig. 5. The pre-diction system, which will be the focus in the following section, is highlighted in a different color.

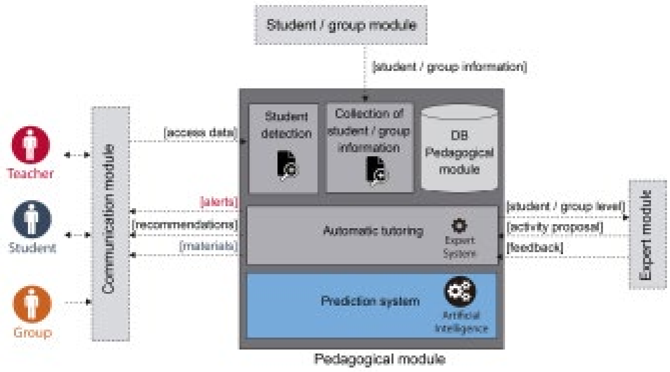

Fig. 5 Detailed architecture of the pedagogical module.

To sum up, TICVA proposes a methodological framework to manage large scale courses that cover multidisciplinary areas and require an automatic monitoring and tracking of students progress by means of an ITS. As shown above, the key element of the pedagogical module from TICVA is the prediction system. The next section details Sagittarius, our implementation proposal of this module for deploying MaaS to train students in Big Data using the TICVA framework.

5 Sagittarius

This section presents Sagittarius, a reliable ITS framework to support ac-tive learning methodologies based on collaborative work using LMSs (i.e., the MaaS). This tool assists the academic staff to carry key actions related to collaborative work by means of data mining techniques: personalized content delivery, individualized student tracking and monitoring, profitable continuous assessment, and outcome prediction.

Exploiting and identifying the specific cognitive needs of every student is fundamental for acquiring the expected learning goals and competences of any curriculum [12]. Therefore, personalized feedback turns out to be a key element to keep a permanent connection with students and, thus, boosting their motivation, reinforcing their knowledge, and extending their skills. How-ever, in those scenarios where there is a physical barrier between the teaching staff and students

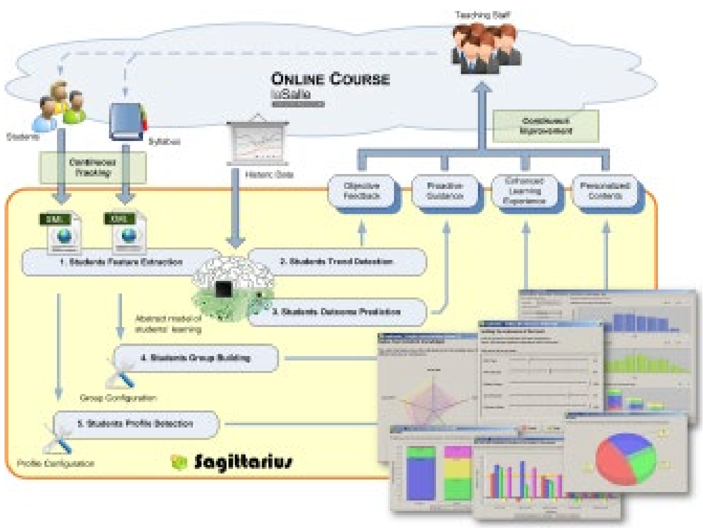

Fig. 6 System architecture of Sagittarius. Students' data is collected from the online course through XML files and delivered to Sagittarius' internal modules. Sagittarius computes a set of valuable metrics and parameters to assist the teaching staff [52].

[10,24,48] exploring these specific needs by means of traditional techniques becomes time costly and inefficient. To address this concern, we have developed Sagittarius, a software tool aimed to reduce the negative effects of homogeneously dealing with broadly heterogeneous students. Specifically, Sagittarius is committed to automatically determining these individual needs and assisting the teaching staff on managing the learning process in heterogeneous educational environments.

As shown in Fig. 6, Sagittarius is deployed on the teaching staff side in order to continuously track the students' performance and improve course contents (e.g., syllabus, exercises, assessment, feedback, [12]). In order to build an abstract model of the students' learning process (symbolized as a brain in Fig. 6), Sagittarius parses a set of XML files—which eases the integration of the Sagittarius system with other existing learning platforms—that contain a set of features that (1) describe students individually (e.g., historic grades, preferences, skill set, past work partners), (2) characterize students globally (e.g., grades obtained in past editions of the course), and (3) provide information concerning the correlation between syllabus and students feedback obtained in the past. With this information, a CSV file is automatically built and later used by the other modules of Sagittarius: trends detection, outcome prediction, group building, and student's profile detection. Note that this meta-knowledge model is continuously adjusted with the new inputs provided by students, which allows the teaching staff to react in advance to possible undesired outcomes and, thus, continuously improve the course materials. These modules are further elaborated in what follows.

This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

Student	CA_1	CA_2	PA_1	PA_2	TL_1	TL_2	FE
student38	6.20	6.50	1.70	1.80	45	52	7.00
student67	2.70	3.30	0.20	0.00	10	11	4.50
student47	6.80	5.20	3.20	4.50	63	80	6.50
student61	8.60	7.00	1.60	2.50	35	40	8.30
student54	8.50	9.50	3.60	4.80	30	35	9.50

$\cancel{if\ PA_1 < 2\ then\ CA_1 < 5.}$
$if\ TL_1 > 60\ then\ FE > 6.$
$if\ CA_1 > 7 \wedge PA_1 > 3\ then\ FE > 4.$

Fig. 7 Student's trend detection. Example of the rules induced by Sagittarius. Chronologically ordered input data collected from students (CA: Continuous Assessment, PA: Practical Assignment, TLi: Test of Lessoni, FE: Final Exam). Induced rules on the right. Note that the first rule is discarded because is non-causal.

5.1 Student's Trend Detection Module

Typically, a syllabus of a course includes exercises, checkpoints, online sessions, distinct types of assignments, and exams or tests. Apparently, the obtained score by a student on a given activity may affect the obtained score of the subsequent quizzes. Discovering these kind of correlations requires a deep expertise of the teaching staff and an accurate analysis of students performance, which becomes barely feasible given the huge amount of students attending the MaaS and their associated data. Hence, Sagittarius uses descriptive data min-ing techniques (i.e., association rules [46]) to reliably find out new behavioral patterns and trends from these data [24] as can be seen in Fig. 7. Specifically, Sagittarius induces production rules, using the Apriori algorithm [1], of the form X → Y , which denotes that when X happens then also Y happens—assuming that both X and Y are any set of feature-value pairs concerning the students or the course. Every rule is provided with a value of support and a value of confidence. The support is an indicator of the ratio of times (i.e., frequency) that the items involved in the rule occur together. The confidence indicates the ratio of times the if-then statements have been found true.

It is important to note that, in the context of teaching, these pairs of characteristics and values (i.e., X and Y) describe events that take place at a specific moment in the teaching activity, that is, they have a temporal order. Taking this into consideration, Sagittarius can guarantee that all the facts used in the rules respect the order of events in time (i.e., causal rules). If the user selects this configuration (which is the recommended one), then the system discards those discovered rules that present in their consequent Y an element that happens before those present in the antecedent X (see Fig. 7).

Also, the teaching staff can adjust the maximum number of rules that need to be discovered. The rules discovering process is as follows:

1. First, Sagittarius configures the Apriori algorithm with the highest confi-dence level, that is 1.0 (i.e., 100%). Then, Apriori is executed with the goal of getting as many rules as possible. If the number of obtained rules with maximum confidence is lower than the number selected by the user, the following step is executed. Otherwise, the last step will be executed.
2. If the number of rules selected by the user is not reached in the previous step, Sagittarius reduces the confidence parameter of the Apriori algorithm in 0.05 (i.e., 5%) and reruns it to discover more rules. This step is repeated until the desired number of rules is obtained. Note that this process would stop when either the maximum number of rules that need to be discovered is reached or the confidence parameter equals to 0.
3. Finally, once the desired number of rules is obtained, they are ordered according to their relevance, confidence and support, respectively. Then, the user can inspect them and—manually—select which ones are of his/her interest.

See, as an example, the following rule that shows an association between a continuous assessment test (CA) and the final exam (F E):

$$\textbf{\textit{if}}\ CA1 > 9 \wedge 5 < CA2 < 6\ \textbf{then}\ FE > 4\ (66\%, 94\%).$$

This rule has a support of 66% and a confidence of 94%. From this information, the teaching staff can infer that mastering the concepts of the first CA quiz and passing the second one is not enough to acquire the required knowledge degree to face final exam with a guarantee of success, which may suggest the teaching staff to act accordingly.

5.2 Outcome Prediction Module

Every student has their very own needs to succeed in a particular test [60]. In this regard, Sagittarius predicts the grade that students may obtain in a particular evaluation based on the internal knowledge model that the platform builds. This model is obtained from the aforementioned CSV file through the application of a top-rated artificial intelligence algorithm, the decision tree C4.5 [61]. Hence, this module generates a human-readable knowledge model from these data and predicts the future student results.

C4.5 is a predictive algorithm that bases its prediction potential on obtaining a knowledge model induced from previous experiences. Therefore, this algorithm, unlike descriptive algorithms (such as the Apriori used in Section 5.1), requires a training process prior to its exploitation. For this reason Sagittarius requires as much data as possible (e.g., student records, grades, interactions with the LMS, etc.) that corresponds to a minimum of one academic year prior to the current one, and that has the same curricular structure as the one for which it is making the prediction. Once this is available, Sagittarius: (1) configures the C4.5 algorithm with the target attribute to be predicted (e.g., obtained grade in the final exam), (2) obtains the knowledge model based on the data available from previous year(s) for that target attribute and (3) predicts the academic results of current students based on the available data.

At the end of the academic year, once all the students data is available, it is possible to add new data to the C4.5 experience data set of previous courses. In this way, the knowledge base in which C4.5 builds its model can be broadened. However, note that this is a critical process that should be thoroughly done in order to avoid introducing noise or imposters into the knowledge base—this type of data is not representative of reality, it may hide other facts and could, in consequence, compromise the level of accuracy of the prediction made with the C4.5 algorithm.

In this way, Sagittarius can perceive any deviation of the course concerning students' performance and, thus, generate an alarm to let the teaching staff move accordingly (i.e., think globally and act locally).

Additionally, teachers can realize which students: (1) are unmotivated—those predicted as Not Attended (NA)—to prevent them from leaving the course, (2) have difficulties on achieving the desired knowledge degree—those classified as Fail (D) or (F)—to give them further exercises and bibliography,(3) will pass the exam with some guarantees that they achieved the desired learning outcomes—those predicted as Pass (B)—to reinforce their knowledge, and (4) apparently have achieved the desired knowledge—those students classified as Excellent (A)—to provide them with additional voluntary challenges.

5.3 Group Building Module

Properly managing working groups (e.g., size, member profiles, targets) is essential for collaborative environments. In fact, there is a noteworthy correlation between the individual skill set of every group member and the final group out-come [57]. Therefore, the teaching staff is committed to pay special attention on configuring these groups in order to maximize the (1) knowledge quality attained by every member, (2) performance of the class, and (3) associated social competences to collaborative work of every student (e.g., conduct teamwork, ability to apply new learning strategies, emphasize problem solving skills) [48, 4]. Thus, according to the specific needs and goals of every learning activity, the teaching staff may select a given group building strategy. For instance, the teaching staff may decide to place all students with similar profile in separate groups in order to build a heterogeneous configuration. However, when facing educational environments such ad the MaaS with hundreds of students and poorly reliable and summarized information concerning their condition, building groups as in face-to-face lectures becomes time-consuming and ineffective. Hence, this module proposes a group configuration layout (i.e., connection between students and groups) that is automatically constructed by using the K-means clustering method [62] over the previously built abstract model of student's learning. Note that, again, K-means is a descriptive technique, so it should be noted that no training process is necessary.

Using this module, the teaching staff can weigh the features (also referred to as attributes) that characterize students (e.g., obtained grade at a given quiz, acquired competences, preferred interests). Thus, two group configurations are possible:

1. Homogeneous groups. Group members have very similar interests, knowledge, or competences. They might be used for those time-constrained activities, where slight social disruptions may drive the group to not succeed. In order to achieve a high degree of homogeneity, Sagittarius configures the K-means algorithm using a K equal to the desired number of members per group. Based on the K partitions obtained, the preparation of each of the groups is done in such a way that all the members of the same cluster are selected in an iterative way until the maximum number of students per group is reached. If all the members of a cluster have been assigned, then the same process is followed with the next cluster

that has elements pending to be assigned. Sagittarius shows for each group its members and their characteristics (see Fig. 8), and also the degree of homogeneity by group, a metric that indicates the number of students in the group who belong to the same cluster (e.g., a degree of homogeneity 4/6 indicates that 4 out of the 6 members are part of the same cluster).

2. Heterogeneous groups. Group members have complementary interests, knowledge, or competences, as shown in Fig. 8. They might be used to explore the social competences required at some subjects and, thus, pursue ambitious goals individually unreachable for several group members.

Student	Java Coding	UML Modeling	Leadership	Programming Fundamentals	User Experience
student38	5.10	7.50	7.05	2.00	6.50
student67	5.80	NA	NA	NA	NA
student47	5.10	NA	3.50	2.00	2.90
student61	9.30	7.00	4.10	2.00	2.15
student54	10.00	8.00	7.80	10.00	4.00

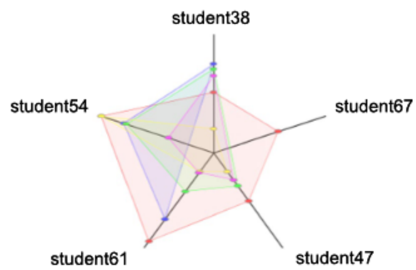

Fig. 8 Group monitoring. Example of how Sagittarius graphically plots the competences of a working group. Input data collected from students records on top. Data visualization provided by Sagittarius on the bottom.

In order to generate heterogeneous groups, Sagittarius also configures the K-means algorithm with a value of K equal to the number of members per group. In this case, using the descriptive model obtained, each group is made by selecting an element from each of the K clusters discovered.

Note that the weight and relevance assigned to every attribute is given by the teaching staff according to the course characteristics, which permits selecting several degrees of homogeneity and heterogeneity.

5.4 Students Profile Detection Module

A deep understanding of the student behavior is critical in order to success-fully attend its demands in terms of knowledge and guidance [24]. For instance, knowing in advance the preferences and shortages of a reduced set of students (also referred to as category), the teaching staff may be able to assign a specific set of exercises and activities to fulfill their needs, which may contribute to changing the group dynamics in order to reach the desired academic targets. Despite the obvious benefits of this methodology, it is very expensive in terms of resources to find out the appropriate category corresponding to every stu-dent. This situation often drives the teaching staff to develop standard and non-personalized contents that lead to courses with poor learning support for students.

To revert this situation, Sagittarius uses the *X-means* clustering method [62] to automatically analyze the aforesaid abstract model of the students' learning process and discover a closed set of profiles inside the classroom. Additionally, Sagittarius allows to weigh every student feature in order to obtain an accurate profile layout of the students. With this information, the teaching staff can develop personalized contents for students.

X-means is an algorithm that belongs to the family of descriptive tech-niques in the field of Data Mining and, thus, does not require the execution of a previous training phase. Sagittarius configures minimum and maximum number of clusters (referred to as K_{lower} and K_{upper}) to discover respectively, depending on the user's preferences. If the user does not specify the minimum and maximum number of profiles, then Sagittarius sets $K_{lower} = 2$ and $K_{upper} = 5$, a suitable range that enables X-means to discover:

1. At least the 2 partitions (i.e., $K_{lower} = 2$), which allows the identification of students with high academic performance and those with low performance.
2. A hypothetical distribution of student academic performance in these 5 categories (i.e., $K_{upper} = 5$): NA (Not Attended), F (Failure), E-D (Approved), C-B (Remarkable) and A (Excellent).
3. Any partition between both values (i.e., 2 and 5) that represents a better distribution of the students.

This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

It should be noted that, if the teaching staff needed a more fine-grained analysis, it would be possible to specify any pair of values as the minimum and maximum number of profiles. In fact, *X−means* will find the most appropriate number of clusters (between K_{lower} and K_{upper}), to minimize the variability of the characteristics of the students in the same cluster, while maximizing the variability of the characteristics of students in different clusters.

Overall, Sagittarius offers a powerful learning management environment to support collaborative work by combining the information provided by these modules. Sagittarius combines the aforementioned modules to (1) intelligently assist in the group creation, (2) automatically monitor the progress of every member in the class, and (3) effectively support the teaching staff on managing all the assessment and teaching actions derived from collaborative-based teaching methodologies.

6 Framework for cross-disciplinary training in Big Data

The MaaS approach, the TICVA framework, and the Sagittarius tool form a learning ecosystem based on a teaching methodology that has been designed to be suitable for teaching Big Data and its associated fields: datacenters, storage and processing technologies, business intelligence, and data analytics.

The MaaS approah enables synergies between students and companies, in which students with different profiles come together to achieve a common goal: solve real-world Big Data challenges in industry. TICVA is a framework aimed at fostering a personalized and social learning process while providing guided learning about a specific knowledge domain in a virtual community. Sagittarius' puts the most representative problems of machine learning at the service of teaching, such as the benefits offered by the modules for the cre-ation of working groups, profile detection, prediction of academic performance as well as discovery of trends in the student learning process. Taking into con-sideration the characteristics and potential of each of them, and having the functional modules of a VLE, a teaching methodology has been designed to address the challenge of learning Big Data and Supercomputing. This frame-work is contextualized within a learning approach based on interdisciplinary projects originated by the MaaS initiative, in which the TICVA framework is implemented based on the main functions of a VLE and the Sagittarius tool.

This section describes on the adoption of the proposed Sagittarius system to deploy the MaaS approach when training multidisciplinary students in the discipline of Big Data and their associated fields.

As shown in Fig. 9, students from the master's programs in Digital Trans-formation, eHealth, Smart Cities, Software and Apps development, User Ex-perience, and Robotics were proposed a joint Master's thesis project aimed to train them in Big Data and Supercomputing. Hence, students with differ-ent backgrounds and heterogeneous profiles had 5 months to come up with a feasible solution to the proposed challenge.

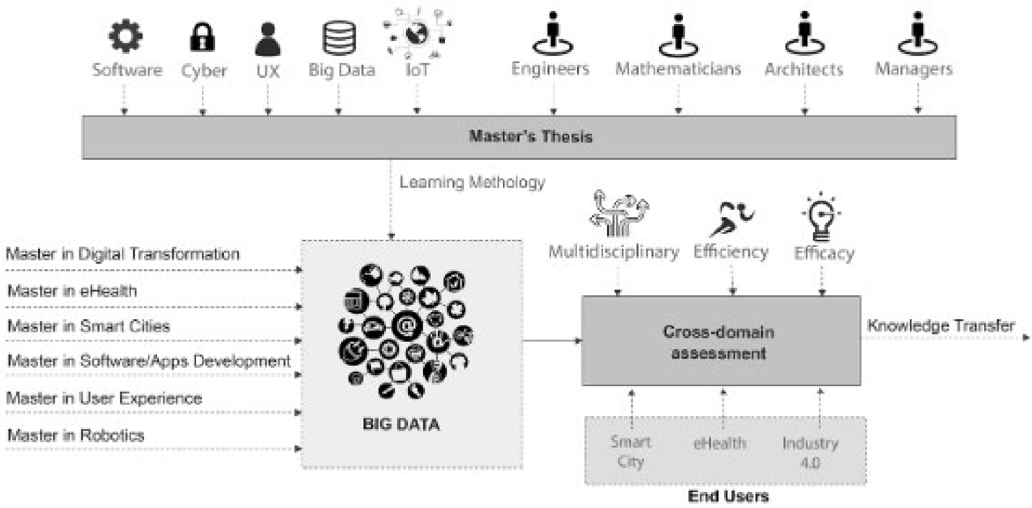

Fig. 9 Conceptualization of the Master as a Service approach for Big Data training

The proposed teaching methodology comprises 6 stages, which are de-scribed below.

This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

6.1 Challenge definition

The process begins by specifying the requirements of the challenges proposed by companies in the Big Data sector. In this phase, professionals, with the guidance of experts in the field, propose real-world challenges. Each of these includes a set of mini challenges to be completed that provide guidance for the complete achievement of the challenge. This "recipe" guides students and forms the basis of project-based learning. In what follows, an example of challenge proposed in the latest edition of MaaS is summarized:

The challenge was posed by a cybersecurity company in charge of securing critical infrastructures. Specifically, the challenge proposed to (1) manage all data generated from the Security Information and Event Management (SIEM) systems of an intercity water management system (which covers the fields of infrastructure design and massive data storage), (2) display the data to cyber-security experts in a human readable way (which includes the fields of business intelligence and data processing), and (3) design a system to forecast threats and faults (which covers the field of data analytics). Additionally, this challenge covers topics from software development, cyber security, user experience, Big Data and Internet of Things.

Also, this challenge, had associated a set of mini challenges that guided its resolution and were designed in an incremental way:

The very first thing they had to understand what a SIEM was, which kind of data they should expect and how these data could affect a city——where students from the masters in Smart Cities and Robotics led their associated groups.

Table 1 Example of group configuration

Group	Student	Incoming academic program
1	*student.1_UX*	Master in User Experience
	student.14_Sw	Master in Development of Software Applications
	student.16_eH	Master in eHealth
	student.7_SC	Master in Smart Cities
	student.1_DT	Master in Digital Transformation
	student.11_Cb	Master in Cybersecurity
2	*student.6_UX*	Master in User Experience
	student.18_Sw	Master in Development of Software Applications
	student.17_eH	Master in eHealth
	student.6_SC	Master in Smart Cities
	student.2_DT	Master in Digital Transformation
	student.12_Cb	Master in Cybersecurity

Next, they had to get familiarized with (private) cloud computing environments——where students from the Master in Digital Transformation led their associated groups. Later, they had to understand the implications of dealing with sensible data——where students from the Master in eHealth lead their associated groups. Next, they had to come up with a software architecture (i.e., based on the Hadoop ecosystem) capable to collect, store, and process all these data in a scalable way——where students from the Master in Software Development led their associated groups. Next, they had to learn how to present massive amounts of data and build a dashboard——where students from the Master in User Experience led their associated groups. Finally, they had to develop a small forecasting system——as this topic was not closely related to any of the graduate programs involved in the MaaS initiatives, all members in every group collaborated to gain the required knowledge.

6.2 Creation of work teams

In the second phase, and through the Sagittarius tool, the teaching team generates the work groups automatically according to the learning objectives and the characteristics of MaaS. Specifically, based on information regarding academic performance, the teaching team selects a heterogeneous configuration aimed at maximizing group performance. As suggested in [51], it is of paramount importance to balance the skill set of students with different ability levels in order to boost the effectiveness of group activities [10].

For the sake of the aforementioned MaaS challenge, the teaching staff run the Sagittarius software to obtain heterogeneous working groups. For instance, Fig. 10 shows a diagram with the characteristics of the following heterogeneous work teams automatically obtained with the Sagittarius tool:

This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

Group	Student	Incoming academic program
1	*student.1_UX*	Master in User Experience
	student.14_Sw	Master in Development of Software Applications
	student.16_eH	Master in eHealth
	student.7_SC	Master in Smart Cities
	student.1_DT	Master in Digital Transformation
	student.11_Cb	Master in Cybersecurity
2	*student.6_UX*	Master in User Experience
	student.18_Sw	Master in Development of Software Applications
	student.17_eH	Master in eHealth
	student.6_SC	Master in Smart Cities
	student.2_DT	Master in Digital Transformation
	student.12_Cb	Master in Cybersecurity

Note that although both groups have the same configuration considering the incoming academic program—it might happen that more than one student from the same incoming academic program would be included in the same group—, Sagittarius has considered the students' academic performance (see Fig. 10) to create heterogeneous, yet balanced, working groups.

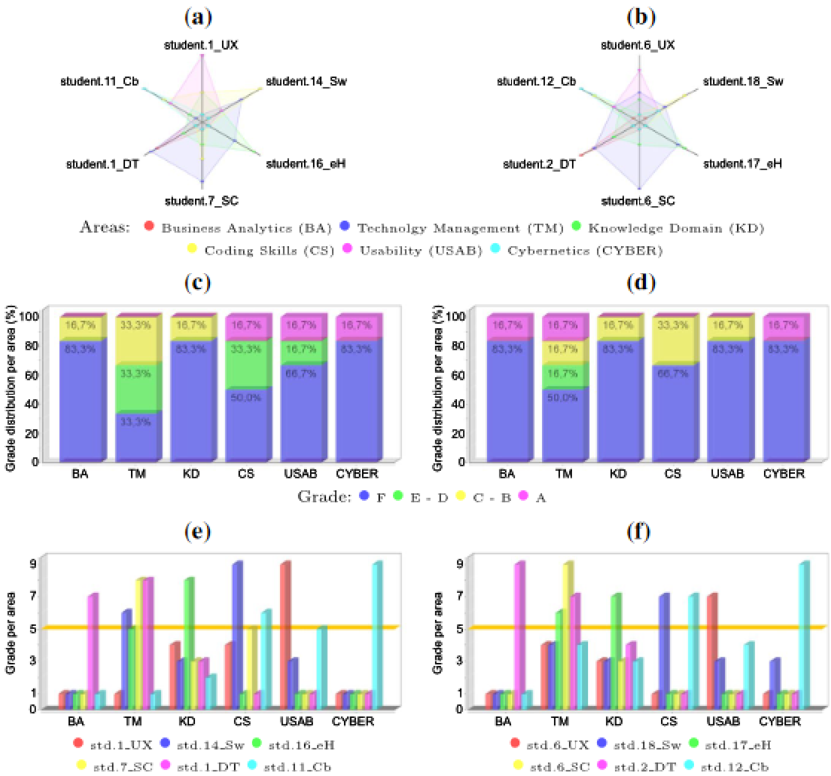

Fig. 10 Level of knowledge representations of the 6 team members—from different academic programs—of two working groups.

Indeed, as shown in Fig. 10, the level of knowledge acquisition of each area of knowledge is represented in a very graphic and intuitive way by each member of a group. According to the graduate programs involved in the MaaS, we have defined the following areas of expertise:

– *Business Analytics*, skills involving data analysis for business decision making based on previous experience.
– *Technology Management*, skills to appropriately use technology in order to obtain a competitive advantage.
– *Knowledge Domain*, skills and knowledge to extract and understand the requirements of a specific domain.
– *Coding Skills*, skills to design and code processes by means of algorithms and computer programming.
– *Usability*, skills to increase the level of the user satisfaction when using a system.
– *Cybernetics*, skills to understand and apply knowledge from different disciplines in order to conceive robotic systems.

Radial graphs (Fig. 10.a and Fig. 10.b), percentage graphs (Fig. 10.c and (Fig. 10.d) and bar graphs (Fig. 10.e and Fig. 10.f) show that Sagittarius has grouped students with different levels of competences into the same group (group 1 or group 2), being able to complement individual weaknesses result-ing in group configurations with an acceptable level of each competence and knowledge. As an example, look at the graphics shown in Fig. 10.a, Fig. 10.c and Fig. 10.e)

In addition, the teaching staff can qualitatively examine the amount of knowledge that the group owns at every individual facet by analyzing the size of every colored area and, thus, obtain a preliminary idea of possible group imbalances. In this way, students with different backgrounds and skill sets——in addition to those they had already gained in their graduate program——such as engineers, mathematicians, architects, managers) are divided into groups to address the Big Data challenge.

Note that the groups generated with Sagittarius are created in the VLE through the main functions of student grouping, that is, those that implement the identified student and group modules of the TICVA framework shown in Fig. 3. Hence, students are automatically informed about who their teammates will be for facing the Big Data challenge.

In the undesirable situation in which a member of a team withdrew from the working group or the Master Program, the knowledge balance in each time proposed by Sagittarius would be affected. In this case, Sagittarius would be still of great use for the teaching staff:

– Proactive student tracking. Using the Trends Detection and Outcome Prediction modules of Sagittarius (see Sections 5.1 and 5.2) the teaching staff can proactively identify those students that may potentially quit the course (e.g., low odds of passing the final exam, little interaction with the LMS, etc.) and provide them with extra tutoring sessions to track their progress.
– Student groups reconfiguration. Using the Students Profile Detection module (see Section 5.4) the teaching staff can look for students with similar characteristics than the one who have left the group. Later, using the Group Building module (see Section 5.3) the teaching staff can quantitatively foresee (see Fig. 10) the effects on moving one student from a group to another group or destroying the group and moving students to other groups.

To sum up, human intervention is unavoidable when this kind of extraordinary circumstances arise. However, the automatic data analysis over students data conducted by Sagittarius saves a lot of time to the teaching staff.

6.3 Publishing and choosing challenges

The companies present their offer of challenges (i.e., the ones that were conceived in the first phase), and the students, previously grouped by the system, choose the one that interests them the most. This choice creates a relationship between a group of students and a mentor or professional who monitors the development of the challenge, which is stored in the VLE through the corresponding modules identified in the TICVA framework. The VLE also contains all those materials and resources that may be used to face each of the challenges that are published.

The following phases, 4 and 5, involve the communication modules, students, groups and mentors that generate the TICVA pedagogical module in-puts. The blocks equipped with Sagittarius machine learning algorithms respond to the submodules of the prediction and automatic tutoring system of the TICVA pedagogical module. Thus, mentors and academics benefit from Sagittarius' potential to (1) detect non-trivial trends in student performance,(2) predict student performance and (3) obtain a descriptive model of student profiles. With these benefits, it is possible to offer personalized attention that improves the development and training of students in the Big Data discipline. It should be noted that these 2 phases are repeated in an iterative way until the challenge that ends in phase 6 of the process is resolved.

6.4 Data collection

The fourth stage of this process is to allow Sagittarius to collect enough information to monitor and evaluate groups' progress and performance. This information emerges from the monthly meetings that mentors hold with their respective groups in order to guide them in the development process and monitor their performance. At this point, the tool updates the abstract learning model according to previous experiences and new ones (e.g., obtained grades, number of posts written in discussion forums (PF), minutes spent reading lectures (TL)). This numerical information allows the teaching staff to have a real and objective perception of what it is going on the course and, further, compare it with previous editions. More specifically, it is possible to (1) detect those group members that can put the group performance at risk, (2) quantify the amount of knowledge that every individual reaches, (3) assess how the collaborative work influences on the individual skills level (i.e., up to what extent the area built on the initial group creation stage changes), and (4) discover the group evolution.

Note that Sagittarius strongly relies on a continuous interaction between student and the LMS provided by the university. If students chase to interact and exchange materials between each other without using this infrastructure the system would build poorly reliable models (i.e., there will always be a mini-mal interaction as a result of the different assessment activities). In this regard, the teaching staff must (1) properly set up all the necessary resources (e.g., forums, chats, virtual classrooms, dialogues) to effectively conduct collaborative work, and (2) encourage students to use the university's infrastructure by explaining them its benefits and advantages.

6.5 Intelligent mentoring

The teaching staff uses its essential experience to supplement the information obtained in the previous phase to perform some actions in order to drive the groups into a desired state. More specifically, the teaching staff can (1) design and deliver complementary material to motivate and regain those critical students for their groups, (2) adjust the amount of knowledge delivered to students in order to achieve an homogeneous knowledge level and prevent possible burn outs, (3) evaluate the effectiveness of the delivered contents and improve them accordingly, and (4) adapt the online environment to current students' behavior (e.g., if forum discussions are underused the teaching staff could consider reviewing the usability of the learning management system platform or find alternatives such as chat rooms).

Bar graphs shown in Fig. 11.a and Fig. 11.b, percentage graphs shown in Fig. 11.c and Fig. 11.d and pie graphs shown in Fig. 11.e and Fig. 11.f represent the student profiles detected by Sagittarius. Specifically, the tool has analyzed 149 students using two configurations: the default one and the weighted one. Also, the student is characterized using 11 attributes that come from the different evaluation activities that he/she has taken:

- *t_week0, ..., t_week5*, results obtained in 6 tests.
- *t_avg*, average of the tests.
- *p1p*, grade obtained in the first assignment.
- *p2p*, grade obtained in the second assignment.
- *exa_a*, grade obtained in the first exam.
- *exa_b*, grade obtained in the second exam.

When the default configuration is applied, Sagittarius groups the student into 3 profiles (graphs (11.a), (11.c) and (11.e)) according to their present academic performance. Profile 1 is made up of 91 students (61.07% of the total) and represents those who have obtained satisfactory results in all the assessment criteria except the exams (*exa_a* y *exa_b*). Profile 2, made up of 25 students (16.78% of the total) represents students with a slightly lower performance than that required in the tests (*t_week0, ..., t_week5* y *t_avg*) and a very poor performance in the practical work (*p1p* y *p2p*) and the exams. Finally, profile 3, made up of 33 students (22.15%), represents those with an unsatisfactory performance in all the graded items. This knowledge is a vital tool which complements the experience of the professor, enabling them to offer the student a more personalized approach in order to achieve better results and help students reach the goals they have been set. For example:

- Students who correspond to profiles 1 and 2 can be assigned additional material which specifically addresses the weak points detected in their tests. However, even though the students may acquire some knowledge, they are still not well enough prepared to successfully complete the practical work which corresponds to the subject (profile 2 students) and/or the exams (profile 1 and 2 students). Specifically, in Fig. 11.a it can be seen that profile 2 students have failed the practical assignments p1p and p2p and, profiles 1 and 2 have failed exams *exa_a* and *exa_b*.
- Profile 3 students, whose performance is poor in all of the areas of the subject (see Fig. 11.a), should be assigned a personalized study plan in order to help them acquire the knowledge associated with the exercises which test the fundamentals of the subject or topic in question.

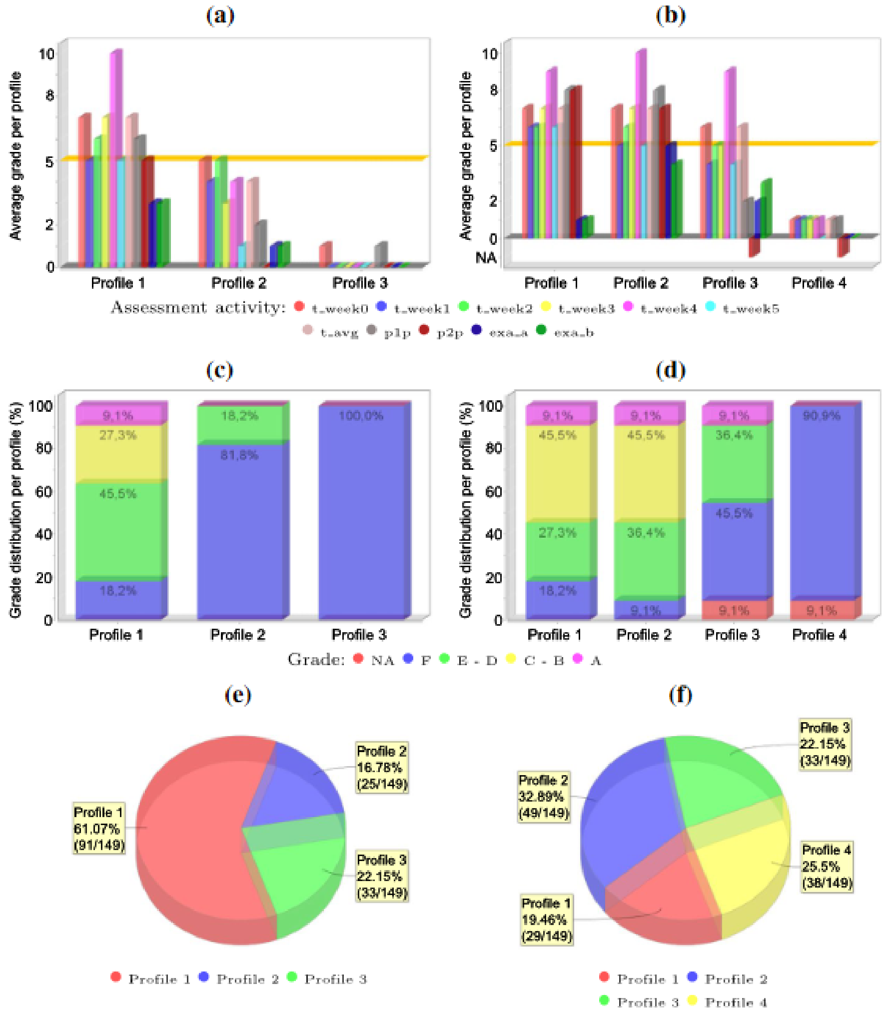

Fig. 11 Academic performance of each student profile using two different configurations. The column on the left (a), (c), (e), corresponds to a manually selected configuration of 3 profiles. The column on the right (b), (d), (f), corresponds to an automatic configuration where Sagittarius has discovered that there are 4 different profiles in the classroom.

Sagittarius enables the professor to configure the degree of importance of each of the characteristics which define the student profile. If we go back to the previous simple of 149 students, a different result is obtained (11.b),(11.d) y (11.f)) when the configuration is modified according to the weight of each gradable item: 0.2 for *t_week0, ..., t_week5* and *t_avg*, 0.4 for p1p, 0.6 for p2p, 0.8 for *exa_a* and 0.4 for *exa_b*. As we can observe in this case, the tool has detected 4 profiles. Profile 2, with 49 students (31.89%), represents the students whose academic performance is considered to be good. Profiles 1 and 3, are antagonistic when it comes to performance in practical work (p1p y p2p) and exams (*exa_a* y *exa_b*). To be precise, the 31 students (20.81%) from profile 1 perform better in practical work than in exams. However, the 30 profile 3 students (20.13%) perform better in exams than in practical work.

As for profile 4, the 39 students (26.17%) from this group are in need of special attention given their poor academic performance in general.

The previous example, if detected within the correct timespan, enables the professor to decide upon the appropriate measures to adapt the course content to each profile of student.

The Sagittarius profile detection is applied at various stages of the academic year in order to observe: (1) the current performance of the student and (2) her or his progress over time. This enables the professor not only to cater for the individual needs of each student, but also to obtain objective feedback on the consequences of the decisions s/he has taken.

In addition, during this iterative process a set of rules are created. From these association rules, the teaching staff can add an extra degree of personalization to the online course. The information derived from these rules can be used to adjust the quality of the online course (i.e., objectively assessing the influence of every content) by adapting its syllabus to the students behavior and demands. For instance, from the aforementioned example of 149 students, the following two rules have been selected:

1. if *p1p=NA* then *p2p=NA* (33%, 100%)
2. if *t_week0=F* and *t_week2=F* then *exa_b=F* (22%, 100%)

Rule 1, expresses the fact that students who do not submit part 1 of the practical work also do not present the part 2 with 33% support and a reliability of 100%. This leads us to the following suppositions (1) review the connections between the 2 phases of the practical work and determine whether they are suitable if the student fails to submit the first part and (2) dedicate their efforts to undertaking a more exhaustive monitoring of the groups that fail to submit the first phase, in order to avoid the same thing happening in the second submission.

Rule 2, establishes the relations between 2 of the weekly tests and the exam for block b. More specifically, note that if a student fails the tests in weeks 0 and 2, s/he fails the block b exam with 22% support and 100% reliability. Thus, it can be observed that the content of the 2 tests are representative and determine the student's performance on the block b test. Therefore, at this point, the professor can decide to provide students who do not pass the initial tests with supplementary material to improve their academic performance. These supplementary materials are often related to the generic key topics that the course covers and evaluates. In addition, the mentor (i.e., industry professional or domain expert) of every working group provides the students with a set of domain-specific supplementary materials related to the challenge they are addressing.

Note that the evaluation activities (e.g., tests) are designed to assess up to what extent students have consolidated the key topics of the course.

As the course progresses, stages 4 and 5 are repeated until achieving an ideal steady state where a unique group type that encompasses all teams (i.e., all the groups own similar properties in terms of knowledge level and performance) is identified. Hence, as a result of using this methodology, the initial heterogeneous layout have become homogeneous.

Once the challenge has been developed and under the acceptance of its completion by the professionals and the teaching team, we arrive at the final phase, phase 6.

6.6 Presenting results

Students present the highlights of the development of the challenge to a panel made up of professionals and academic experts in the field of Big Data. After presenting the results and the conclusions obtained, the panel asks the students the questions they deem appropriate.

As it can be seen, with the application of the MaaS initiative supported by the implementation of the TICVA framework through a VLE and the use of the Sagittarius tool, a specific methodology has been proposed aimed at success-fully conducting interdisciplinary project-based learning for Big Data learning. This methodology covers aspects that go beyond the individual follow-up or group of students for an adaptation to the individual needs, offering key information for the continuous improvement of the student follow-up process and the implementation of initiatives focused on the learning of Big Data.

Fig. 12 shows the number of students enrolled to the Master's Programs that delivered the Final Master's Project (FMP) on top and the results of the student satisfaction surveys after having delivered the Final Master's Project are shown on the bottom. As far as the number of successful delivered FMPs there we have been unable to find any significant correlation between the delivered FMPs and the adoption of the MaaS approach. As shown in data from years 2015 and 2016 in Fig. 12, we believe that in the context of Master's programs students already put a lot of effort on finishing their studies to obtain their diploma and, thus, rapidly get a job. In this survey students rate their level of satisfaction from 1 (Terrible) to 5 (Great). Taking into account that the implementation of the MaaS initiative began in 2017, it can be seen that thanks to MaaS, the degree of satisfaction has been significantly improved (from 3.5 to 4.7) and, at the same time, has achieved a lower dispersion in the survey results (that is, a greater number of satisfied students). In addition to the grades, students had the opportunity to leave written comments, In these comments, some of the students found it very interesting to spend the Master's thesis project on a topic that was outside of the standard syllabus rather than repeating what they had already done and learnt in their regular classes. Also, some of them mentioned their reluctance to work with students with such a different background, but they rapidly changed their minds when they realized that it was an effective way to address challenges that they would otherwise find impossible to solve individually.

This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

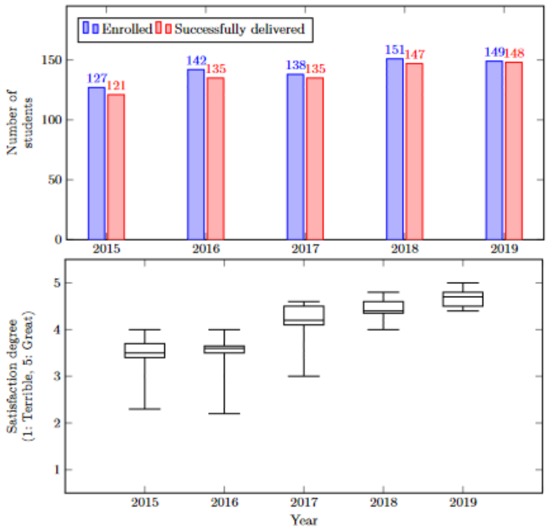

Fig. 12 Results of the delivered Final Master's Project (top) and MaaS satisfaction surveys (bottom) conducted in 2015, 2016, 2017, 2018, and 2019.

7 Discussion and Conclusion

This paper presents TICVA, a methodological teaching and learning frame-work to include an automatic and intelligent tutoring system, coined as Sagittarius, to train graduate students in Big Data using the MaaS approach.

The MaaS paradigm aims to combine students from different academic programs to train them in a specific topic (i.e., Big Data) by exploiting collaborative work. For the sake, of this work, we have taken 149 students from different programs such as Digital Transformation, eHealth, Smart Cities, Soft-ware and Apps development, User Experience, and Robotics. When running the MaaS approach in previous editions, we found out that it was very hard to individually analyze every student profile—in order to build effective working groups—and performance—in order to optimize his/her development during the course. Also, we found that it was unfeasible to track all the interactions between students themselves and also with the LMS due to the ever growing number of data that the MaaS philosophy generates. Overall, all of this, strongly penalized on the quality of the feedback delivered to students and their experience during their training.

Therefore, we have proposed TICVA, a methodological framework aimed to include an automatic ITS inside the teaching and learning ecosystem, which enables the teaching staff to automatically collect data and provide individualized feedback to students. Indeed, TICVA abstracts the learning process proposes a set of modules to combine the advantages of VLEs with the advantages of ITSs. Such a combination, results in a very convenient approach to effectively deploy the MaaS over a large number of students.

In this regard, we have developed Sagittarius, a software tool that uses ma-chine learning and data mining algorithms to assist the teaching staff on the management and operation of the course. Indeed, Sagittarius can be best seen as a tangible implementation of the TICVA framework to address the issues posed by the MaaS when facing a large number of students. More specifically, Sagittarius is devoted to (1) assist in configuring and setting collaborative working groups from heterogeneous profiles, (2) effectively manage the course contents according to every individual student learning demands by automatically analyzing their performance, (3) track and simulate students' learning progress, and (4) mine the massive data generated by students and their inter-actions. Sagittarius, as a result of this automatic data mining process, is able to discover hidden rules in the vast amount of data generated by the MaaS, which assists the teaching staff in understanding the concrete effects of every particular teaching action.

It can be seen that Sagittarius strongly relies on the amount—and quality—of data collected from students. Typically, these are very sensitive data, such as historical grade records, that are under the General Data Protection Regulation (GDPR) and other ethical directives. Therefore, special care has to be taken when putting this tool into production environments. To avoid jeopardizing the data security policies of the campus, the following recommendations should be taken:

– Secure deployment. Sagittarius should be installed at the same server where the grades database is hosted, which enables itself to automatically inherit all its cyber security defenses and protocols already installed.

– Anonymize students identities. The four modules of Sagittarius (student's trend detection, outcome prediction, group building, and students pro-file detection) can build their knowledge models with anonymous (i.e., hashed) students identities. Therefore, data supplied to Sagittarius should be anonymous.
– Access control. Access to the Sagittarius platform should be only granted to the teaching staff (who already have access to the students records) in order to avoid leaking sensible data.

In this way, Sagittarius can adapt to the personal data protection standards of educational centres.

Overall, the combination of MaaS, TICVA and Sagittarius to train students in Big Data provides valuable outcomes from all the actors involved in the learning process: from the instructor's side it contributes to the continuous improvement of the course materials to make them appealing and effective, from the industry experts side it contributes to understand the progress of their students, and from the students' side it eases the process of satisfying every individual needs, which improves the learning experience.

8 Future work

During the last three years, the implementation of the MaaS initiative by means of the TICVA paradigm and Sagittarius has enhanced the learning and teaching process in our campus. Along these years we have identified some aspects that should be addressed in the forthcoming future.

First and foremost, it would be very convenient to have a seamless in-tegration of Sagittarius with the LMS. So far, student records are manually extracted from the LMS, converted to XML files, and manually uploaded to Sagittarius. It would be very interesting to develop a plugin for existing LMSs (e.g., Moodle, edX, Chamilo, Canvas) to interface with Sagittarius and, thus, avoid the tedious process of managing multiple files and formats.

Furthermore, it may be worthy to analyze the impact and effects of auto-matically publishing all the rules and knowledge inferred by Sagittarius not only to the teaching staff but to the students as well, in order to let them know how their progress is being tracked. In this way, the student would see as a retroaction what is his/her progression forecast according to Sagittarius. This could be used as a stimulus to make students more conscious on their achievements and enrich their continuous assessment process.

Indeed, engagement is fundamental in working groups. As mentioned in Section 6.2, one member of the group could decide to withdraw from the team, which would break the knowledge balance of the group and would risk its success. With the advent of chatbots and their massive improvements, authors believe that it would be challenging to explore the possibility of conceiving a virtual partner (i.e., bot) to replace the member who quit the group. This virtual partner should have the same amount of knowledge than the physical person who is replacing and behave similarly. We believe that this would be of great help when adjusting unbalanced working teams or giving them some reinforcements.

Sagittarius is currently committed to identify generic knowledge needs rather than domain-specific demands (which must be detected and fulfilled by the mentor of the group). Therefore, the recommended supplementary materials mostly cover generic, yet necessary, aspects of Big Data. It would be very interesting to supply additional metadata for each working group related to the specific challenge they are addressing. In this way, it would be possible to adjust Sagittarius to make specific recommendations for supplementary materials—supervised by the domain expert—related to the individual challenge that each working group is solving.

Last but not least, it would be interesting to consider the weight effects of social interactions in the teaching and learning process. It may seem that Sagittarius automatically builds knowledge models from students and, potentially, would be able to interact with them autonomously, which may question the existence of the teaching staff. Nonetheless, authors believe that trainers still play a very important role designing the learning activities, assessment strategies, addressing especial student demands, or conceiving new challenges among many others. Sagittarius should be deemed as a tool to enable the teaching staff on these activities rather than the repetitive and time-consuming ones.

Acknowledgments. Authors acknowledge the help and collaboration of Guillem Villa on the restyling of the graphical artworks of this paper. This research was partially supported by Secretaria d'Universitats i Recerca of the Department of Business and Knowledge of the Generalitat de Catalunya under grants 2017-SGR-934 and 2017-SGR-977. Also, authors gratefully acknowledge the support of NVIDIA Corpo-ration with the donation of the Titan Xp GPU used for this research.

This is a post-print (final draft post-refeering) published in final edited form as
Solé-Beteta, X., Navarro, J., Vernet, D., Zaballos, A., Torres-Kompen, R., Fonseca, D., Briones, A.
Automatic tutoring system to support cross-disciplinary training in Big Data. J Supercomput 77, 1818–1852 (2021). https://doi.org/10.1007/s11227-020-03330-x

Conflict of interest. The authors declare that they have no conflict of interest.

References

1. Agrawal, R., Imieli´nski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on Management of data, pp. 207–216 (1993)
2. Alliance, I.E.: Graduate attributes and professional competencies. Tech. Rep. 3, Inter-national Engineering Alliance (2013)
3. Attwell, G.: The personal learning environments-the future of elearning? elearn-ing papers, vol. 2 no. 1. issn 1887-1542. En ligne: http://www. elearningeuropa. info/files/media/media11561. pdf (2007)
4. Barkhuus, L., Tashiro, J.: Student socialization in the age of Facebook. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’10, pp. 133–142. ACM (2010)
5. Bartual Figueras, M.T., Turmo Garuz, J.: Higher education and employment related skills. the point of view of employers. Revista Complutense de Educaci´on 27(3), 1211–1228 (2016)
6. Beas, M.I., Burriel, R., Lapeña, L.: Ajuste de competencias al puesto de trabajo en los titulados uji y sus experiencias previas a la inserci´on laboral. Proceedings from II Jornadas Nacionales de Estudios Universitarios. Castelló de la Plana: Publicacions de la Universitat Jaume I [CD-ROM] (2009)
7. Becerik-Gerber, B., Gerber, D.J., Ku, K.: The pace of technological innovation in ar-chitecture, engineering, and construction education: integrating recent trends into the curricula. Journal of Information Technology in Construction (ITcon) 16(24), 411–432 (2011)
8. Bell, S.: Project-based learning for the 21st century: Skills for the future. The clearing house 83(2), 39–43 (2010)
9. Bishop, J.L., Verleger, M.A., et al.: The flipped classroom: A survey of the research. In: ASEE national conference proceedings, Atlanta, GA, vol. 30, pp. 1–18 (2013)
10. Bonastre, O.M., Pe˜nalver, A., Hernansaez, J.M., Lopez, J.J., Lazcorreta, E., Bia, A.: UCD/SW: A distributed system to promote cooperative learning using semantic web and ubiquitous computing. In: Proc. of the CDVE 2008, pp. 225–228 (2008)
11. Bourdeau, J., Grandbastien, M.: Modeling tutoring knowledge. In: Advances in intelli-gent tutoring systems, pp. 123–143. Springer (2010)
12. Burk, E.: Online learning indicators. eLearn Magazine 2011(7) (2011)
13. Burriel Calvet, J.R., Beas Collado, M.: Ajuste de competencias transversales entre grad-uados y mercado de trabajo: Resultados en la Universitat Jaume I. Publicacions de la Universitat Jaume I (2017)
14. Cámara De La-Fuente, M., Cano-Rodriguez, M., Castilla-Polo, F., Chamorro-Rufián, E., Moreno-Aguayo, A.: Conocimientos y competencias contables demandados por los empleadores de los alumnos de la universidad de jaén. Proceedings from X Jornadas de Docencia en Contabilidad de la Universidad de Jaén pp. 24–26 (2015)
15. Campanini, A.: Bologna process. In: J.D. Wright (ed.) International Encyclopedia of the Social & Behavioral Sciences (Second Edition), second edition edn., pp. 741–746. Elsevier, Oxford (2015)
16. Carracedo, F.S., Soler, A., Mart´ın, C., López, D., Ageno, A., Cabré, J., Garcia, J., Aranda, J., Gibert, K.: Competency maps: An effective model to integrate professional competencies across a stem curriculum. Journal of Science Education and Technology 27(5), 448–468 (2018)
17. Chrysafiadi, K., Virvou, M.: Dynamically personalized e-training in computer programming and the language C. IEEE Trans. Educ. PP(99), 1–1 (2012). DOI 10.1109/TE.2013.2243914
18. Cipollone, S., Flores, C.A., Ahualli, N., Fatica, I.F., Denardi, P.: Valoración de competencias no técnicas en cirujanos en formación de la fuerza aérea argentina y sus estrategias de manejo. Hosp. Aeronáut. Cent pp. 44–52 (2017)
19. Collazos, C.A., Guerrero, L.A., Pino, J.A., Ochoa, S.F.: Evaluating collaborative learning processes. In: International Conference on Collaboration and Technology, pp. 203–221.Springer (2002)
20. Crouch, C.H., Mazur, E.: Peer instruction: Ten years of experience and results. American journal of physics 69(9), 970–977 (2001)
21. Espinar, S.R., Nebot, A.P., Arjona, L.B., Castiñeira, S.S.: Sobre la empleabilidad de los graduados universitarios en catalunya: del diagnóstico a la acción graduate employability in catalonia. from diagnosis to action. Revista de Educación 351, 107–137 (2010)
22. Everson, M.: What can you gain by teaching online? eLearn Magazine 2011(1) (2011)

23. Farrell, G.: The Development of Virtual Education: A global perspective: A study of current trends in the virtual delivery of education, conducted with funding provided by the Department for International Development, London, U.K. Commonwealth of Learning (1999)
24. Fini, A.: The technological dimension of a Massive Open Online Course: The case of the CCK08 course tools. International review of research in open and distance learning 12(1) (2009)
25. Flanagin, A.J., Metzger, M.J.: Trusting expert-versus user-generated ratings online: The role of information volume, valence, and consumer characteristics. Computers in Human Behavior 29(4), 1626–1634 (2013)
26. Fonseca, D., Conde, M. A., García-Peñalvo, F.J.: Improving the information society skills: Is knowledge accessible for all? Universal Access in the Information Society 17(2), 229–245 (2018)
27. Freire Seoane, M.J., Teijeiro Alvarez, M.M., Pais Montes, C.: The match between graduate-acquired and employer-required competencies. Revista de Educación 1(362), 13–41 (2013)
28. Hamblen, J., van Bekkum, G.: An embedded systems laboratory to support rapid prototyping of robotics and the Internet of Things. IEEE Trans. Educ. 56(1), 121–128 (2013)
29. Hernick, M., Hylton, A.C., Justice, M.: A process for curricular improvement based on evaluation of student performance on milestone examinations. American journal of pharmaceutical education 80(9), 159 (2016)
30. Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R., Shahabi, C.: Big data and its technical challenges. Communications of the ACM 57(7), 86–94 (2014)
31. Jaimes, E.I.G.: Competencias académicas de los egresados universitarios y su predicción de ocupación laboral. RICSH Revista Iberoamericana de las Ciencias Sociales y Humanísticas 5(10) (2016)
32. Jong, B.S., Lai, C.H., Hsia, Y.T., Lin, T.W.: Effects of anonymity in group discussion on peer interaction and learning achievement. IEEE Trans. Educ. PP(99), 1–1 (2012)
33. Kaisler, S., Armour, F., Espinosa, J.A., Money, W.: Big data: Issues and challenges moving forward. In: 2013 46th Hawaii International Conference on System Sciences, pp. 995–1004. IEEE (2013)
34. Kop, R.: The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. The International Review of Research in Open and Distance Learning 12(3), 19–38 (2011)
35. Kulich, M., Chudoba, J., Kosnar, K., Krajnik, T., Faigl, J., Preucil, L.: SyRoTek—distance teaching of mobile robotics. IEEE Trans. Educ. 56(1), 18–23 (2013)
36. Labrinidis, A., Jagadish, H.V.: Challenges and opportunities with big data. Proceedings of the VLDB Endowment 5(12), 2032–2033 (2012)
37. Lacleta, M.L.S.E., Blanco, A.F., Peñalvo, F.J.G.: Buenas prácticas de innovación educativa: Artículos seleccionados del ii congreso internacional sobre aprendizaje, innovación y competitividad, cinaic 2013. Revista de Educación a distancia 1(44) (2014)
38. Laurillard, D.: Rethinking University Teaching: A Framework for the Effective Use of Educational Technology. Routledge (2000)
39. Lemons, J.: The urgent need for universities to comprehensively address global climate change across disciplines and programs. Environmental management 48(3), 379–391 (2011)
40. Martín, S., Díaz, G., Cristóbal, E.S., Gil, R., Castro, M., Peire, J.: New technology trends in education: Seven years of forecasts and convergence. Computers & Education 57(3), 1893–1906 (2011)
41. Massy, W.F., Zemsky, R.: Using information technology to enhance academic produc-tivity. Educom Washington, DC (1995)
42. Moreira, M.A., Vargas, E.F., et al.: Buenas prácticas de aulas virtuales en la docencia universitaria semipresencial. Teoríaa de la Educación. Educación y Cultura en la Sociedad de la Información 11(1), 7–31 (2010)
43. Moreno, L., González, C.S., Castilla, I., González, E.J., Sigut, J.F.: Applying a constructivist and collaborative methodological approach in engineering education. Computers & Education 49(3), 891–915 (2007)
44. Navarro, J., Zaballos, A., Fonseca, D., Torres-Kompen, R.: Master as a service: A multi-disciplinary approach to big data teaching. In: Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality, pp. 534–538. ACM (2019)
45. O'Donoghue, J., Singh, G., Dorward, L.: Virtual education in universities: a technological imperative. British Journal of Educational Technology 32(5), 511–523 (2001)
46. Orriols-Puig, A., Casillas, J.: Evolution of interesting association rules online with learning classifier systems. In: IWLCS 2008, pp. 21–37 (2010)

47. Park, J.H., Choi, H.J.: Factors influencing adult learners' decision to drop out or persist in online learning. Educational Technology & Society 12(4), 207–217 (2009)
48. Piedra, N., Chicaiza, J., L´opez, J., Tovar, E., Bonastre, O.M.: Open educational prac-tices and resources based on social software, UTPL experience. In: Proc. of the EATIS 2009, p. 34 (2009)
49. Qiu, M., Hewitt, J., Brett, C.: Online class size, note reading, note writing and collab-orative discourse. International Journal Computer-Supported Collaborative Learning 7(3), 423–442 (2012)
50. Searson, M., Monty Jones, W., Wold, K.: Editorial: Reimagining schools: The potential of virtual education. British Journal of Educational Technology 42(3), 363–371 (2011)
51. Sigman, B.P., Garr, W., Pongsajapan, R., Selvanadin, M., Bolling, K., Marsh, G.: Teaching big data: Experiences, lessons learned, and future directions. decision line 45(1), 10–15 (2014)
52. Solé-Beteta, X., Navarro, J., Sancho-Asensio, A., Zaballos, A., Jiménez-Ruano, V., Canaleta, X., Vernet, D.: Sagittarius: A tool to enhance the collaborative work in virtual learning environments. In: 2014 9th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–6. IEEE (2014)
53. Tsai, W.T., Li, W., Elston, J., Chen, Y.: Collaborative learning using wiki web sites for computer science undergraduate education: A case study. IEEE Trans. Educ. 54(1), 114–124 (2011)
54. Vanderlinde, R., van Braak, J.: The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. British educational research journal 36(2), 299–316 (2010)
55. Vernet, D., Canaleta, X., Navarro, J., Zaballos, A., Pallas, G., et al.: Setting up and tutoring the working groups of a virtual learning community. Journal of Research and Practice in Information Technology 45(3/4), 219 (2013)
56. Vernet, D., Canaleta, X., Pallàs, G.: Intelligent tutoring of virtual learning communities. In: 2012 International Symposium on Computers in Education (SIIE), pp. 1–6. IEEE (2012)
57. Vernet, D., Canaleta, X., Pallas, G.: Intelligent tutoring of virtual learning communities. In: Computers in education (SIIE), 2012 international symposium on, pp. 1–6 (Oct.)
58. Vernet Bellet, D., et al.: Tutoritzaci´o intel· ligent de comunitats virtuals d'aprenentatge. Ph.D. thesis, Universitat Ramon Llull (2016)
59. Wells, J., Barry, R.M., Spence, A.: Using video tutorials as a carrot-and-stick approach to learning. IEEE Trans. Educ. 55(4), 453–458 (2012)
60. Wojciechowski, A., Palmer, L.B.: Individual student characteristics: Can any be predic-tors of success in online classes? Online Journal of Distance Learning Administration 8(2) (2012)
61. Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A.F.M., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., Steinberg, D.: Top 10 algorithms in data mining. Knowl. Inf. Syst. 14(1), 1–37 (2008)
62. Xu, R., II, D.C.W.: Survey of clustering algorithms. IEEE Trans. Neural Netw Learn. Syst. 16(3), 645–678 (2005)