Business Methodology for the Application in University Environments of Predictive Machine Learning Models Based on an Ethical Taxonomy of the Student’s Digital Twin

Luis Miguel Garay Gallastegui 1,*, and Ricardo Francisco Reier Forradellas 2,*,*

1 Department of Marketing, ESIC University, Pozuelo de Alarcón, 28223 Madrid, Spain
2 DEKIS Research Group, Department of Economics, Catholic University of Ávila, 05005 Ávila, Spain
* Correspondence: luismiguel.garay@esic.university (L.M.G.G.); ricardo.reier@ucavila.es (R.F.R.F.)

Abstract: Educational institutions are undergoing an internal process of strategic transformation to adapt to the challenges caused by the growing impact of digitization and the continuous development of student and labor market expectations. Consequently, it is essential to obtain more accurate knowledge of students to improve their learning experience and their relationship with the educational institution, and in this way also contribute to evolving those students’ skills that will be useful in their next professional future. For this to happen, the entire academic community faces obstacles related to data capture, analysis, and subsequent activation. This article establishes a methodology to design, from a business point of view, the application in educational environments of predictive machine learning models based on Artificial Intelligence (AI), focusing on the student and their experience when interacting physically and emotionally with the educational ecosystem. This methodology focuses on the educational offer, relying on a taxonomy based on learning objects to automate the construction of analytical models. This methodology serves as a motivating backdrop to several challenges facing educational institutions, such as the exciting crossroads of data fusion and the ethics of data use. Our ultimate goal is to encourage education experts and practitioners to take full advantage of applying this methodology to make data-driven decisions without any preconceived bias due to the lack of contrasting information.

Keywords: competences; higher education; machine learning

1. Introduction

Educational institutions are undergoing an internal process of strategic transformation to adapt to the challenges posed by the growing impact of digitalization and the continuous development of student and labor market expectations.

This global transformation is contextualized within a broader story of profound societal change (Eurofound 2018), the recent impact of the 2020 pandemic and emerging technology trends that requires the construction and execution of a digital strategy in education, clarity of vision and the ability to make timely decisions.

To cope with this transformation, educational institutions are focusing their efforts on specific lines of work. For example, the need to teach in almost any face-to-face or distance condition, the adoption of technologies focused on the digital management and assessment of learning, and the promotion of collaborative communication and work technologies (Interrupción y Respuesta Educativa 2021).

At the heart of this transformation is the student experience, understood as the total interactions between the student and the educational institutions, from the time of interest and admission to graduation and subsequent engagement as alumni.

Before the pandemic, the higher education landscape in Spain was relatively stable, with a mixed style of public and private education coexisting in Spain. According to the study CYD (2021), which measures and compares all universities by evaluating institutional...
performance and various fields of knowledge, with a scope of 77 Spanish universities and almost 3000 degrees, the pattern repeats throughout the country. Public universities stand out in aspects such as funds obtained and articles published and research, while private centers present better results in teaching, performance, or graduation rates.

The indicators used in (CYD (2021)) are structured in five dimensions: teaching and learning, research, knowledge transfer, international orientation, and contribution to regional development. The main elements of these evaluation axes are the student and the teaching staff.

According to Sheehan et al. (2020), university education has traditionally been linked to university experiences based on a high level of personal contact and relatively low use of technology. Only those institutions seeking to align and identify more closely with a growing wave of digitally literate students have evolved from such face-to-face models.

Most higher education institutions had approached the digital transformation with curiosity rather than a firm commitment (Yanckello 2021). Brand equity was key to stability, and in many cases, those brands dated back dozens of years. Furthermore, while there were pockets of good practice, most of the higher education sector had not been stressed and challenged to change beyond the competition of student recruitment on entry to university. The rate of digital transformation in higher education, therefore, lagged behind other sectors.

The COVID-19 pandemic has forced an initial response from the sector to the continuity of learning and has stimulated considerable interest in online learning models. The response has been extraordinary, and initiatives that had not been consolidated for years were implemented within days (UNESCO 2021). These online models are far from perfect, but creating the world’s largest online learning experiment has built a solid foundation for future changes to the sector. Face-to-face learning remains the hallmark of universities, but they need to think about longer-term aspirational directions and plausible futures while addressing short-term recovery and renewal has been reinforced (Yanckello et al. 2020).

Unfortunately, this survival-oriented and sometimes haphazard transition has, in many cases, provided poor experiences for college students (Diez-Gutierrez and Galardo-Espinoza 2020) and raised short-term costs.

This situation magnifies the need to optimize costs across the sector, integrating technologies at the heart of institutional strategy and resolutely addressing deep digital transformation processes (Las nuevas Tecnologías en el Desarrollo Académico Universitario 2019). Similarly, there is a need to reclaim excellence in student experiences, explore new learning designs, and better understand students’ real expectations and experiences at all contact points with this new learning environment.

This research work is framed around student experiences, with the development of a methodology for applying supervised predictive models based on AI, oriented towards the processing of data generated by student and university contacts, standardized in an organized information structure or taxonomy. This taxonomy is based on the virtualization of the student as an entity and the digitalization of interactions through the development of the student’s digital twin.

One of the unique events of this COVID-19 crisis has been the sudden shutdown of significant components of public life. The behavioral change imposed by the pandemic response is unprecedented in human history. Since that imposition, we have learned new behaviors such as how we buy things, how and where we work, and even how we interact with other people. We have also had to cope with imposed changes such as physical distancing, the use of masks or lockdowns (Lasa et al. 2020).

Moreover, this is not just individual behavior. These behaviors have reflected a sense of collective health and safety. Our new behaviors will not only affect individuals, but, if they persist, they will aggregate and amplify and eventually become encoded in our built environment, in regulations, in our business processes, and our social and cultural norms (Lasa et al. 2020).
In university education, student experience and education have always been seen as fundamental elements. They are at the heart of the model, alongside the advancement and generation of new knowledge itself. COVID-19 has also impacted the student experience, but this only reinforces its importance (WeWork and Brightspot Strategy 2021).

Universities develop their mission in a world where students’ expectations of brands are high. Students want consistent messages, information, and service when interacting with a brand, including university brands. They may not know whether they are receiving information from a marketing team, an admissions officer, or an IT manager informing them of an issue.

The most advanced universities are therefore looking at the student experience with their brand as a way of meeting expectations and providing personalized, seamless experiences that offer real value to students, moving away from the traditional approach based on selling a product (Morgan et al. 2021a).

Millions of students choose to study at university each year, which significantly impacts the rest of their lives. In an increasingly competitive higher education environment, many universities are increasingly prioritizing digital student interaction to increase the number of applications and increase their overall satisfaction (Yanckello et al. 2019). However, with the wide range of platforms and channels for student engagement, it can be challenging to assess whether universities’ efforts make a difference. Making sense of this digital landscape requires consideration, strategy, planning, and agility. It is clear which data are relevant sources of information and how to analyze and activate them to generate value.

This generation of digitally savvy students expects a sophisticated user experience in all their online interactions, and the university environment is no exception (Morgan et al. 2021a):

- Students can find countless university options through digital channels;
- Universities must provide the best teaching experience, also now in digital format;
- Students value data privacy;
- Immediate and simplified access to information is essential.

To meet these expectations, universities around the world are in the midst of a digital transformation facing significant internal challenges such as decentralized structures with varying levels of digital expertise, siloed faculties and departments with the capacity and autonomy to make decisions about the experiences they provide to students, and inconsistencies in internal workflows (Morgan et al. 2021b).

In this complex environment, digital student engagement takes many forms and encompasses all the touchpoints a university has with its students in the digital ecosystem (Lowendahl 2019). Essentially, it can be any element that involves online interaction, from opening an email to clicking on its links and requesting a training program’s content. It can include responses to students from student service and care or even maintaining perceptions of the university’s brand from the admissions or internship department, and because digital student engagement covers a wide terrain, analyzing and tracking these interactions can become a daunting task (Lowendahl 2019). Additionally, combining these digital contacts with more traditional contacts such as telephone or face-to-face contact is necessary. Therefore, it is needed and relevant to develop methodologies to automate these experiences and help determine which efforts on the part of the university are meaningful. Likewise, measurement is vital to understand how a university is positioned in terms of digital student engagement. Without it, there is no way of knowing what works and what does not (Reier Forradellas and Gallastegui 2021).

Universities also encourage student engagement. According to the study of Kuik (2020), optimizing the digital student experience is particularly relevant. Once digital data has been collected, analyzed, and measured, it should be activated and used to optimize the digital presence. This means updating web content and functions to provide relevant and targeted information and resources that meet students’ needs, with decisions based on and supported by data.
1.1. Opportunities in University Educational Institutions

For universities, an essential dimension of digital transformation is to leverage data to understand the processes students go through and their needs, ensuring a long-term improvement of the student experience.

The study by Sheehan et al. (2020) identifies opportunities in higher education institutions that give rise to four scenarios:

- **Consolidation**
  
  This is a short-term scenario with a little transformational impact. It is characterized by high competition in student recruitment, the creation and reconstruction of experiences, and the alignment of new educational models with the needs of teachers and students.

  As the pandemic becomes less critical for strategic planning, we will see a polarization between institutions that choose to duplicate their online models and those that seek a safe path back to campus.

  Those institutions with government funding and a strong reputation will seek a robust on-campus educational experience that remains attractive to students. Those with mature online strategies will leverage their pandemic experiences to push them further and gain market share.

  In general, a return to a traditional education model is being sought, but with some online variants being accelerated. Cost optimization is relevant, as is the mobilization of learning platforms and investments towards the construction of immersive experiences, the implementation of CRM, and resources for student retention. In this consolidation scenario, there is a risk of less student engagement with the university.

- **Standardization and adjustment**

  It is a long-term scenario with a significant capacity to transform the education sector. It is characterized by a reorientation of the higher education sector towards the skills demanded by business and the economic and industrial sectors. The pandemic has created significant damage to the economy, and industries are targeted in national recovery plans. Critical skills are identified to support these industries, and governments drive a sector-wide employability approach by aligning investment in higher education towards lifelong learning. Micro-credentials or specific quality training is making a great deal of sense, in many cases displacing undergraduate studies.

  Education is more online, on-demand, connected, and personalized, driven by technologies such as Artificial Intelligence. Flexible learning platforms capable of powering anytime, anywhere learning are needed. Learner analytics and AI are vital to creating advanced learner experiences.

- **Chaos management**

  This is a long-term scenario with little transformation of the education sector, where complete normality has not yet arrived. Both teachers and pupils comply with the rules but are uncomfortable with the experience.

  The sector will tend to polarize into elite institutions, with a strong and renewed research orientation and attracting students, but tends to preserve traditional teaching practice. The other extreme comprises highly industry-oriented universities with considerable financial restructuring and essential optimization of their operational costs.

- **Discovery of new paradigms**

  This is a short-term scenario with a significant capacity to transform the education sector. Some universities are evolving rapidly with alternative training pathways to meet the need for specific, skills-oriented training for flexible learners seeking on-demand and personalized training. Branding as an institution remains essential as long as the activity in this format is also of quality. This type of training leads to agreements and partnerships with content providers and companies specializing in this type of skill.
In any of these scenarios, the learner experience and the analytical capacities of institutions are essential to understand student behavior and outcomes better.

1.2. Opportunities AI Project Management

A project is a temporary effort to create a single product or service (AENOR 2013). Projects should be clearly distinguished from operations. Making a new road is a project. Maintaining a road in a roadworthy condition is part of maintenance operations. Operations end when the life of the product or service in question is over. The project ends when the product or service has been created and is usually delivered to the customer (Baccarini 1999).

Project management is the application of knowledge, skills, tools, and techniques to plan activities aimed at satisfying the needs and expectations of the stakeholders (Wang and Huang 2006) of a project (Silvius and Batenburg 2009). Among these stakeholders, we can find, in addition to the clients or users themselves, other actors such as managers, legal services of the company, security departments, marketing departments, regulatory bodies, legislators, governments, pressure groups, etc.

When we think of an AI project, regardless of the cognitive and physical elements it may require (Castillo 2009), these projects always have one thing in common: software. Software is the most common product that a project of these characteristics generates.

Several software methodologies and procedures can be used for any type of software project, from a virtual environment to accounting management software. However, the typology of projects related to AI and in particular, machine learning has a series of particularities that need to be reflected (Cheng et al. 2012):

- Firstly, because artificial intelligence, as a science, requires us to follow the scientific method (Rech and Althoff 2004). The scientific method presents a set of orderly steps, from hypothesis formulation to the analysis of results and planning of conclusions, which are mainly used to discover new knowledge in science. These steps involve an iterative process if the starting hypothesis is not correct;
- Secondly, machine learning projects are data-intensive. Traditionally in data mining, the CRISP-DM methodology is used, which stands for Cross Industry Standard Process for Data Mining. This methodology starts with understanding the business, understanding the data and its preparation, and continuing with the modeling process (Moine 2016). If the evaluation of the model is positive, it is deployed. If not, it is necessary to rethink the understanding of the business and start again in an iterative process where needs and opportunities that can be solved with an innovative approach can always be detected.

The data science ecosystem brings a new approach. The generation of models capable of obtaining patterns from the data and making predictions becomes a central part of the scheme. Just as the visualization and communication of results acquire particular relevance compared to other more generalist models, an essential difference between modeling and a traditional software project is that once the model is made available to users, it may be necessary to evaluate its behavior and adjust its parameters periodically. The model users may change their habits, and the solution itself may induce behaviors that were not foreseen at the beginning.

1.3. Taxonomies

The use of taxonomies allows us to reflect on the conceptual model and the relationships between the elements. In this research, we will use them as a means of reflecting student satisfaction.

Satisfaction refers to the evaluation of service and considers cognitive, affective, and attitudinal factors (Petruzzellis et al. 2006). High satisfaction leads to elements of loyalty and a positive effect (Baumann et al. 2012), leading to recommendations and higher enrolment and retention rates (Abdelmaaboud et al. 2020).
Although satisfaction varies across disciplines (Green et al. 2015), Butt and Rehman’s (Butt and Rehman 2010) study specifically identifies (Figure 1) teacher experience and knowledge, course and subject offerings, learning environment, and classroom/campus facilities as having the most significant impact on student satisfaction.

![Figure 1. Elements impacting student satisfaction. Source: (Butt and Rehman 2010).](image)

Other authors complement these satisfaction factors with new elements such as the teaching style of the lecturer (Oldfield and Baron 2000), assessment tests and workload (Ginns et al. 2007), or affective interactions between the university and the student (Malouff et al. 2010).

Other complementary studies related to satisfaction in learning also include the interests, needs, and motivations of students, the practice of what they have learned in contextual situations, the sense of initiative, or the development of collaborative and cooperative interpersonal relationships (Villarreal-Villa et al. 2019; Silva Quiroz and Castillo 2017; García and Gago 2019).

This research study focuses on the dimension of learning provision as an element of learner satisfaction and relies explicitly on a taxonomy based on learning objects (Cuervo 2011; Castrillón 2011; Sánchez Medina 2014).

This taxonomy based on learning objects is in line with the improvement of the learner experience in the dimension of learning provision. It is a framework that helps us to identify the relevance of a data model and its availability.

To adapt personalized content to the needs of learners, especially in asynchronous environments, learning objects or information resources (an image, a page of text, a slide, an interactive simulation, a complete course) need to be reusable and assembled in different contexts.

These components and their use present a series of characteristics (Figure 2) such as the capacity to be digitally identified, show a granularity level that allows their reuse in different fields, and be self-contained, not depending on other objects be related.

Every learning object has external and internal components (Figure 3). The external components (Figure 4) are based on metadata compatible with the Learning Object Metadata (LOM) data model. The internal components (Figure 5) can be content, learning activities, and contextualization elements. In addition, every learning object is also assigned an external information structure (metadata) to facilitate its storage, identification, and retrieval.
Every learning object has external and internal components (Figure 3). The external components (Figure 4) are based on metadata compatible with the Learning Object Metadata (LOM) data model. The internal components (Figure 5) can be content, learning activities, and contextualization elements. In addition, every learning object is also assigned an external information structure (metadata) to facilitate its storage, identification, and retrieval.
Figure 4. Taxonomy of learning objects. External components. Fuente: Basado en Cuervo (2011); Castrillón (2011); Sánchez Medina (2014).

Figure 5. Taxonomy of learning objects. Internal components. Source: based on Cuervo (2011); Castrillón (2011); Sánchez Medina (2014).
Taxonomies contain both the terms of interest and the set of rules that establish how the words relate to each other and tend to be tree-like with a hierarchical structure.

In education, another well-known taxonomy is Bloom’s Taxonomy (Bloom 1956), which is a framework for setting learning objectives and skills that students should master. It is a well-known theory in education that is suitable for assessing the cognitive level acquired in a subject and includes knowledge, comprehension, application, analysis, synthesis, and evaluation.

This learning method is organized into levels:
- Remember: memorizing and storing data and concepts;
- Understand: enabling knowledge to be explained and discussed;
- Apply: putting into practice what has been learned and being able to demonstrate a hypothesis;
- Analyze: contrasting the information and content used;
- Evaluate: assessing what has been learned with a critical approach;
- Create: generating new contributions based on their ability to construct and generate original knowledge. This is the highest level of the hierarchy to which a student should aspire.

Bloom’s Taxonomy and the revised version by Anderson and Krathwohl (2001) have been used extensively in different educational settings, such as children’s art education (Rodríguez 2019). They have generated various versions and adaptations, such as Churches (2009) study that evolves traditional classroom practices towards new technology-related environments.

1.4. Digital Twins

Digital twins are digital replicas of products, services, or processes that are already in use, especially in the field of Industry 4.0 (O’Sullivan et al. 2020). According to Parris et al., a digital twin is a living model that drives a business outcome. The idea is to digitally replicate an actual physical (single or composite) element to synchronize with it perfectly. This capability allows for an increase in the possibilities of analysis on the digital twin, understanding its behavior in different scenarios, and provides valuable information leading to improvement and corrective actions (i-SCOOP 2021).

In education, digital twins are a new tool that contribute to better and faster learning of simulated environments. Instead of studying the real thing, one can learn its digital representation. Here technologies such as virtual reality provide new tools that amplify the use of digital twins.

In perspective, we should be able to aggregate fragments of scattered information into a more complete, more accurate representation of our “self”. A more accurate picture of such a digital mirror would be achieved, representing our knowledge and skills and our fading knowledge (forgetting) and skills (what we lose by not practicing), which could become the starting point for a proactive education program.

Educational institutions could mirror their students based on their digital twins, modulating and balancing knowledge, education, and exchanging interactions and experiences based on the evolution of these digital representations and the analysis performed by intelligent models based on Artificial Intelligence.

2. Materials and Methods

The main objective of this work is to develop a methodology for the pre-validation, from a business point of view, of projects based on supervised Machine Learning models in the field of university students’ experiences.

In addition, it develops the following specific objectives:
- Design a process map that defines the relationships between the student and a university in order to apply the methodology in the specific use case of satisfaction with learning;
2.1. Context for Technological Development of an AI Model

Development of AI capabilities initially involves a whole set of tasks for obtaining, exploring, and preparing data and the parameterization and optimization of algorithms, and the validation of the results obtained. Subsequently, it is necessary to make this AI capacity available to ensure that it is available for use by an application regularly and is applicable to new data. Finally, it is necessary to monitor the application in this phase so that its efficiency does not decline over time and lose its effectiveness.

In AI projects, the data domain is the most relevant to consider. It is necessary to evaluate the amount and distribution of available data, as a lack of data is a reason that can make an AI project unfeasible. The same applies if there is a significant imbalance between the number of positive and negative samples. Although this may be a representative characteristic of the reality to be analyzed, it presents substantial modeling challenges. In addition, consistency must be considered between the data used for training the model and the data on which the model is applied. This consistency must consider factors such as the intrinsic characteristics of the dataset and the transformations applied to the data. When an entity develops a project in which it trains a model to apply to its data, it is relatively straightforward to maintain consistency. The training data comes from the same source as the application data.

The presence of bias in the data is also a very relevant aspect. This is a factor that must be taken into account and minimized by taking appropriate control measures.

In summary, the development and technical deployment of a model based on AI (Figure 6) is a sometimes complex, resource-consuming process that, based on the scientific method, does not always ensure the success of its objective due to influential aspects such as the type of data available or used, their consistency and treatment, the characteristics of the data used to train the underlying model, their distribution, and possible inherent bias. Additionally, once the model is available, the type of deployment of this capability and continuous monitoring of its performance and maintenance are relevant.

![Figure 6. Methodology for the technological development of an AI Model. Source: own elaboration.](image)

However, the typical cost of developing AI models is generally framed in contexts that do not belong to a technological approach, but rather to a point of view of business design and the need to validate the business results obtained.

2.2. Context for Business Design of IA Models

Therefore, starting from this need, the determination to transform a business problem from a business perspective into an analytical/technical problem to be solved by AI requires prior design decisions regarding the business objective itself, which will subsequently impact the development of the AI model (Figure 7).
Consequently, the decisions to be followed in the roadmap for the 360° construction of a specific AI application (Figure 8) must be the result of a set of prior design decisions and approaches to the business problems, which constitute the basis of the business, to subsequently proceed to its development with traditional methodologies (Agile, CRISP-DM, DevOps...).

The methodology for the business design of predictive machine learning models developed in this research is framed in the previous context of business understanding, preliminary considerations and design principles of Figure 8. Its objective is to establish a methodological framework to understand and provide a solution to the real needs in an environment of student relations with the university, which will be the basis for the subsequent development of models based on Artificial Intelligence.

This methodology incorporates practices such as the following:

- Creating AI hypotheses and assess which ones provide the most value in a prototype design;
- Identifying potential AI capabilities that support the required functionality;
- Rationalizing what capabilities already exist in off-the-shelf solutions or whether they can be sourced from public assets (open source) or must be developed in-house. The sourcing model decision impacts the cost and availability of AI models. Generally, more strategic functionalities require less commoditized AI capabilities;
- Determining the incorporation of such model functionalities in specific development cycles based on prioritized capability planning and cost-benefit balance;
• Giving personality to the cognitive system, which will modulate how the system will respond, how it will adapt its responses to the different contexts (place and time of the interaction, type of user...), the style and tone of the language (formal, informal...);
• Considering all data sources, even if they seem absurd. Start with the most obvious ones and extend considerations to public data sources, unstructured and non-standard data types (images, videos, phone calls...), and then filter and prioritize.

2.3. Methodology for the Prevalidation from a Business Point of View of Projects Based on Supervised Machine Learning Models in the Field of University Students’ Experiences

The application of this methodology for the business design of predictive machine learning models (Figure 9) enables the design and construction of specific business scenarios for the subsequent construction of supervised learning AI models based on students’ experiences with the university. It also provides a systematic and scalable framework for consideration of other AI use cases and models at the university level.

Figure 9. Detailed Methodology Business Design IA Models. Source: own elaboration.

The detailed application of this methodology for the Business Design of IA models in the environment of university entities is carried out with the execution of the following phases and vectors of analysis:

1. **Identify the business opportunity before data and models**

   First, business objectives must be defined as a preliminary step before even thinking about data and machine learning models. Starting with data means taking a significant risk because the simple exploration of data always provides interesting aspects by its very nature. However, the search for insights without a clear objective implies going down a line of work without a horizon.

   University business opportunities are on a two-speed track. In the short term, tactically, they are looking to return to a traditional education model but accelerating some online variants and cost optimization. In the medium term, work is being undertaken to reorient the sector towards the skills demanded by companies through a continuous learning approach and with micro-credentials or specific quality training complementary to the degree, developed autonomously or with agreements and alliances with content providers and specialized companies.
Each university should know the profile of its students and determine which are its strengths as an institution and which need to be reinforced. Consequently, the points of contact and participation to be measured should be correlated with them.

For example:

- Retention and graduation rates;
- Program completion times;
- Monitoring of academic performance;
- Registration for information sessions or campus tours;
- Personalized academic contacts with the student;
- Content and marketing materials that attract and engage learners more effectively than others;
- Patterns of student participation: how long it takes a student to complete a full interaction with the university, barriers to continuous student participation, etc.

2. **Build the business case**

   Develop a business case for the project. The objective is to anticipate the cost–benefit ratio involved in creating, implementing, and deploying the model, serving as a basis for refining the scope, functional requirements, and appropriate capabilities to make the model viable. At this stage, activities are also prioritized, and orderly planning of initiatives aligned to the business case is established.

   To identify the **cost–benefit ratio of creating, implementing, and deploying the IA model**, the specific use case to be developed must be determined:

   - The selection of the use case is first made based on the design of the university’s process map on which to fit all student contacts and possible use cases based on these interactions.

   The design of the student contact and experience map covers the student’s extended relationship with the university, from pre-entry with the application through to graduation and on to the alumni relationship. It includes the main stages of the customer journey, such as awareness and consideration, assessment and application, post-acceptance and alumni care, and alumni service.

   This digital customer journey survey of current and prospective students shows the various paths they may take when exploring the university’s digital environments. It also identifies the various phases of the student lifecycle and how they interact with your university during each phase.

   After identifying each phase, for example, relevant content will be created for each stage, targeting the crucial moments when learners need support. Content creation includes brand consistency, applying the right tone, and keeping it up to date and consistent.

   In this way, the university can address weaknesses in student relations and deepen engagement.

   - Secondly, all potential use cases that can be addressed from an AI point of view are placed on the process map and prioritized and selected. This selection of the specific use case allows refining the scope, functional requirements, and appropriate capabilities to make the AI model viable.

   In order to identify the use cases, it is necessary to rely on a university process map that reflects the iterations with the students. The value chain of a university institution (Figure 10) serves its core student-centric mission and is the perfect framework for sorting out the possible use cases in which Machine Learning-based learning models can be applied.
In an extended way, this value chain of a university educational institution begins even before the student becomes a student, with the market analysis and discovery phases. It continues until the student ceases to be a student and becomes alumni, thus continuing the customer journey of relations with the university.

All potential use cases reflected in the graph above should be prioritized based on their overall importance for the organization and not only from a departmental point of view, but also marketing, admissions... Other factors such as complexity, cost, and execution time should also be taken into account.

Based on the above criteria and based on the opportunities in educational institutions developed in the Gartner (2021) study, the use case associated with student satisfaction in the learning phase has been selected as it is the opportunity present in all four scenarios developed by Gartner (2021) and is considered especially relevant in environments where high impact transformation is required.

The deployment of this use case implies some requirements that must be considered in the cost–benefit ratio of the deployment of the associated IA model. These requirements are as follows:

- Big Data infrastructure (data collection, storage, and processing), with a planned and sustainable architecture roadmap;
- The use of tools for data exploration, data integration, and advanced real-time analytics will be commonplace;
- A Data Governance policy will be in place and systematically applied for the development of advanced use cases with analytical, predictive, and prescriptive models;
- Relevant structured and unstructured data and information sources shall exist on a dedicated, flexible, robust, and scalable technological infrastructure in cloud/on-premises environments;
- In the case of deployment of cloud solutions, these will tend to be initially hybrid infrastructures, with a tendency to migrate to multi-cloud (hyperscale) solutions.

In short, the importance of the cloud in the university’s IT strategy should be high and relevant, with a policy of service management and development of IA models linked to Big Data and Cloud environments.

Identifying the data sources needed to build the use cases involves assessing the effort required to operationalize them and then include them in the technical design of the
AI models. The actual treatment of the unstructured data present in these environments measures the real effort of their inclusion in the technical AI models.

In a digital-only environment, universities can engage students through various channels, including emails, chatbots, text messaging, and mobile apps. Initially, efforts can be concentrated on two key channels: the web, and social media platforms.

- **Website**: the design and interface of the website are critical in determining how students perceive the university. Indeed, a university’s website is the ultimate brand statement, an important component of the student experience, and can greatly influence a student’s decision to apply as a student of the university or subsequent attractiveness as a student.

- Therefore, the web should be analyzed from a student’s perspective to see how they feel about using the platform as a tool for participation and work on aspects that improve it to achieve an attractive and easy-to-navigate information environment.

- The importance of improving the student’s digital experience is decisive, as digital aspects dominate their decision-making to a large extent in the new generations of students. Some of the lines of work are based on the following pillars:
  - Content quality control, thorough content inventory, and problem reporting. High-quality content is critical to providing students, staff, alumni, and the general public with the information they need online, and doing so with a frictionless experience is critical.
  - Improving the accessibility of their web portals to remove all barriers that prevent students from accessing essential information online.
  - Prioritization of web sections and digital content, based on the analysis of student behavior and interaction and the identification of actions that have the greatest impact on improving the user experience.

Social networks play a very important role in students’ perception of universities. Institutions now make greater use of social media platforms such as Facebook, Twitter, YouTube, and Instagram to market their programs and interact with students. For prospective students, the way a university responds to their questions or comments makes a difference.

3. **Define actionable models**

How we will apply the model, how we will use it in the organization, and whether it will affect the way people work. These are premises and approaches that involve reflecting on the impact of these models in real organizations and anticipating the definition of the necessary actions for their implementation.

The implementation of the model requires the unconditional support of the university’s business areas and academic directors. Especially when it comes to student satisfaction, it is a strategic issue and therefore must be led from the highest level.

The data needed to feed the model come from different areas, so the cross-cutting use of the data and its availability requires a Data Governance Model, which is not usually present, as well as a cross-cutting data manager.

The technical construction of learner impact models is strongly dependent on the business, and their activation and performance monitoring are also the responsibility of the business, so their design and development must be strongly supported and led by the business to be actionable.

4. **Consider Machine Learning as an experimental science**

Machine Learning is an experimental science that implies feedback in the model creation and evaluation phases, including data preparation, based on continuous monitoring of the continuously obtained results in the model implementation phase. Therefore, it is necessary to monitor deviations in the ROI of the project (Figure 11). This continuous
monitoring action is critical as it allows us to evaluate how useful the model is for the business.

![Methodology of IA models](image)

**Figure 11.** ROI feedback via IA Model results monitoring. Source: own elaboration.

Reviewing AI models is a key activity to optimize their ongoing performance. However, Machine Learning reinforced learning models being continuously optimizable can deliver value and results that increase student knowledge already at early stages. This should guide Universities to compare planned vs. actual ROI at all times and shape a true picture of the model’s usefulness from the very first iterations of the AI model.

5. **Evaluate the cost of false positives or false negatives**

Both false positives and false negatives are incorrect predictions but have different costs for organizations. An aggressive business case will prefer to minimize false negatives because opportunities are lost depending on the business strategy. In contrast, a conservative business case will choose to minimize false positives because money is lost.

In student recruitment, the choice will be to minimize false negatives. In terms of satisfaction, the option is to minimize false positives of dissatisfaction to focus on the dissatisfied students and not lose focus and waste resources.

6. **Find the right balance between false positives and false negatives**

Obtaining the break-even point between the two is a fine-tuning process as decreasing false positives implies increasing false negatives, and vice versa. The model with the best ratio for the business strategy must be found and is a relevant input for the business case.

Dealing with dissatisfied learners always requires personalized attention. Action plans aim to eliminate or alleviate the root causes of dissatisfaction, but AI models also allow us to identify dissatisfied students individually. The balance of false positives and false negatives is determined by the university’s ability to address dissatisfied students in a personalized way.

7. **Transform regression models into classification models**

The analysis of models and decision trees, looking for the reasons for the predictions and the relevance of the different variables, is key to be able to orientate the IA models towards classification models. This analysis of variables is fundamental to build and apply rules and conditions for categorical decisions.

Predictive models provide a numerical value (e.g., sales to be made in a month, the value of a property for sale, or the number of students applying to university). However, this numerical prediction can be transformed into a categorical variable and therefore evolve into a classification model which, in general, is easier to operate and with which better results are obtained in machine learning. Moreover, in many business opportunities, a Boolean decision is needed as the decisions made are also categorical (yes/no, segment A, B, C, learner profile X, Y). Therefore, it is advisable to reflect on whether we can apply rules and conditions that transform the characteristics of the predictions through data engineering.
8. **Understand data from a business perspective**

Machine learning models identify patterns and trends in data, looking for correlations in the data. However, they do not understand the meaning of the data. The model design methodology must identify whether the available data are relevant and reliable. Relevance is understood as the characteristic that defines whether the data are relative to the objective that the model seeks to predict. Reliability is an indicator that measures whether the data are correct. At this stage, the suitability of normalization, standardization, identification of outliers, or the proportionality of the available data can be identified.

The relevance and reliability of the data are determined by business judgment. In this phase, it is necessary to define and use a data taxonomy adjusted to solve the business problem. It is a reference framework that helps us identify the relevance of the data model and its availability. It is also necessary to identify the need or complementarity of the available information sources with other internal or external data sources. The task associated with the data engineering involved in creating the models is an enormously sensitive task on which the viability of the predictions depends to a large extent. Therefore, a prior reflection of the data from a business point of view is very critical.

On the other hand, compliance with the standards determined by data privacy laws obliges universities to take precautions about student data and ensure that they are treated appropriately. Data processing itself is already performed in the AI development methodologies. Still, in this design phase of the AI Models, transversal actions are conceived that go beyond the specific scope of their processing in a particular model:

- Establish privacy policies, standards, and secure data collection processes across the university;
- Include the ethical component in the Data Governance Model;
- Comply with applicable regional, national and international data privacy laws;
- Develop and implement data privacy training programs;
- Respond promptly to breaches or privacy incidents and be prepared for them;
- A catalog that has access to student data and collects data transparently;
- Review data privacy policies and procedures regularly;
- Create an action plan to improve them periodically: data privacy laws are constantly changing.

Data are the basis for decision-making, and therefore the understanding of the results and their sensitivity and potential impact must be guided by the active participation of the business. Professors, academic directors, area directors must be part of the day-to-day running of the models. In this phase of the model, the objectives pursued and the business relevance of the different categories of data must be clear. As an example of the importance of data, some business considerations to develop would be the following:

- The data generated by digital platforms contain valuable information on what is working, what is not working, and areas for optimization. Access to student engagement data can help the university communicate with students and, for example, trigger an automated chatbot response that provides meaningful information (such as a link to a specific degree program) when the student needs it.
- Another relevant use case could be the reduction in university dropouts during the first year. By collecting student data from the LMS (Learning Management System), such as grades and attendance records, and combining it with their demographic data, a student dropout profile can be constructed, and the stages in their university career at which they typically drop out can be determined.
- With this pattern, it will be possible to target communication efforts towards those students who fit this profile at the most at-risk stages of their university career and thus help them to orientate themselves appropriately towards graduation.
- Data are only useful if they are accurate. Therefore, use only clean data that have been validated, especially concerning the following two dimensions:
- Student data: collected from digital interactions, such as clicks and website visits, and then analyzed to determine how and when to interact to improve the student’s digital experience;
- Student engagement: student data enables relevant communications at the right time, in the right place, and on the right platforms.

The student’s digital twin is susceptible to collect a wide range of data categories. Of all these categories, those related to the construction of the learner’s experience have been selected for the reasons explained above, and among the wide variety of factors that influence the learner experience, those that specifically refer to the dimension of training offerings or educational content (Butt and Rehman 2010).

The application of the taxonomy based on learning objects contributes to the student’s learning, as this does not depend so much on the modality or grade in which they are, but on the presentation of the thematic contents and the activities that reinforce their learning, both elements contemplated in the taxonomy. Moreover, the coherence of the learning objects with the needs of the student and the educational objectives present in the subject guides are reflected in the very definition of the characteristics of the educational objects in the taxonomy.

Additionally, AI models pose ethical challenges that must be resolved in advance and be guided by deep reflections of the Universities defined in their ethical principles in the use of AI. These are, again, business-driven decisions and should not be left to free development on the technical side. Data processing, within the framework of the privacy policies of each university, must comply with the principles of legality, impartiality, and transparency. Their use must be guided by an approach of minimized use only for the purpose for which they were collected and processed with accuracy, integrity, and confidentiality, proceeding to their anonymous or aggregated use where sensitive data cannot be used on an individual basis.

Consequently, the taxonomy of educational objects should be expanded (Figure 12) to include not only compliance with current regulations on data processing but also considering their ethical use.

![Extended taxonomy of learning objects. Characteristics of data. Source: own elaboration.](image)

Beyond educational objects, in the context of data, bias is an inherent feature, and often universities themselves, as well as companies, are not aware that it leads to discriminatory and unfair results. The university must ensure that there are no discriminatory elements when it is the AI that learns and the algorithms that suggest decisions. This rule is developed based on the next stage of the methodology.

9. **Monitor the impact of the IA Model’s performance on business objectives.**

Typical methodologies in Machine Learning model development already include continuous monitoring of the models, which involves feedback from the model creation and evaluation phases.
This study proposes to also extend the scope of feedback to this Model Design Methodology, as it can be substantially involved, changing the objective of some of its phases or even affecting the business case.

The feedback of AI Models must go beyond their technical construction because depending on their impact, they can alter the scope of the business objectives and the business case itself. This is why this Business Design Methodology includes the performance monitoring phase of the model.

The university must decide in this feedback if it should change the objective of some of its phases or even start again with a completely different approach. Additionally, to track the relevant metrics that each university has identified in the business opportunity phase, a measurement plan should be created to see if the efforts are paying off.

A measurement plan categorizes the metrics to be tracked, on which platforms they are tracked, and when and how success will be measured for each metric. It is important to establish benchmarks throughout the execution of the IA Model so that they can be continuously improved. As an additional benefit, the measurement plan can align university teams and practitioners towards the same goal, sharing the lines of work to achieve them.

An additional aspect to consider when setting business objectives that are then translated into the measurement plan is the identification and designation of responsibility for the university’s digital strategy. Many universities are not clear about which department is ultimately responsible for digital content, digital asset mapping, and establishing digital governance processes. It is also often not sufficiently clear who is responsible for establishing digital business indicators that are transparent to internal stakeholders or for analyzing the positive and negative impacts that contribute to their achievement.

10. **Change management**

Machine Learning supervised learning models are only one piece in the model. It must be integrated in terms of data architecture, data engineering, software lifecycle, training in its use, new methods and procedures that affect the day-to-day, considerations in the capture and privacy of data, alignment in the moments of retraining of the models to speed up the availability of the data... These disruptive technologies in the field of AI (Machine Learning, Neural Networks, Natural Language Processing,...) can transform how tasks have traditionally been performed, and therefore the management of organizational change is present as an essential element of the methodology.

These AI-based models can transform the way tasks have traditionally been performed. Decision-making is data-driven, and a culture of learning and growth is fostered. These models require a flexible and agile organization, open to continuous feedback based on results, and working collaboratively both internally and with open collaborators. Innovation ecosystems emerge and contribute to engaged teams relentlessly pursuing business impact.

Therefore, organizational change management is an important element of this methodology as it helps people to adjust value propositions by making them more personalized for the learner, more complete, and more transparent, based on the results of the IA models.

### 3. Results

This research work develops a methodology for pre-validation from a business point of view of projects based on supervised Machine Learning models in the field of student experiences. It provides a business methodology composed of 10 stages designed and adapted specifically to the characteristics of AI models, their operational performance, and their transformative impact on the business incorporating in its design the identification of the business opportunity, the construction of the business case, and the need to develop models that are actionable in universities. The methodology also incorporates in its design the premise that AI is an experimental science and the continuous need to validate the starting hypotheses and model feedback.

The application of this methodology (Figure 9) enables the design and construction of specific business scenarios for the subsequent construction of supervised learning AI
models based on students’ experiences with the university. It also provides a systematic and scalable framework for consideration of other AI use cases and models at the university level and ensures that the validation of sensitivity and error of Machine Learning supervised learning models, such as the evaluation of the cost to the business of false positives and negatives, as well as the need to find a balance in them, is taken into account. It also provides context-optimized results by addressing the suitability of transforming regression models into classification models.

These methodology and framework accelerate the opportunities identified by Sheehan et al. (2020) and aligns with the sophisticated digital user experiences demanded by learners (Morgan et al. 2021b).

Likewise, as complementary objectives, the process map of the relationship between the university and the students is designed in an extended relationship cycle, from the application consideration stage to the alumni relationship stage, which allows the application of supervised predictive models and where different use cases have been located that are susceptible to applying AI-based models.

The use of the taxonomy based on learning objects (Cuervo 2011; Castrillón 2011; Sánchez Medina 2014) allows us to reflect on the conceptual model and the relationships between the elements. This is why it is the vehicle used in this methodology of business design of IA models to standardize the data to be used and accurately reflect the variables related to the training offer that influence learner satisfaction (Butt and Rehman 2010). This taxonomy of learning objects has been extended in its characteristics so that it can consider and comply with the principles of legality, fairness, and transparency. Their use should be guided by an approach of minimized use only for the purpose for which they were procured and processed with accuracy, integrity, and confidentiality, proceeding to anonymous or aggregated use where sensitive data cannot be used on an individual basis.

Change management is not an element that is usually incorporated in AI project management (Project Management Institute 2019), as its usual orientation is that of sequential execution of tasks with an end goal at the end of the project. However, due to the experimental nature of AI Models and the impact they have on the organizations where they are deployed, this AI Models Business Design Methodology includes a change management stage as a necessary element as it helps people to continuously adjust the value propositions making them more personalized for the learner, more complete and more transparent, and above all, because AI models can transform the way decisions have traditionally been made, fostering a culture of continuous learning and adaptation (Reier Forradellas and Gallastegui 2021).

4. Discussion

The development of AI based models has been a practice that usually resembles traditional project development based on software lifecycle management (analysis, design, construction, testing, monitoring) and where waterfall, spiral, or agile lifecycle implementation methodologies are applied.

These methodologies are not sufficiently adapted to the scientific method on which AI research is based, as they do not include the definition and design criteria from a business point of view, which are decisive for their influence on the sensitivity of the models to be built and on the definition of what is expected in terms of their performance. The business design methodology for the AI models developed, from the identification of the business opportunity to the evaluation of the cost of false positives and negatives, incorporates a series of stages whose objective is to determine, directly and simply, those design principles that are relevant to the business and that condition the subsequent development of the AI models.

Additionally, the construction of AI capabilities requires intensive use of data and, consequently, the typology of the data, its consistency, quantity, and distribution, among other factors, largely conditions the evolution of the models. In this sense, methodologies such as CRISP-DM used in statistical and data mining models (Moine 2016), are a reference
for the development of Machine Learning models, including the analysis of available data, modeling, and validation of the results.

The CRISP-DM methodology includes an initial phase that focuses on understanding the objectives and requirements of the project from a business perspective and how it directly affects the data. However, this stage is based on the understanding of the data and the impact on the data as a driver for the data-driven models and analyses that have already been decided to be developed and lacks influence on the business in the sense of modifying the project objectives or even modifying the business case. The business design methodology of the proposed IA models includes this necessary business transcendence in a much more natural environment where the discussion is about business, and the technological development of the Model where CRISP-DM is framed has not yet been addressed.

This work does incorporate the business point of view intrinsically in the development of supervised learning models in Machine Learning and provides a methodology whose design is oriented to identify opportunities and contribute to the construction of the business case.

5. Conclusions

Educational institutions are undergoing an internal process of strategic transformation to adapt to the challenges posed by the growing impact of digitization and the continuous development of student and labor market expectations. This need has been accelerated by the impact of the 2020 pandemic and emerging technology trends that require the construction and execution of a digital strategy in education, clarity of vision, and the ability to make timely decisions.

The student experience is a vital element for universities working to improve the scope of faculty knowledge, expertise, and research work and to improve the range of courses, learning environments, and facilities available to them on campus and in classrooms.

This student experience is vital as a driving element for the development of basic skills/soft skills that are considered as a fundamental element of learning and with even greater value than the training contents themselves.

The construction of AI models to predict and improve student satisfaction allows for a better understanding of student satisfaction patterns and the factors that influence them in a personalized way. However, this approach to models should not only be made from a technological point of view of building and developing them, but it is also key to approaching them from a previous business vision.

In the development of the methodology, the taxonomy based on learning objects is applied to the learner’s digital twin, and improvements to it are included, taking into account compliance with the principles of legality, impartiality, and transparency. Its use is guided by ethical principles of accuracy, integrity, and confidentiality, proceeding to its anonymous or aggregate use where sensitive data cannot be used on an individual basis.

The business design methodology of AI models incorporates a change management stage as a necessary element, as AI-based technologies can transform the way decisions have traditionally been made, fostering a culture of continuous learning and adaptation.

As lines of work for the future, the following are identified:

(1) Expansion of the use cases for the application of the methodology to the student lead funnel. Some additional cases could be the profiling of potential students to classify leads into homogeneous groups according to their intrinsic characteristics, or the prioritization of potential students to optimize and accelerate the conversion of students who have shown interest in joining the university and have applied but have not formalized their enrolment;

(2) Expansion of the methodology based on the development of a complete Learning Analytics ecosystem, which allows learning from the key factors for the courses, accompanying students, obtaining early warnings of abandonment or lack of in-
terest, understanding how students interact, how to improve pedagogically, being differential as a university in its core business;

3. Incorporating data from users who are not yet university students but potential students (discovery stage) and analyzing the implications in the field of data processing;

4. Development of new taxonomies applied to continuous training environments, micro-studies, and courses for adaptation to the labor market, given at university level but without the consideration of a degree/master’s degree;

5. Development of non-university scenarios such as early childhood education, primary education, and continuous training environments in companies and self-learning environments.

Author Contributions: Investigation, L.M.G.G. and R.F.R.F. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References


Las nuevas Tecnologías en el Desarrollo Académico Universitario. 2019. Inter-Cambios. Dilemas y Transiciones de la Educación Superior, 4. [CrossRef]


Rodríguez, Martín Caero. 2019. Recreando la taxonomía de Bloom para niños artistas. Hacia una educación artística metacognitiva, metaemotiva y metaafectiva. Artseduca 24: 65–84. [CrossRef]

Sánchez Medina, Irlesa Indira. 2014. Estado del arte de las metodologías y modelos de los Objetos Virtuales de Aprendizaje (OVAS) en Colombia. Entornos 28: 93. [CrossRef]


