electronics

Article

Designing an Intelligent Virtual Educational System to Improve the Efficiency of Primary Education in Developing Countries

Vidal Alonso-Secades [1,*], Alfonso-José López-Rivero [1], Manuel Martín-Merino-Acera [1], Manuel-José Ruiz-García [2] and Olga Arranz-García [3]

[1] Computer Science Faculty, Pontifical University of Salamanca, C/Compañia, 5, 37002 Salamanca, Spain; ajlopezri@upsa.es (A.-J.L.-R.); mmartinmac@upsa.es (M.M.-M.-A.)
[2] ProFuturo Foundation, C/Gran Vía 28, 28013 Madrid, Spain; manueljose.ruizgarcia@telefonica.com
[3] Education Faculty, Pontifical University of Salamanca, C/Compañia, 5, 37002 Salamanca, Spain; oarranzga@upsa.es
* Correspondence: valonsose@upsa.es; Tel.: +34-923277119 (ext. 7676)

Abstract: Incorporating technology into virtual education encourages educational institutions to demand a migration from the current learning management system towards an intelligent virtual educational system, seeking greater benefit by exploiting the data generated by students in their day-to-day activities. Therefore, the design of these intelligent systems must be performed from a new perspective, which will take advantage of the new analytical functions provided by technologies such as artificial intelligence, big data, educational data mining techniques, and web analytics. This paper focuses on primary education in developing countries, showing the design of an intelligent virtual educational system to improve the efficiency of primary education through recommendations based on reliable data. The intelligent system is formed of four subsystems: data warehousing, analytical data processing, monitoring process and recommender system for educational agents. To illustrate this, the paper contains two dashboards that analyze, respectively, the digital resources usage time and an aggregate profile of teachers' digital skills, in order to infer new activities that improve efficiency. These intelligent virtual educational systems focus the teaching–learning process on new forms of interaction on an educational future oriented to personalized teaching for the students, and new evaluation and teaching processes for each professor.

Keywords: intelligent systems; virtual education; educational data mining; data cleaning; machine learning; learning analytics; dashboard; web analytics

check for updates

Citation: Alonso-Secades, V.; López-Rivero, A.-J.; Martín-Merino-Acera, M.; Ruiz-García, M.-J.; Arranz-García, O. Designing an Intelligent Virtual Educational System to Improve the Efficiency of Primary Education in Developing Countries. *Electronics* **2022**, *11*, 1487. https://doi.org/10.3390/electronics11091487

Academic Editors: Claudio Savaglio, María-Luisa Pérez-Delgado, Roberto García Martín, José Escuadra Burrieza and Jesús Ángel Román Gallego

Received: 31 March 2022
Accepted: 2 May 2022
Published: 6 May 2022

1. Introduction

The incorporation of technology in different social spheres is bringing about relevant changes to improve current processes. It is a fact that education is one of those areas where technology has been introduced to improve educational processes and achieve greater effectiveness in the teaching–learning process [1].

Augmented reality, robotics, artificial intelligence, big data, machine learning, and the Internet of Things (IoT) are some examples of technologies that are, in fact, already acting in the educational process, allowing new forms of student–teacher interaction to be established, creating novel assessment processes, or providing analyzed information that is relevant to educational decision-making [2,3].

One of the fields that has benefited most from the availability of new technologies is the field of virtual education. Taking advantage of new forms of communication, education has seen an opportunity for expansion and has turned to the implementation and use of virtual educational platforms that enable the distribution of knowledge beyond the location of an educational institution.

The use of intelligent systems in virtual education is not a new concept. As early as 2010, research was underway to incorporate intelligent systems into education through virtual reality [4]. However, the focus of the educational aspects of intelligent systems is now on many more factors, such as the students and their use of the activities offered by the new technologies [5].

As a background, and for further study of the subject, a recent survey is provided that compiles the main existing documentation on intelligent educational virtual systems [6].

The aim of this paper is to show how an intelligent virtual educational system based on data analytics has been designed to improve the efficiency of teaching and learning processes being delivered virtually in primary education in developing countries. This design of the intelligent system is carried out within the framework of the Telefonica Chair "Data analytics for educational projects in vulnerable environments" that the Pontifical University of Salamanca and the ProFuturo Foundation of the Telefonica Group maintain.

1.1. Intelligent Systems and Education

There is evidence that the education sector is increasingly migrating towards virtual educational processes. The emergence of COVID-19 forced all educational institutions to develop their training activities online, and this fact led to the detection of a new virtual communication channel to spread its activity to a global world.

This move towards online teaching leads to substantial changes in institutional processes and in the teaching–learning processes themselves. If institutions have to change their structures to ensure the development of their online exams and efficient communication with their students, teachers now have a large amount of information at their disposal, which needs to be filtered and organized coherently to achieve efficient learning [7].

This type of learning generates a large amount of data where the possible exploitation enables the irruption of intelligent educational systems in the virtual environment, which will result in fewer errors in the manual educational tasks or in the possibility of providing personalized teaching more in line with the progress and knowledge demonstrated by the students [8].

Intelligent virtual educational systems are not systems that will finish with the classical teaching–learning process. These systems should be seen as a complement to the process that focuses on improving the quality of learning through the automation and analysis of available educational data [6], seeking to generate knowledge that can be used by institutions or faculty in their strategic or teaching planning.

Its emergence in the area of education has been late in coming, but it is here to stay thanks, according to [9], to the explosion in the use of mobile technologies and the emergence of procedures capable of detecting the state of learning of students and the characteristics of the contexts where they are developing.

These intelligent systems will make it possible to discover the behavior of students in order to enable personalized teaching where each individual student can have the right information at the right time and in the right place [5,10].

1.2. ProFuturo Foundation

The ProFuturo Foundation is a digital education program promoted by Telefonica Foundation and La Caixa Foundation that aims to reduce the educational gap in the world providing quality digital education aimed at enhancing the futures of millions of children living in vulnerable environments in Latin America, the Caribbean, Africa and Asia [11].

Its educational proposal focuses the action on technology, in line with goal 4 of the United Nations 2030 Agenda for Sustainable Development, using innovative digital teaching–learning experiences to enhance the development of digital competencies for the 21st century.

It uses a proprietary platform with customizable educational resources, which are adapted to the local learning context, and serves primary schools in more than 40 countries with low socio-economic development indices or territories with low school enrolment.

Together with the learning platform, the different educational centers receive basic technological equipment that allows them to access the platform's contents: server, router, laptop, tablets, etc. The interaction idea is that students access the platform and play an active role in the learning process through the principle of learning by doing. This digital ecosystem allows the teacher to assess students continuously, and to personalize teaching according to the progress of each student.

In this project, teachers are the key figure in the education system, as educational guides for their students, and for this reason, they must have the necessary educational tools for the development of their own digital skills and their subsequent use in the learning process.

Through the platform, they are in a continuous period of training in the use of essential digital tools for their application in the classroom or in the use of active methodologies that foster digital skills and enable them to select educational content that is efficiently adapted to the social context where they are.

1.3. Education & New Technologies

The irruption that new technologies are having in the educational field is making it possible to infer a broad knowledge of the educational institutions functioning, the students' behavior or the teachers' teaching–learning process. The application of this knowledge to the modification of educational processes or institutional structures will lead to a substantial improvement in the training of individuals.

The virtual field is no stranger to this fact. The application of data mining techniques to the vast amount of data from students' virtual interactions has created the term Educational Data Mining, which searches for patterns or inference rules among the stored records. A generalization of this term is given by Learning Analytics, which extends the search for patterns to the use of other analytical techniques such as statistics, classifications or decision trees [12,13].

Another technology that has entered the educational field is Big Data. Its high potential for action is underlined by the great possibilities of this field, focusing on improving students' behavior during face-to-face and virtual classes [14], and on identifying students' educational priorities [15].

Their application, combined with Educational Data Mining or Learning Analytics techniques, allows intelligent virtual educational systems to focus their action on educational aspects such as the improvement of the learning process, the analysis of the learners' learning profile, or the ability to acquire the skills according to the tasks performed [16].

The latest technology to be incorporated into the field of education is artificial intelligence, through its predictive algorithms implemented with machine learning techniques, which allows the incorporation of automated processes that assist users in the decision-making processes. In their operation, automated processes classify records by relationships or related activities in order to detect similar personal patterns that can be extrapolated to similar students [17].

These artificial intelligence techniques are also beneficial for teachers to improve their interaction with students. The increase in the quantity and quality of communications and the availability of strictly necessary information, selected by automatic procedures of artificial intelligence, provoke an effective interaction that fosters students' enthusiasm to the learning process [18].

2. Dataset, Materials, and Methods

2.1. Dataset Source

The origin of the data that make up the dataset of the intelligent virtual educational system comes from each of the educational centers that belong to the Profuturo educational project, which means that the system has a set of heterogeneous data sources.

While the process of data collection is similar in all cases, the different circumstances in each country (network connection, right communications infrastructure provision, culture of action) and of each center determine the quality of the data collected.

The implementation of the ProFuturo Foundation in these developing countries has not stopped evolving, and new emerging countries are joining the education system designed. In the beginning, countries such as Bolivia, Brazil, Ecuador, and Colombia were the main data providers for the system. However, the Foundation's expansion into the Asian and African continents is now making it possible to increase the number of different data sources [11]. Data are currently collected in 40 countries, which are reflected in the map in Figure 1.

Figure 1. ProFuturo map.

The intelligent virtual educational system in use captures data from all these sources, and must undergo a transformation process to provide reliable data. This process is developed in two stages, a data cleaning stage and a data conversion stage, which aim to make the data available in a format that allows them to be analyzed by the algorithms present in the intelligent system [19].

The analysis will work with homogeneous data records, highlighting the main fields to be stored as those contained in Table 1.

Table 1. Main fields to be collected for each activity.

Field	Description
Record Origin	Educational center origin
Record Date	Collection date
Record Subject	Subject or type of content worked on
Activity Identification	Code to identify the activity
Teacher Identification	Code to identify the teacher
Student Identification	Code to identify the student
Activity Duration	Duration in minutes and seconds of the activity
Digital Resource Employed	Digital resources used in the activity
Activity Assessment	Result of the assessment of the activity
Comments	Any further comments

The collection of this data in a common format will enable further processing of the data by intelligent algorithms, in order to achieve, through data analytical techniques, a knowledge that will allow decisions to be taken for a substantial improvement of education in that country or area.

2.2. Data Cleaning Process

Within intelligent systems, there is a need to ensure the availability of reliable data to work with, in order to avoid errors, both in the actions executed automatically and in the results provided.

This need makes it indispensable to activate a data cleaning process that enables or converts the ingested amount of available data into clean data ready to be processed by the selected algorithms [20]. The actions of the data cleaning process can be implemented in different ways. Duplicate data, incomplete data, missing types are some of the problems that appear very frequently according to [21].

Studies acknowledge that the time spent on the data cleaning process consumes between 50% and 70% of the data analytics processing time [22]. This has led to a search for intelligent procedures, based on algorithms and statistical techniques, in order to automate data cleaning tasks [23].

Applying this fact to the intelligent virtual educational system designed, the existence of inhomogeneous data sources means that the sample presents an excessive level of noise that causes a workload of around 60% of the total time spent.

In order to reduce this time, the focus has been on two areas of action: the origin of the sample and the intelligent system, where data cleaning measures are established to provide reliable data. At the source of the sample, in addition to the training received by data handlers, automatic controls have been set up to respect the type of information to be sent, the range of values or the absence of information. In the intelligent system, similar images of the dataset have been created in order to observe the differences between them and proceed to unification of samples. Furthermore, procedures based on the classification of previous records have also been implemented to fill in missing data without distorting the sample received.

This data cleansing and conversion process can be seen in Figure 2, which is a consequence of adapting various existing data cleansing processes to the intelligent system [24,25]:

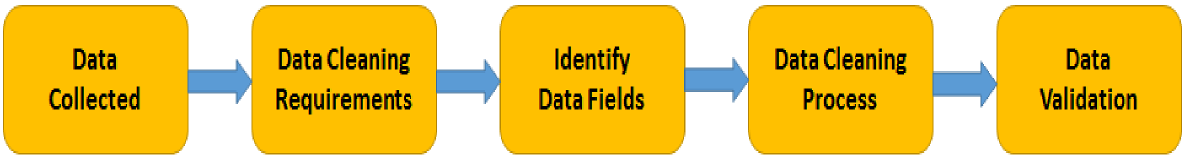

Figure 2. Data Cleaning Process.

- Data Collected: Data collection should be done using automated procedures that minimise the potential for errors to be transmitted. Forms should be created to assist the data sender (students, teachers, or institutions) to rectify possible errors at source
- Data Cleaning Requirements: It is necessary to establish the requirements to be met by the data within the intelligent system. Types, mandatory data, value ranges, presence of information, unduplicated codes, and referential integrity are some of the aspects to be determined.
- Identify Data Fields: The automatic procedures used in data cleansing need to be matched to the fields being checked, in a way that each field to be recorded has its own associated process.
- Data Cleaning Process: The data cleaning process activates the enabled procedures for the detection and correction of data and filters and transforms the data into the corresponding format, deleting duplicate, conflicting, incomplete or invalid data from user interaction.
- Data Validation: Data validation is the final stage and is used to ensure that the dataset is consistent, with high quality and in the necessary format prior to data processing.

This data cleaning process will allow to increase productivity and lead to faster and more efficient decision making, which will redound in a better functioning of the intelligent virtual educational system in order to provide the relevant findings to the groups concerned.

2.3. Algorithms and Analytical Tools Selection

After the data cleaning and conversion process, the data is now in a format suitable for the future application of algorithms and analytical tools. Among the wide range of possible techniques to be used [26] the following, associated with artificial intelligence, big data, data mining, or fuzzy logic, should be highlighted:

- Predictive models: They are data-driven models that, using statistical techniques, predict an outcome [27]. Regression, clustering, Bayesian networks, or predictive analytics are some of the techniques involved.
- Machine learning: These algorithms enable academic institutions to make decisions based on knowledge rather than intuition. Techniques such as artificial neural network or support vector machine are some of the most commonly used algorithms [28].
- Visualization: Visualization analytics tools help deliver knowledge in the format demanded. In this analytical technique, it is worth mentioning the use of Principal Component Analysis (PCA), which allows dimensionality reduction by projecting onto a space while preserving the original distances [29].
- Association rules: They are tools that create correlations to look for associations between all datasets records [30]. They work in transactional mode and provide insight into the patterns followed by students in their learning process.
- Classification: It is one of the most commonly used techniques, which identifies and assigns a category to a data collection within a context [31]. It uses decision trees, clustering, neural networks, linear programming, and statistical technique.
- Fuzzy Logic: Fuzzy logic techniques work with approximate values, rather than exact values, where the range of values is between 0 and 1. This type of technique is often used to make recommendations that can be adjusted to the learner's profile to help them improve their learning [32].

3. Designing the Intelligent Virtual Educational System

The design of an intelligent virtual educational system must incorporate not only the automatic procedures inherent to an intelligent system, but must meet the requirements that the educational area is demanding in the virtual environment. Personalization of platforms and digital resources, technological analysis of stored data, activity monitoring or feedback to improve the efficiency of the intelligent system are some of the aspects to be incorporated.

This design should also include the technological aspects necessary to add the required functionality to the intelligent system. Artificial intelligence algorithms, machine learning techniques, agent technology, or the potential of Big Data or IoT, should be considered when implementing the design.

Thus, the functionalities to be achieved by the system include the automatic acquisition of all the educational transactions performed, its cleaning and conversion to a format that allows the application of the selected algorithmic techniques and their interpretation to generate usable results that can be applied to improve the teaching and learning process.

With these guidelines in mind, and considering some of the contributions to the design of intelligent systems in the scientific literature [33–35], is presented the design of the intelligent virtual educational system, which is structured in four fundamental subsystems. The design sequencing can be seen in Figure 3.

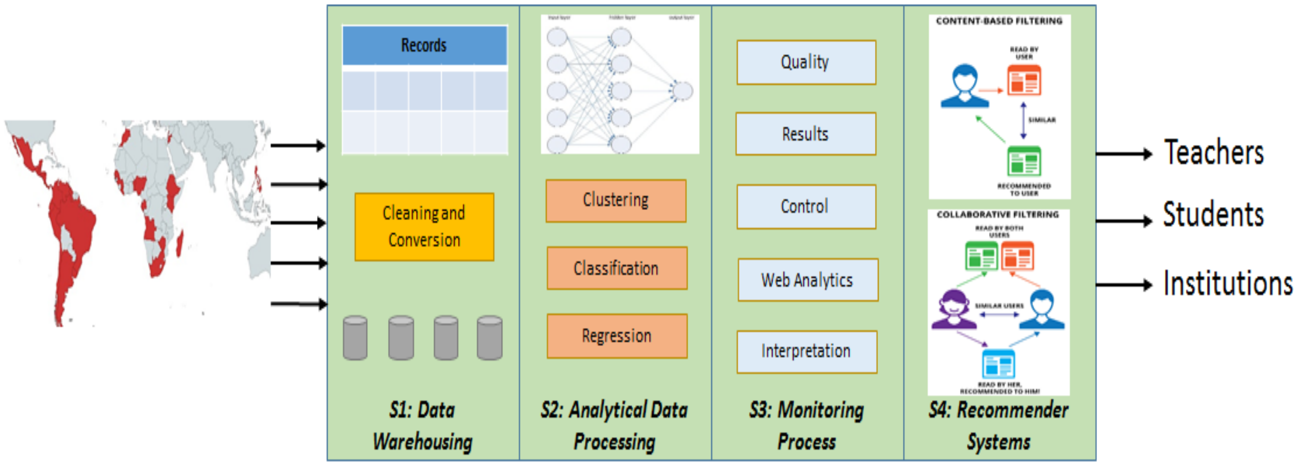

Figure 3. Intelligent virtual educational system design.

- Data warehousing. (S1)
- Analytical data processing. (S2)
- Monitoring process. (S3)
- Recommender systems. (S4)

3.1. Data Warehousing Subsystem

This subsystem is responsible for capturing and storing the data transferred from the different developing countries. The process is activated when students or teachers interact with any of the activities or educational resources available in the system and takes care of recording the development of this activity or the use that has been made of the educational resource available.

This data collection is performed automatically under the standard formats defined, which contain the main information to be processed by the intelligent virtual educational system.

This data transfer is stored in several data entry tables that contain the multiple records generated by the daily educational interactions of students in each of their subjects and activities.

Based on these records, and before their final storage in the reliable storage space of the intelligent system, a data cleaning and conversion process is required, according with Section 2.2, to provide quality-assured data in a comfortable format in order to the data analytical process can extrapolate certainties rather than uncertainties.

Two technologies are being used in this data warehouse: the MongoDB database, as a NoSQL database, which handles the processing of unstructured data, and, on the other hand, SQL Server is used as a relational database management system for the processing of structured data.

This subsystem aims to provide reliable data to the analytical subsystem to avoid the scope of erroneous analytical data and so, to be able to safely infer or focus data relationships. Figure 4 shows graphically the description of the operation of this subsystem:

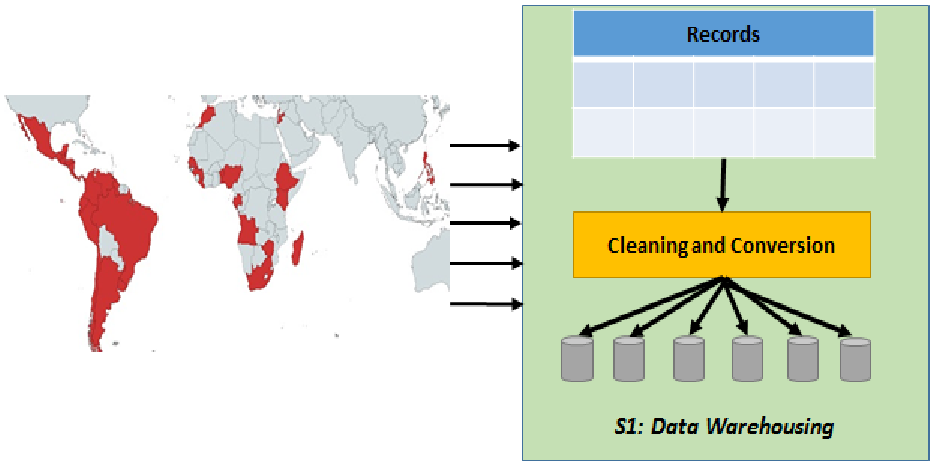

Figure 4. Data warehousing subsystem.

3.2. Analytical Data Processing Subsystem

Once the process of collecting and storing information has been completed and the reliability of the data is guaranteed, the intelligent system analytically processes all the information using data mining techniques that facilitate machine learning.

Machine learning is an inductive process of knowledge that creates algorithms capable of generalizing behavior and recognizing patterns based on available information or storage records. Therefore, it takes as a basis for its training the experience provided by the different activities carried out by the students and applies data mining techniques in a search of recognizable patterns.

This subsystem processes the information by semi-supervised learning combining the algorithms of supervised learning, which establishes a correspondence between inputs and outputs, and from the unsupervised one, which works on a set of input records in search of innovative patterns [36]. In this learning combination, it should be noted that this intelligent virtual education system shows a greater predominance of unsupervised learning over supervised learning.

Techniques such as classification, regression, clustering, decision trees, and neural networks are activated in order to find stable relationships between data that allow inferring new knowledge that can be applied by external agents to the intelligent virtual educational system [37].

For the application of these techniques, a first selection of the available dataset is made by setting up a dashboard, which is oriented towards an educational aspect to focus the analytics. Sometimes this dashboard focuses on subjects, countries, teachers or educational resources, which gives a great dynamism to the intelligent system that can analyze a set of dashboards for different purposes.

This dashboard, which contains the activity performed in the intelligent educational virtual system, is processed with the most appropriate algorithmic techniques. The choice of these algorithms is dynamic and their application depends on the results provided by the intelligent educational virtual system.

Some dashboards employ classification algorithms, which, based on past experience, classify records through a complex system of values. The most commonly used algorithm is the supervised machine learning algorithm Random Forest, which combines predictions made by multiple decision trees, where each tree is generated based on the values of an independent set of random vectors. In other cases, clustering techniques are used to group the records according to the value of certain fields in order to establish homogeneous samples. This subsystem uses the Agglomerative Hierarchical Clustering algorithm, which takes the data belonging to a cluster as a starting point, by joining the closest data and creating a cluster in a hierarchical format. Regression techniques also allow us to infer values for empty fields based on accumulated experience. These regression techniques employ, also, Random Forest, that uses ensemble learning method, combining predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.

Furthermore, the intelligent virtual educational system makes use of neural network techniques that learn from stored records in order to establish a predictive model that can probabilistically infer the classification or clustering of the collected records. The type of neural network used is a feedforward multi-layer model with a single intermediate layer, since multi-layer models with several intermediate layers are prone to over-fitting of data and, given the high volume of data available, it would involve a very high computational cost.

Finally, it is important to highlight the use of machine learning algorithms to find behavioral patterns in the fields of the records that can infer analogous knowledge among students in order to incorporate such knowledge into the teaching and learning process.

Figure 5 shows the analytical subsystem with the dashboard created and the most commonly used analytical techniques:

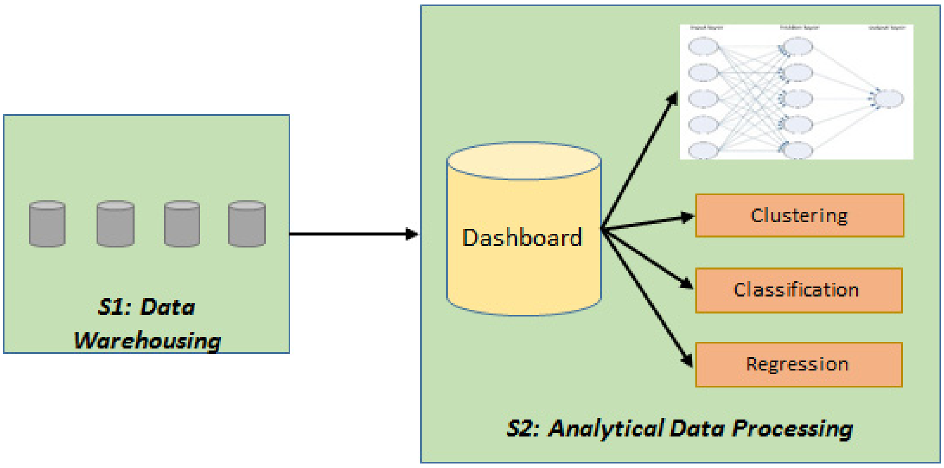

Figure 5. Analytical data processing subsystem.

3.3. Monitoring Process Subsystem

Although the activities of analytical techniques infuse usable results and point out possible unknown paths to follow, a monitoring subsystem is required to track the evolution of the intelligent system and via monitoring controls, it can observe the progress of the implemented teaching–learning process oriented to the acquisition of digital skills.

These monitoring controls aim to obtain data to assess the quality of the intelligent system and thus implement measures to improve those aspects of education that are not evolving as the institutions wish. Its implementation in the system must be done in an orderly manner, in order to provide usable results in the intelligent system [38].

Teachers' skills, the use of digital resources, connection times or connection failures, are often some of the aspects to be monitored so that the intelligent virtual educational system can take further actions to improve its performance [39].

Although there are so many possible monitoring controls to be implemented, the main controls that are being monitored in the current system are as follows:

- Time spent in each activity.
- Average resolution time for each activity.
- Number of accesses in absolute value for each activity.
- Total number of activities by content area.
- Ratio of participants completing training actions.
- Number of technology solutions reporting data.
- Median number of active users per technology solution offered.
- Average number of incidences reported per educational center.

Along with these specific controls, there is also the use of web analytics that are responsible for searching and inferring a pattern of behavior in the students' navigation through the different activities and educational resources offered by the intelligent system. Its purpose is to detect which activities are the object of students' attention in order to extrapolate them to the proposal of new or improved educational pedagogical activities [40].

These analytical techniques are of great help for the recommender system to establish, through association rules, personalized processes based on previous experiences that allow students to achieve the desired skills.

Figure 6 represents the different elements of the monitoring subsystem, where quality controls and web analytics cohabit to generate results that must be interpreted by the intelligent system to be sent to the recommender subsystem.

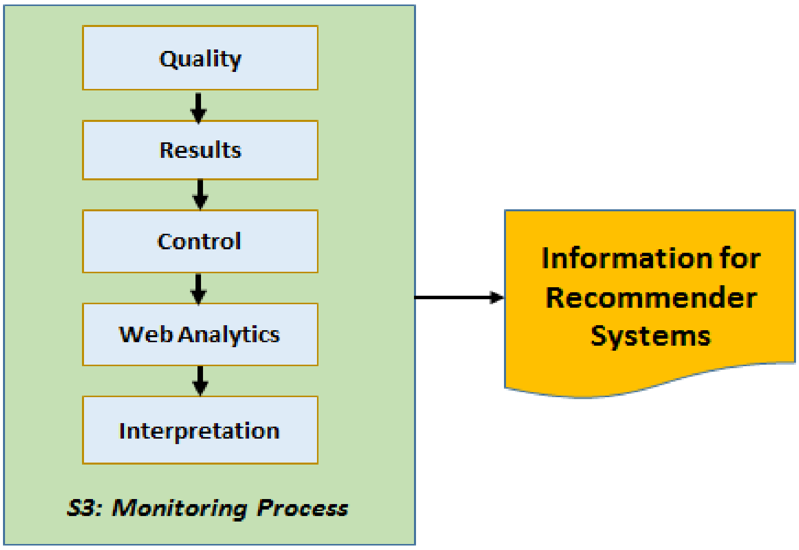

Figure 6. Monitoring process subsystem.

3.4. Recommender Systems

Considered as an output subsystem, it is responsible for transmitting to teachers, students and educational centers the recommendations developed by the intelligent virtual educational system in order to be able to apply them in new iterations of the teaching–learning process.

This subsystem elaborates the recommendations based on the data obtained in the monitoring system and on the rules inferred as a recommender system. The inference of these rules is in a continuous dynamic process, as the values of the monitoring processes themselves are used in their refinement from the executed controls. This rules generation takes place under two different recommender systems, depending on the filter used:

The content-based filtering recommender system recommends users who have made use of an educational resource, new educational resources with similar content or grouped in the same subject in order to go more deeply into the subject. It also provides teachers with new assessment or learning procedures, which have been obtained from the clustering detected by the intelligent system [41].

The recommender system based on collaborative filtering uses the didactic resources or learning procedures that a person has used, for recommendation to persons who have previously had access to the same resources. This system is based on the similarity between users and filters according to the intelligent system's classification of users based on detected preferences [41].

Along with these recommender systems, there are other recommender systems based on hybrid methods [42–44], where they take as a reference the experiences of activities based on existing records.

In its execution, the subsystem uses visualization and fuzzy logic techniques to recommend the activities or educational resources to be used to improve the efficiency of the teaching–learning process. Thus, visualization techniques make it possible to visualize, in certain groups, which activities tend to achieve the established skills and which other scattered activities do not tend to achieve these skills.

Figure 7 shows how the monitoring subsystem provides the information input and how the recommender systems deliver new guidelines to the different actors involved.

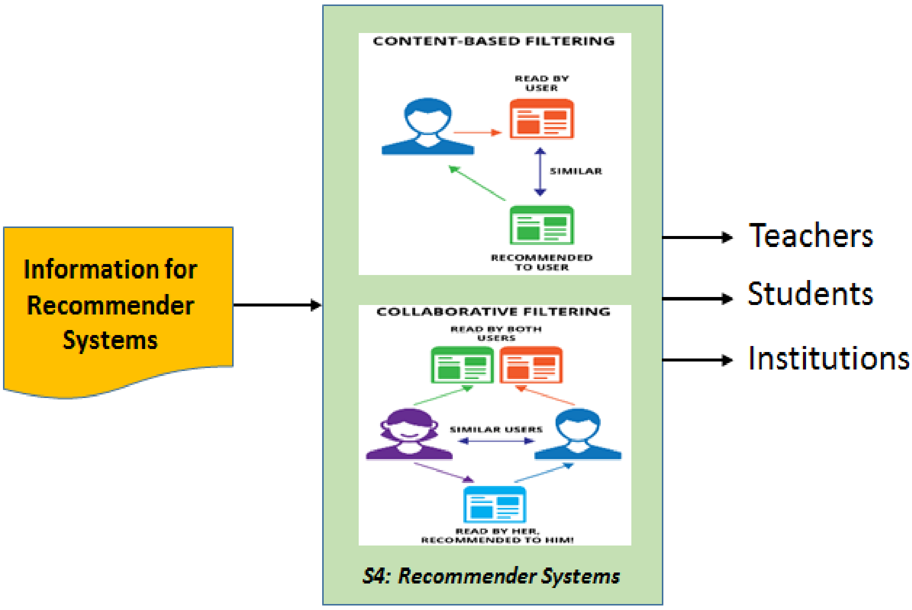

Figure 7. Recommender systems.

4. Results

As a sample of the results offered by the intelligent virtual educational system designed, we provide the results obtained in two dashboards whose respective objectives are the analysis of the time spent using the applications and the teaching skills of the professors to be evaluated. Each dashboard will provide a description of the dashboard, the analytical results obtained and an interpretation by the intelligent virtual educational system.

The confidentiality agreement signed with the Telefonica Chairs reduces the sample used to the minimum necessary to support results that can be presented. In addition, since these two dashboards are focused on Spanish-speaking countries, the figures present the text in Spanish to avoid manipulation. In any case, the text has an immediate correspondence with the English terminology.

4.1. Dashboard 1: Application Usage Time

The knowledge about how students in schools are using the different digital resources provided is a relevant circumstance that allows us to analyze the use made of these resources. In this way, it will be possible to understand and examine the utilization of the most commonly used digital resources and to analyze the reasons for the underuse of other resources that have been practically discarded.

For this data analysis about the application usage time, a dashboard has been created to visualize the use of the resources, which has been divided into three aspects: use per month (Figure 8), use per country (Figure 9) and use per educational institution (Figure 10).

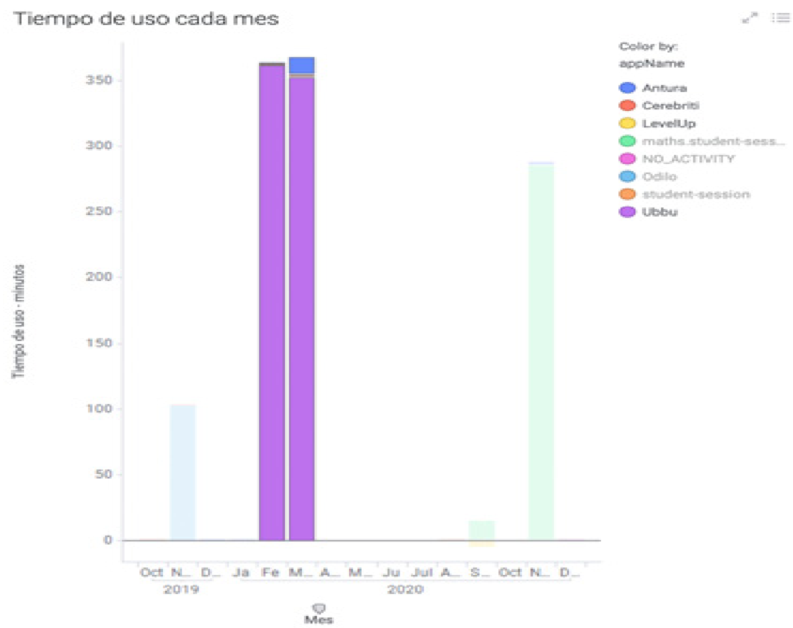

Figure 8. Digital resources usage time per month.

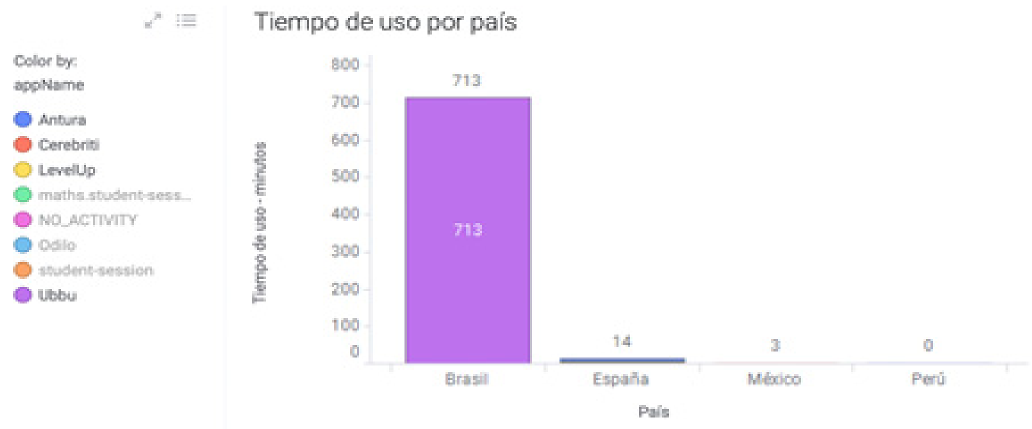

Figure 9. Digital resources usage time per country.

Figure 10. Digital resources usage time per educational institution.

This dashboard comprises activity records obtained from a variable sample of students, with a potential number of 2000–4000 students. The variability of the students participating in the dashboard depends on the activities that the teacher deems appropriate to develop using technological equipment and the students' own interest in its use.

From the graphs provided by the intelligent system, it is clear that Brazil is the country that has made most use of the activities and educational resources available through the platform. Furthermore, it can be seen that the time in which these activities have been executed has been concentrated in February and March, being the most used resource the technological solution "Ubbu".

Likewise, it can be observed that there have been very few or no connections from countries such as Peru and Mexico. This observation will allow us to focus our research on the reasons for such low participation, whether it was due to the non-utilization of available resources, or whether it was due to connectivity problems that did not allow the storage of the records generated.

On the other hand, the data in Figure 10 allow us to distinguish the most active schools from those that hardly use digital resources. Thus, improvement actions will focus on the planning of activities or training courses to increase the participation of these educational centers.

4.2. Dashboard 2: Self-Assessing Teachers' Digital Skills

Assessing digital skills and their pedagogical appropriation by teachers is a cornerstone for training teachers capable of transforming their pedagogical practice to underpin high-quality education that meets the challenges of the digital era.

For this purpose, the Self-Evaluation Tool of Digital Skills for Teachers (developed by the Centre of Innovation for Brazilian Education-CIEB- and adopted and adapted by ProFuturo) has designed a dashboard, with the participation of 280 teachers from all around the ProFuturo Foundation. In the dashboard we can visualise the assessment of 12 relevant skills for teachers, and the level that the teacher has reached in each of them. To proceed with the analysis, the skills to be assessed are distributed, equally, in three areas: pedagogical, digital citizenship, and professional development. Table 2 shows the distribution of the skills to be assessed:

Table 2. Teaching skills areas to assess.

Area	Skills
Pedagogical	Pedagogical Practice
	Personalization
	Evaluation
	Curation and Creation
Digital Citizenship	Responsible use
	Critical use
	Safe use
	Inclusion
Professional Development	Self-development
	Self-evaluation
	Sharing
	Communication

The assessment of these skills is executed according to a scale, which offers a scope of five possible levels of appropriation that display the teacher´s progression through each of these skills. These levels of appropriation are shown in Table 3.

Table 3. Levels of appropriation.

Level of Appropriation	Value	Description
Exposure	1	There is no use of technologies in the pedagogical practice, or the teacher requires support from third parties to use them
Familiarization	2	The teacher begins to know and use technology in his/her activities
Adaptation	3	Technologies are used regularly and can be integrated into the planning of pedagogical activities
Integration	4	The use of technology is frequent in planning the activities and the interaction with the students
Transformation	5	The teacher uses technologies in an innovative way

Based on the records received from the teachers' performance in their respective subjects, the intelligent system analyses the information using a diagnostic tool and displays the results in a dashboard with aggregated, anonymous, and nominal data and information on the teachers´ appropriation levels and use of technologies, in order to focus efforts on improving the least developed skills. Figure 11 shows the skills section of the Teacher Self-Evaluation tool dashboard:

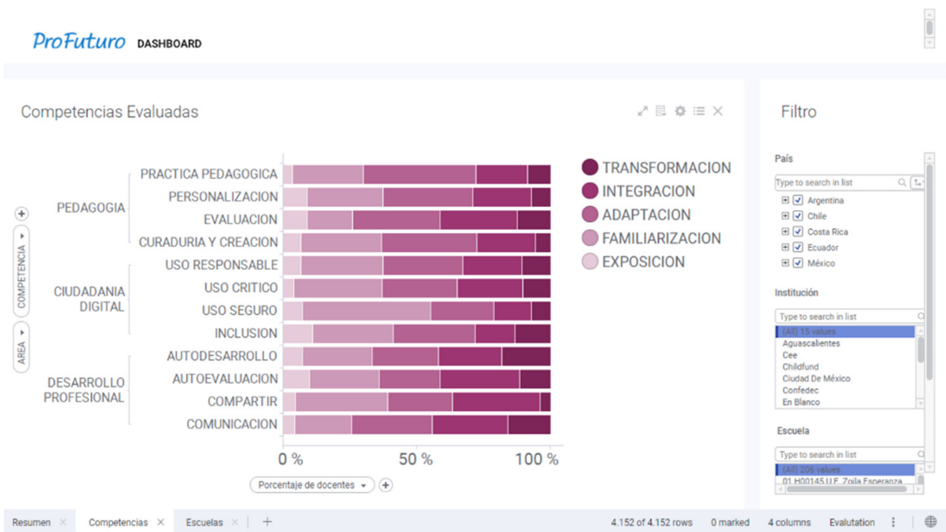

Figure 11. Distribution of teachers in each skill in accordance with the level of appropriation.

This aggregated feedback can help guide public education networks, private educational organisations, civil society organisations, and other research and cooperation institutions to the planning of continuous training programmes and courses to promote the teachers´ professional development. For example, the analysis in the figure shows that most of the teachers who took the self-assessment test, can be placed at the familiarization and adaptation levels.

It also shows that the teaching skills that have experienced the greatest transformation are those related to self-development and communication, both grouped under professional development. On the other hand, it can also be deduced that digital citizenship skills are not very developed, as all of them are in the low range, with a value close to 50%. This analysis has certified that the efforts to improve the teacher´s skills have to be focused on the four skills related to digital citizenship with the aim of achieving, at least, the range of adaptation or integration in the assimilation of these skills.

5. Discussion

The increasing use of e-learning educational systems has led technology to move fast into the education field, looking for improving the efficiency of virtual educational procedures. In the beginning, this virtual education resided in learning platforms. Nowadays, with the appearance of artificial intelligence and big data techniques, there is an approach to intelligent virtual educational systems in order to take advantage of the large amount of generated data.

In the references section, there are a lot of authors who show how clustering [27], classification [31], machine learning [28], or data mining techniques [8,12,13,15] perform analytical learning tasks [45,46] providing more useful knowledge to educational centers heads to make decisions based on the reliable data generated by their students.

Although many of these techniques are already running properly, there are still some future research that must be highlighted in order to improve an intelligent virtual educational system.

The need to adapt collection methods to each particular intelligent system, the thinning and automation of data cleansing processes, or the optimal selection of datasets that will be analyzed in order to find patterns or useful knowledge, are just a few aspects to research. The creation of new forms of monitoring to improve intelligent system tracking, the implementation of innovative visualizations, or the improvement of the effectiveness of recommender systems, are other aspects to highlight for further research in our intelligent virtual educational system.

Focusing on the students it is possible to look for the implementation of new forms of assessment to allow them self-assessment or to help them make an accurate assessment of all available resources, without the teacher's intervention.

Finally, and meeting the needs of educational institutions, it is necessary to focus research on a better integration of these intelligent systems with the different existing learning platforms in order to reuse the large number of teaching resources that have been developed.

There is no doubt that these intelligent virtual educational systems are here to stay and give a new impulse and a radical change in the known procedures. The incorporation of artificial intelligence techniques into learning analytics processes is still at an incipient stage. Even so, it allows us to pointer an educational future with personalized teaching offering each student, at any given moment, the digital resources they need to continue learning the subject being studied, and each teacher will receive a new evaluation and teaching processes that will allow them to check their students are acquiring the required skills. Moreover, and just to finish the discussion, it is necessary to appoint that all these tasks will have to be developed in these intelligent systems considering an ethical perspective to preserve citizen privacy.

6. Conclusions

The main aim of this paper was to show the design of an intelligent virtual educational system based on data analytics to improve the efficiency of primary education virtual learning processes in developing countries.

This design starts collecting data, automatically, from different educational centers in developing countries belonging to the ProFuturo program. Then, in an analytical way, the intelligent system processes the transformed and cleaning data in order to generate useful knowledge to infer new educational processes to be applied in primary education.

The intelligent virtual educational system is providing encouraging results that are allowing us to focus the teaching–learning process on new forms of interaction. Furthermore, through the recommendations provided, it will also help teachers to make better decisions in the selection of technological tools and educational resources to work with.

Given the good performance of the intelligent virtual educational system, both, the members of the Chair and the ProFuturo Foundation, are convinced that the research and the thinning of the algorithms will continue in order to obtain greater efficiency in the recommendations provided to primary education in developing countries.

Author Contributions: Conceptualization, V.A.-S., A.-J.L.-R. and M.-J.R.-G.; methodology, A.-J.L.-R., M.M.-M.-A. and O.A.-G.; software, V.A.-S. and M.-J.R.-G.; validation, V.A.-S., A.-J.L.-R. and M.-J.R.-G.; formal analysis, A.-J.L.-R. and M.M.-M.-A.; investigation, V.A.-S., A.-J.L.-R., M.M.-M.-A. and O.A.-G.; resources, V.A.-S. and M.-J.R.-G.; data curation, V.A.-S., A.-J.L.-R., and M.M.-M.-A.; writing—original draft preparation, V.A.-S. and O.A.-G.; writing—review and editing, V.A.-S., A.-J.L.-R., M.M.-M.-A. and O.A.-G.; visualization, A.-J.L.-R. and O.A.-G.; supervision, V.A.-S., M.M.-M.-A. and M.-J.R.-G.; project administration, M.M.-M.-A. and M.-J.R.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out in collaboration with Telefonica Chair "Data analytics for educational projects in vulnerable environments" that the Pontifical University of Salamanca and the ProFuturo Foundation have signed.

Data Availability Statement: The agreement signed with Telefonica Chair restricts the availability of the data to a confidential environment to safeguard the rights and privacy of the persons involved.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Elatia, S.; Ipperciel, D.; Zaiane, O.R. (Eds.) *Data Mining and Learning Analytics: Applications in Educational Research*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
2. Berendt, B.; Littlejohn, A.; Kern, P.; Mitros, P.; Shacklock, X.; Blakemore, M. *Big Data for Monitoring Educational Systems*; Publications Office of the European Union: Luxembourg, 2017. [CrossRef]
3. Tuomi, I. *The impact of Artificial Intelligence on Learning, Teaching and Education. Policies for the Future*; Cabrera, M., Vuorikari, R., Punie, Y., Eds.; Publications Office of the European Union: Luxembourg, 2018. [CrossRef]
4. Buche, C.; Bossard, C.; Querrec, R.; Chevalier, P. PEGASE: A Generic and adaptable intelligent system for virtual reality learning environments. *Int. J. Virtual Real.* **2010**, *9*, 73–85. [CrossRef]
5. Dhanalakshmi, S.; Komalavalli, K.; Hemalatha, R.; Kalyani, S. A study on impact of virtual intelligence among the students of Higher Education with special reference to Chennai City. *Shanlax Int. J. Educ.* **2021**, *9*, 196–200. [CrossRef]
6. Da Silva, L.M.; Dias, L.P.; Rigo, S.; Barbosa, J.L.; Leithardt, A.R.; Leithardt, V.R. A literature review on Intelligent Services Applied to Distance Learning. *Educ. Sci.* **2021**, *11*, 666. [CrossRef]
7. Zapata-Ros, M. The smart university. The transition from Learning Management Systems (LMS) to Smart Learning Systems (SLS) in Higher Education. *RED Rev. De Educ. A Distancia* **2018**, *57*, 1–43. [CrossRef]
8. Baker, R.S. Educational Data Mining: An Advance for Intelligent Systems in Education. *IEEE Intell. Syst.* **2014**, *29*, 78–82. [CrossRef]
9. Hwang, G.J. Definition, framework and research issues of smart learning environments- a context-aware ubiquitous learning perspective. *Smart Learn. Environ.* **2014**, *1*, 4. [CrossRef]
10. Cruz-Benito, J.; Theron, R.; Garcia-Peñalvo, F.J.; Pizarro, E. Discovering usage behaviors and engagement in an Educational Virtual World. *Comput. Hum. Behav.* **2015**, *47*, 18–25. [CrossRef]
11. Profuturo Foundation. Available online: https://profuturo.education/ (accessed on 22 March 2022).
12. Romero, C.; Ventura, S. Educational data mining and learning analytics: An updated survey. *WIREs Data Min. Knowl. Discov.* **2020**, *10*, e1355. [CrossRef]
13. Kavitha, G.; Raj, L. Educational Data Mining and Learning Analytics: Educational Assistance for Teaching and Learning. *Int. J. Comput. Organ. Trends* **2017**, *7*, 21–25. [CrossRef]
14. UNESCO; Wyatt-Smith, C.; Lingard, B.; Heck, E. *Digital Learning Assessments and Big Data: Implications for Teacher Professionalism*; UNESCO Working Paper; UNESCO: Paris, France, 2019. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000370940 (accessed on 8 March 2022).
15. Javidi, G.; Rajabion, L.; Sheybani, E. Educational Data Mining and Learning Analytics: Overview of Benefits and Challenges. In Proceedings of the 2017 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 14–16 December 2017; pp. 1102–1107. [CrossRef]
16. Yunita, A.; Santoso, H.B.; A Hasibuan, Z. Research Review on Big Data Usage for Learning Analytics and Educational Data Mining: A Way Forward to Develop an Intelligent Automation System. *J. Phys. Conf. Ser.* **2021**, *1898*, 012044. [CrossRef]
17. Kashive, N.; Powale, L.; Kashive, K. Understanding user perception toward artificial intelligence (AI) enabled e-learning. *Int. J. Inf. Learn. Technol.* **2021**, *38*, 1–19. [CrossRef]
18. Seo, K.; Tang, J.; Roll, I.; Fels, S.; Yoon, D. The impact of artificial intelligence on learner-instructor interaction in online learning. *Int. J. Educ. Technol. High. Educ.* **2021**, *18*, 1–23. [CrossRef] [PubMed]

19. Rahman, A.U. What is Data Cleaning? How to Process Data Analytics and Machine Learning Modeling? 2019. Available online: https://towardsdatascience.com/what-is-data-cleaning-how-to-process-data-for-analytics-and-machine-learning-modeling-c2afcf4fbf45 (accessed on 8 March 2022).
20. Christine, P. The importance of data cleaning: Three visualization examples. *Chance* **2020**, *33*, 4–9. [CrossRef]
21. Osborne, J.W. *Best Practices in Data Cleaning*; SAGE Publications: Thousand Oaks, CA, USA, 2013.
22. Dodds, L. Do Data Scientists Spend 80% of Their Time Cleaning Data? Turns Out, No? Available online: https://blog.ldodds.com/2020/01/31/do-data-scientists-spend-80-of-their-time-cleaning-data-turns-out-no/ (accessed on 8 March 2022).
23. LakshmiMphil, S. An overview study on data cleaning, its types and its methods for data mining. *Int. J. Pure Appl. Math.* **2018**, *119*, 16837–16848.
24. Wang, J.; Wang, X.; Yang, Y.; Zhang, H.; Fang, B. A review of data cleaning methods for Web Information System. *Comput. Mater. Contin.* **2020**, *62*, 1053–1075. [CrossRef]
25. Dilmegani, C. Data Cleaning in 2022: Steps to Clean Data & Tools. Available online: https://research.aimultiple.com/data-cleaning/ (accessed on 8 March 2022).
26. Romero, C.; Ventura, S.; Pechenizkiy, M.; Baker, R.S. (Eds.) *Handbook of Educational Data Mining*; CRC Press: Boca Raton, FL, USA, 2011.
27. Martinez, A.; Moreno-Ger, P. Comparison of Clustering Algorithms for Learning Analytics with Educational Datasets. *Int. J. Interact. Multimed. Artif. Intell.* **2018**, *5*, 9–16. [CrossRef]
28. Nieto, Y.; García-Díaz, V.; Montenegro, C.; Gonzalez-Crespo, R. Supporting academic decision making at higher educational institutions using machine learning-based algorithms. *Soft Comput.* **2019**, *23*, 4145–4153. [CrossRef]
29. Gil, C. Principal Component Analysis. Available online: https://rpubs.com/Cristina_Gil/PCA (accessed on 17 March 2022).
30. Garg, A. Complete Guide to Association Rules. Available online: https://towardsdatascience.com/association-rules-2-aa9a77241654 (accessed on 17 March 2022).
31. Geeksforgeeks. Basic Concept of Classification (Data Mining). Available online: https://www.geeksforgeeks.org/basic-concept-classification-data-mining/ (accessed on 17 March 2022).
32. Ozdemir, A.; Alaybeyoglu, A.; Mulayim, N.; Uysal, M. An Intelligent Systems for determining learning style. *Int. J. Res. Educ. Sci.* **2018**, *4*, 208–214. [CrossRef]
33. Al-Hudhud, G. Intelligent system design requirements for personalizing e-learning systems: Applications of AI to education. *Int. J. Eng. Educ.* **2012**, *28*, 1353–1359.
34. Deena, G.; Raja, K. Designing an Automated Intelligent e-Learning system to enhance the knowledge using machine learning techniques. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 112–119. [CrossRef]
35. Yin, W. An artificial intelligent virtual reality interactive Model for Distance education. *J. Math.* **2022**, *2022*, 7099963. [CrossRef]
36. Sancho-Caparrini, F. (University of Seville, Andalucia, Spain) Aprendizaje Supervisado y no Supervisado. 2020. Available online: http://www.cs.us.es/~{}fsancho/?e=77 (accessed on 22 March 2022).
37. Bogarin, A.; Cerezo, R.; Romero, C. A survey on educational process mining. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1230. [CrossRef]
38. Roman, J.A.; Perez-Delgado, M.L. A proposal for the Organizational Measure in Intelligent Systems. *Appl. Sci.* **2020**, *10*, 1806. [CrossRef]
39. Torres-Porras, J.; Alcántara, J.; Rubio, S.J. Virtual platforms use: A useful monitoring tool. *EDMETIC* **2018**, *7*, 1. [CrossRef]
40. Miñan-Olivos, G.; Dios-Castillo, C.A.; Cardoza-Sernaque, M.A.; Pulido-Joo, L.A. Web analytics to develop a Learning Analytics environment and its relationship with academic performance in virtual courses. *Rev. Innov. Educ.* **2021**, *23*, 82–94. [CrossRef]
41. Agarwal, D.K.; Chen, B.C. *Statistical Methods for Recommender Systems*; Cambridge University Press: Cambridge, UK, 2016.
42. Aggarwal, C.C. *Recommender Systems. The Textbook*; Springer International Publishing: Cham, Switzerland, 2016.
43. Manouselis, N.; Drachsler, H.; Verbert, K.; Santos, O.C. (Eds.) *Recommender Systems for Technology Enhanced Learning*; Springer Science + Business Media: New York, NY, USA, 2014.
44. Mendoza, G.; Laureano, Y.; Perez, M. Similarity and evaluation metrics for collaborative based recommender systems. *Rev. De Investig. En Tecnol. De La Inf.* **2019**, *7*, 224–240. [CrossRef]
45. Papamitsiou, Z.; Economides, A.A. Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *J. Educ. Technol. Soc.* **2022**, *17*, 49–64. Available online: https://www.learntechlib.org/p/156100/ (accessed on 22 March 2022).
46. Caballé, S.; Gómez-Sánchez, E.; Weinberger, A.; Demetriadis, S.; Papadopoulos, P. (Eds.) *Intelligent Systems and Learning Data Analytics in Online Education*; Elsevier Inc.: London, UK, 2021. [CrossRef]