



# Article Towards Using Unsupervised Learning for Comparing Traditional and Synchronous Online Learning in Assessing Students' Academic Performance

Mariana-Ioana Maier<sup>+</sup>, Gabriela Czibula<sup>\*,†</sup> and Zsuzsanna-Edit Onet-Marian<sup>†</sup>

Department of Computer Science, Babeş-Bolyai University, 400084 Cluj-Napoca, Romania; mariana.dindelegan@ubbcluj.ro (M.-I.M.); zsuzsanna.onet@ubbcluj.ro (Z.-E.O.-M.) \* Correspondence: gabriela.czibula@ubbcluj.ro or gabis@cs.ubbcluj.ro; Tel.: +40-264-405327

+ These authors contributed equally to this work.

Abstract: Understanding students' learning processes and education-related phenomena by extracting knowledge from educational data sets represents a continuous interest in the educational data mining domain. Due to an accelerated expansion of online learning and digitalisation in education, there is a growing interest in understanding the impact of online learning on the academic performance of students. In this study, we comparatively investigate traditional and synchronous online learning methods to assess students' performance through the use of deep autoencoders. Experiments performed on real data sets collected in both online and traditional learning environments showed that autoencoders are able to detect hidden patterns in academic data sets unsupervised; these patterns are valuable for the prediction of students' performance. The obtained results emphasized that, for the considered case studies, traditional evaluations are a little more accurate than online evaluations. Still, after applying a one-tailed paired Wilcoxon signed-rank test, no statistically significant difference between the traditional and online evaluations was observed.

**Keywords:** unsupervised learning; autoencoders; t-SNE; educational data mining; students' performance analysis; online learning

# 1. Introduction

Within the *educational data mining* domain (EDM) [1], there is a continuous interest in extracting knowledge from educational data sets. Academic institutions are interested in improving their teaching methodologies, learning processes [2] and the academic performance of their students and instructors [3]. Thus, uncovering meaningful patterns from academic data sets using *machine learning* techniques, both *supervised* and *unsupervised*, may help educational institutions to understand and improve their education-related processes. One of the most important missions of EDM is to predict students' learning outcomes.

We are witnessing the rapid evolution of *online learning* and digitalisation in education due to various environmental factors. Pandemics, such as the COVID-19 pandemic, change almost all human activities, including education. Education providers need to change their *traditional learning* approach, with online learning being a solution to continue providing education. The success of learning in online settings depends on the quality of teaching and the motivation of students. However, the quality of teaching does not guarantee the students' motivation or vice versa, because the latter depends on other factors, both intrinsic and extrinsic [4]. Within this framework, when both teaching and evaluation are moved online, there is concern regarding how to analyse the students' learning process and improve their learning results. Tsipianitis et al. [5] describe two categories of online learning: *synchronous* and *asynchronous*. The synchronous online learning setting requires simultaneous participation of instructors and learners, while asynchronous learning does not take place in real time [5].



Citation: Maier, M.-I; Czibula, G.; Oneţ-Marian, Z.-E. Towards Using Unsupervised Learning for Comparing Traditional and Synchronous Online Learning in Assessing Students' Academic Performance. *Mathematics* **2021**, *9*, 2870. https://doi.org/10.3390/ math9222870

Academic Editors: Paula Catarino, Maria M. Nascimento, Ana Paula Aires and Helena Campos

Received: 9 October 2021 Accepted: 8 November 2021 Published: 11 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Most of the existing work in the EDM literature related to students' performance analysis was connected to predictive modelling. Classical machine learning models, such as decision trees, neural networks, support vector machines, random forests (RFs) [6], and *relational association rules* [7,8] have been applied. Recently, deep learning methods have been applied for *students' performance prediction* (SPP). Tsiakmaki et al. [9] pointed out the efficiency of transfer learning methods in the EDM domain and investigated whether using student data from a course for training a deep learning model would lead to a model which is applicable for other related courses. *Unsupervised learning* (UL)-based descriptive models have also been applied in EDM for analysing the academic performance of students. Partitional and hierarchical agglomerative clustering methods [10], expectation maximization (EM) and particle swarm optimization-based clustering [11] have been applied to discover student profiles and patterns connected to their academic performance. Variational autoencoders (VAEs) were used by Klingler et al. [12] for learning efficient feature embeddings that were useful for increasing the performance of standard machine learning-based classifiers by up to 28%.

Regarding the supervised learning techniques used in EDM for students' performance prediction, we remarked that logistic regression, naive Bayes and decision trees gave good results in SPP (the *area under the ROC curve* (AUC) more than 0.8) [13]. We also noted that linear discriminant analysis (LDA) and autoencoders (AEs) provided classification accuracies more than 0.75 [14]). Concerning the use of unsupervised learning methods in the EDM literature, we remarked that SOMs are good tools for visualising clusters of students determined based on various student characteristics. Unlike existing related work, our work focuses on the clustering provided by AEs and t-SNE models.

The performance of students in both online and traditional environments has been of great interest in various studies in educational data mining [15]. E-learning activities have been analysed using unsupervised learning methods with the goal of identifying clusters of students who have similar learning behaviours [16,17], while recent papers have investigated how online classes impact the performance and satisfaction of students during the COVID-19 pandemic [18,19].

We conduct, in this paper, a study on applying unsupervised learning techniques to comparatively analyse the academic performance of students in traditional and synchronous online learning environments. Unsupervised learning techniques are addressed in our study due to their usefulness for mining patterns from unlabeled data. The main goal of our study is to analyse if the patterns learned from students' performance data sets in traditional learning environments are preserved in the case of synchronous online learning as well.

In the machine learning literature, AEs are used as neural network-based architectures aimed at learning how to approximate the identity function by learning to reconstruct the input. AEs were applied in numerous data analysis and mining applications for learning meaningful features from data [20], analysing images [21], speech processing [22], uncovering patterns and structural relationships between proteins [23], analysing protein functional dynamics [24] or text summarization [25].

The main contributions of the paper are as follows. Firstly, the ability of AEs to uncover hidden patterns in data will be further investigated with the aim of analysing if the students' learning patterns are preserved in both traditional and online learning. We expect that AEs are able to uncover the underlying structure of student performance-related data and find a separation, even if not a perfect one, between students that belong to different performance classes. Experiments are performed on real case studies and data sets collected from Babeş-Bolyai University in traditional and synchronous online learning environments. To support the interpretation of the learning patterns provided by AEs, a 3D *t-distributed stochastic neighbor embedding* (t-SNE) [26] analysis is then conducted. Afterwards, for a more suitable evaluation of the results and for strengthening the unsupervised learning-based analysis, *linear discriminant analysis* (LDA) is applied as a supervised classification algorithm for estimating the students' performance classes.

assess, on the considered case studies, the influence and impact of online learning upon the academic performance of students when compared to a traditional learning environment. To the best of our knowledge, there is no study in the educational data mining literature using AEs with the goal of comparing traditional and synchronous learning.

To summarize, the following research questions are targeted in our contribution.

- **RQ1**To what extent are AEs, through their latent representation, able to uncover, in an unsupervised fashion, students' learning patterns in academic data sets that are relevant for predicting their academic performance?
- **RQ2** Are the results of the unsupervised learning-based analysis correlated with the performance of a supervised classification model trained for predicting the students' performance?
- **RQ3**To what extent are the students' learning patterns on the considered case studies preserved in virtual environments when compared to traditional learning?

Since the encoder part of an AE may be useful for compressing the input data and reducing its dimensionality, the latent (encoded) representation learned by an AE will be used for answering the previous RQs and expressing relevant patterns in students' performance data.

The rest of the paper is organized as follows. Section 2 reviews existing approaches from the EDM literature related to student performance analysis and prediction in both traditional and online learning environments. The methodology applied for conducting our unsupervised learning-based investigation on students' academic performance in traditional and synchronous online learning environments is introduced in Section 3. Section 4 presents the experimental results, while Section 5 discusses the relevance of the unsupervised learning-based analysis. The conclusions of the paper and several directions to further extend it are summarized in Section 6.

# 2. Related Work

In the early 2000s, the development of technology had increased the amount of online courses for students, so numerous studies appeared in the EDM field regarding student performance analysis in online learning environments.

In 2004, McDonald et al. [15] compared the results of students in traditional versus online courses in a computer science department, specifically within a Database Systems discipline, utilizing data collected over two years, starting from 2001. There were 134 observations in traditional versus 63 observations in online courses. The authors used two statistical methods on the grades obtained by students on the final course to establish the difference between traditional and online teaching for students. The first method was the *t*-test, which was used to compare the means, and the second method consisted of regression analysis, which was used to linearly model the input data and observe which are the most important variables for the model. The conclusion of the paper was that students in traditional learning are more successful than those online. One of the reasons was the incipient phase of online learning compared to the traditional's centuries of experience.

In 2007, Durfee et al. [27] used factor analysis and self-organizing maps (SOMs) to inspect the correlation between student characteristics and their openness for computerbased learning. Lee [28] pre-processed his data with SOMs; then, principal component analysis (PCA) and k-means clustering were applied for measuring the stage of students' knowledge in online learning.

In the next decade, Tamhane et al. [13] developed predictive models using logistic regression, naive Bayes and decision trees in order to identify 8th grade students at risk of failure and to predict the students' success or failure in standardised tests. The best results overall were given by logistic regression. Aguiar et al. [29] used logistic regression in their study, but they considered random forest models the most accurate for identifying failure risk for students graduating high school. These studies refer only to traditional learning.

Regarding online learning, the literature abounds in studies related to clustering algorithms applied on massive online open courses (MOOC) data. Kizilcec et al. [30], Khalil

and Ebner [31] focused on detecting some patterns in the behavior of students enrolled in MOOCs using clustering analysis. Ezen-Can et al. [32] used a k-methods clustering algorithm, and Rodrigues et al. [33] applied hierarchical and k-means algorithms to analyse the discussion forums in the MOOCs. Bara et al. [16] and Youngjin Lee [17] analysed students' behavior in online learning using SOMs on data obtained from MOOC log files. Students' learning behaviors and their relation with their academic performance has been investigated by Bara et al. [16]. Youngjin Lee [17] observed that hierarchical clustering algorithms, together with SOMs, lead to good results in identifying clusters of students with similarities in solving problems. In 2017, Bosch [34] studied the affect detection of students with unsupervised deep autoencoders, which were applied to students' interaction-log data. In 2020, Du et al. [35] analysed the early prediction of at-risk students using data from more than 600 online courses of a K-12 virtual school. Their analysis was based on latent variational autoencoder (LVAE) with a deep neural network (DNN), and they used 2D t-SNE visualization.

In 2020, Li et al. [36] proposed sequential prediction based on deep networks (SPDN) in students' performance prediction. This instrument was used to model students' behavioral sequences online by utilizing multi-source fusion convolutional neural network (CNN) techniques and to incorporate static information based on bidirectional long short-term memory networks (LSTMs). The proposed model outperformed the baseline and demonstrated a significant improvement in early warning. Additionally, it revealed the greater impact of Internet access patterns on students' performance than the online learning activities. Poudyal et al. [14] investigated the level of standard data mining techniques to discriminate between student attention patterns in lectures where computers were used. For feature extraction, they applied Haar wavelets, principal component analysis (PCA), and linear discriminant analysis (LDA). They obtained high accuracy values by using these algorithms followed by classification algorithms.

The pandemic of COVID-19 brought changes in the educational system, because many courses have switched to an online format. The concern related to the influence of this problem is explicable. Recently, Gopal et al. [18] studied the performance and satisfaction of 544 students from Indian universities in online courses during the pandemic through a satisfaction questionnaire. The authors concluded that the instructor's quality is the most important factor that affects a student's satisfaction during online classes. Hermiza Mardesci [37] ran a study on 16 students from an Indonesian university and concluded that online learning has a negative influence on students' motivation for learning. The impact of the adoption of an online system by the educational system in the world has been analysed by Patra and Sahu [38]. The study is optimistic, because the authors observed the following benefits after the imposition of online learning: interaction, convenience, enhanced learning, economy, innovations in teaching, etc. Shivangi Dhawan [39] showed the growth of EdTech start-ups in difficult times for humanity and made a SWOC (strengths, weaknesses, opportunities, & challenges) analysis of online learning. Coman et al. [40] ran a study on 762 students from Romanian universities, and the conclusion was that the higher education system from Romania is not ready for purely online learning. The approach was based on a questionnaire applied to students; IBM SPSS Statistics was used to analyse the data. The questionnaire contained items related to technical issues online, the usage of the educational platform, the schedule and tasks.

In the studies discussed above, we notice that great importance was given to the students' activities on the learning platforms, especially in MOOCs. Because switching to an online format happened very suddenly, the course that will be discussed in Section 3.2.1 provides more synchronous learning than asynchronous, and this aspect is different from the literature, which presents courses written especially for MOOCs and asynchronous learning.

Another tool that has been given importance so far is the questionnaire, but it has a subjective dimension. In our study we aim to be more objective, so we will consider the scores obtained by the students during the semester activities to verify their impact in SPP.

We aim to continue and enrich the presented work by studying SPP at the current moment. The switch to online courses which occurred in 2020 has drawn our attention to the difficulties of students who have had to adapt to a different learning style and participate only online in academic activities. At the same time, there is a need to consider traditional education, as it has been the basis for student training until 2020, so we can continue McDonald's study [15] for nowadays. Compared to the existing approaches, our proposal provides a comparative analysis between traditional and online synchronous learning, starting from objective features (scores obtained by students during the semester) and using different instruments for SPP.

# 3. Methodology

This section introduces the methodology applied for conducting our unsupervised learning-based analysis on students' academic performance in both online and traditional learning environments. Section 3.1 formalizes the student classification problem and describes the two classification tasks we further use in our study. Section 3.2 presents an overview of our approach and details its main stages.

## 3.1. Problem Formalisation

Let us consider the following formalization of the students' performance classification problem we are focusing on. Let us denote by  $Stud = \{st_1, st_2, ..., st_n\}$  a data set consisting of *n* instances, an instance  $st_i$  characterizing a student's performance in a certain academic discipline (course)  $\mathcal{D}$  during an academic semester. Each of these instances is composed of a list of features  $\mathcal{F} = \{f_1, f_2, ..., f_k\}$  representing the grades received by the students during the semester evaluations of the given course. Accordingly, each student  $st_i$  is represented as a *k*-dimensional vector  $st_i = (st_{i1}, st_{i2}, ..., st_{ik})$ , where  $st_{ij}$  represents the value of feature  $f_i$  for the student  $st_i$ .

Given a set of classes  $C = \{Cl_1, Cl_2, ..., Cl_m\}$  expressing categories of students' performance, the classification problem can be formalized as learning to approximate a function  $f : Stud \rightarrow C$ , such that for each instance  $st \in Stud$ , its performance category ( $f(st) \in C$ ) will be predicted.

We assume that the *final grade* received by a student in the course  $\mathcal{D}$  is computed by considering his/her semester grades, as well as the grade received on the exam in the examination session. We endeavor to predict the final performance of students based only on their results obtained during semester evaluations without knowing their final exam grade. This task is complex and difficult, due to unpredictabilities in both the students' learning and instructors' evaluation processes. Moreover, estimating the exact final grade for a student is more difficult than predicting broader categories of grades. Therefore, we decided to investigate two classification schemes, as follows: (1) Grade-based performance. In this classification scheme, seven classes are considered (i.e., m = 7), corresponding to the classes of grades: 10, 9, 8, 7, 6, 5 and  $\leq 4$  (the "fail" class); (2) Category-based performance. In this classification scheme, m = 4 since we are considering only four classes (categories of grades) corresponding to the following categories of grades: Excellent (E)-this class includes students with grades of 10 and 9; Good (G)-this class includes students with grades of 8 and 7; Satisfactory (S)-this class includes the students with grades of 6 and 5; and Fail (F)-this is the "fail" class, which includes students with grades of 4 and below 4. The second classification scheme, (the *category*-based one), is certainly easier, from a machine learning perspective, than the grade-based classification.

#### 3.2. Our Approach

As stated in Section 1, the main objective of our study is to analyze and compare, using *unsupervised learning*, the students' learning patterns in traditional and synchronous online environments. AEs will be used for discovering relevant patterns in students' academic data sets collected from both traditional and online learning settings with the aim of comparing how the type of learning methods (i.e., traditional and online) impact the

students' learning patterns. Our approach consists of three stages: *data collection, building the AE model* and *results evaluation*. These stages will be further detailed.

# 3.2.1. Data Collection

For the analysis presented in this paper, we have used real data sets, gathered from an undergraduate course, "Logic and functional programming", held for second year students of the Faculty of Mathematics and Computer Science, Babeş-Bolyai University, in the autumn semester. This course is compulsory for the Computer Science students, but is also optional for students in Mathematics.

The "Logic and functional programming" (LFP) course aims at introducing the declarative programming paradigm, specifically the logic and functional paradigms. The declarative programming paradigm is strongly connected with mathematical and logical modelling. Within the logic programming paradigm, the Prolog language is introduced, while Common Lisp is used for illustrating the functional programming paradigm. The programs written in declarative programming languages are generated from mathematical principles and thus, from a programming perspective, the design, implementation, abstraction and reasoning become more and more formal activities. We decided to use the LFP course in our study in order to empirically test to what extent the declarative programming learning skills that are based on mathematical principles are influenced by the student's learning environment (traditional learning vs. synchronous online learning, in our case).

We will consider data for two academic years: the 2019–2020 academic year, when all teaching and evaluation activities were performed face-to-face, and the 2020–2021 academic year, when all activities were moved online. We mention that the course we are focusing on was designed for traditional learning, but was adapted in 2020–2021 to synchronous online learning because of the pandemic. Since this course is taught both in English and Romanian, with the exact same content, organization, and evaluation, we will consider data for both languages.

For the 2019–2020 academic year, k = 10 since there are 10 features in the data sets for both languages. The first seven of them  $(f_1, \ldots, f_7)$  are grades obtained by students for homework assignments prepared at home and presented during the laboratory classes. The next two features are grades received by students for two practical exams. These exams were held during the laboratories, when students had to solve, without using any help, a problem similar to their homework assignments. Since the first nine features are all grades, their values are between 0 and 10 (if a student did not turn in a lab assignment or did not participate at the practical exam, the corresponding grade was 0). The tenth feature is the number of seminar activity points. The goal of the seminar was to practice the theoretical concepts presented at the lectures by discussing and solving together problems similar to the students' homework assignments. During the seminar, students could volunteer to solve a problem in front of their colleagues at the blackboard, under the supervision of the teacher, for which they received an activity point. During a seminar, a student could receive only one activity point, and since the considered course has 7 seminars, this feature has a value between 0 and 7.

Table 1 contains a description of the features for the 2019–2020 academic year.

At the end of the semester, students had a written exam, but this grade is not part of the data set, since we only consider as features the grades received during the semester. This written exam grade was worth 60% of the final grade, while each of the practical exams, the average lab grade and a seminar grade (computed based on the number of seminar activity points) was worth 10% of the final grade.

#	Feature Type	Feature Description	Range of Values
1		Grade for the 1st laboratory assignment: "Recursive programming in C++"	
2		Grade for the 2nd laboratory assignment: "Lists in Prolog"	
3		Grade for the 3rd laboratory assignment: "Processing of heterogeneous lists in Prolog"	
4	Laboratory assignments	Grade for the 4th laboratory assignment: "Backtracking in Prolog"	0–10
5		Grade for the 5th laboratory assignment: "Recursive programming in Lisp (1)"	
6		Grade for the 6th laboratory assignment: "Recursive programming in Lisp (2)"	
7		Grade for the 7th laboratory assignment: "Using MAP functions in Lisp"	
8	Practical	Grade for the practical exam in Prolog	0–10
9	exams	Grade for the practical test in Lisp	-
10	Seminar activity	Number of seminar points	0–7

Table 1. Feature descriptions for the 2019–2020 academic year.

For the 2020–2021 academic year, there are only 8 features (for both languages), so k = 8, because practical exams were not organized in the online setting. Therefore, the grades received for the homework assignments represent the first seven features (as described in Table 1), while the 8th feature is, instead of the number of seminar activities, a bonus score computed based on the seminar activities. Seminar activity points were given to students in the same conditions as in the previous year, but they were not part of the final grade, which was computed from the written exam at the end of the semester (60%)and the average lab grades (40%). In this way, students who were too shy to solve problems in front of others were not disadvantaged, as they could still have the maximum grade of 10 at the end of the semester. However, in order to reward students who volunteered and solved problems, the activity points were transformed into a bonus score, between 0 and 0.5 points, which was added directly to the final grade, so the maximum possible grade was 10.5. The bonus score was computed by multiplying the number of seminar activities by 0.125 and limiting the result to, at most, 0.5. In this way, students with at least 4 activity points received the maximum bonus value. Table 2 contains a description of the features for the 2020-2021 academic year.

Table 2. Feature descriptions for the 2020-2021 academic year.

#	Feature Type	Feature Description	Range of Values
1–7	Laboratory assignments	The same description as in Table 1	0–10
8	Seminar activity	Seminar bonus	0–0.5

Besides the features previously described, all data sets contain the final grade received by the student for this course (after the retake session). This final grade is considered to be the class label, which was not taken into consideration when training the unsupervised learning model; however, it was used to visualize the results. As presented above, the final grade was computed as the weighted average of the grades received during the semester (the ones considered as features in the data sets) and the grade received for a final written exam which is not part of the data set at all. Therefore, while the final grade depends on the values of the features, it is not determined by them: a student having the maximum value for all features can still have a final grade between 4 and 10, depending on the grade received for the written exam, while a student with a value of 0 for all features can still have a final grade of 6. Students that were absent from the final exam were removed from the data sets.

Two case studies will be further used in our experiments:

- 1. **First case study**. The first case study represents the data sets for the Romanian language. The data set for the 2019–2020 academic year, denoted by  $D_t$ , contains the grades for 183 students, while the one for the 2020–2021 academic year, denoted by  $D_o$ , contains the grades for 209 students.
- 2. Second case study. The second case study represents the data sets for the English language. The data set for the 2019–2020 academic year, denoted by  $D'_t$ , contains the grades for 169 students, while the one for the 2020–2021 academic year, denoted by  $D'_o$ , contains the grades for 204 students.

The data sets are publicly available at [41], and in Section 4.1, their detailed analysis is presented.

#### 3.2.2. Building the AE Model

*Autoencoders* (AE) represent a type of artificial neural network trained in a selfsupervised manner [42]. The network is made of two parts, an *encoder* and a *decoder*. The main goal of the encoder part is to take an input and map it into a so-called *latent code*, while the decoder part takes this latent code and rebuilds the original input from it. During the training, the goal is to learn the weights for both the encoder and decoder so that the difference between the original input and the reconstructed one is as small as possible. Once an autoencoder is trained, the latent code of the input data can be considered a low-dimensional representation of the data since it contains all information needed for the decoder to rebuild the original input from it. In this manner, autoencoders can be used for dimensionality reduction and data visualization.

AEs are used in our study to encode the input space  $\mathbb{R}^k$  (i.e., the set of *k*-dimensional vectors, as shown in Section 3.1) into the 2D space. Thus, the encoder models a function  $f : \mathbb{R}^k \to \mathbb{R}^2$ , while the decoder models the function  $g : \mathbb{R}^2 \to \mathbb{R}^k$  such that  $g(f(x)) \approx x$  for each input instance x.

AEs are used for supporting the hypothesis that the latent representation will encode features relevant for distinguishing between students' performances. For the AE, we have used an *undercomplete autoencoder* architecture [42] implemented using Keras [43]. The AE architecture used for the data sets from the 2019–2020 academic year is presented in Table 3, while the architecture of the autoencoder used for the 2020–2021 data sets is presented in Table 4. Since the data sets for the 2019–2020 academic year contain 10 features, while the data sets for 2020–2021 contain only 8, the only difference in the architectures is the presence of two extra layers (the first and the last) with 10 nodes for the autoencoder from Table 3. The architectures presented in Tables 3 and 4 include the structure of both the encoder and the decoder.

Layer (Type)	Output Shape	Param #	Activation
dense_1 (Dense)	(None; 10)	110	ReLU
dense_2 (Dense)	(None; 8)	88	ReLU
dense_3 (Dense)	(None; 6)	54	ReLU
dense_4 (Dense)	(None; 4)	28	ReLU
encoding (Dense)	(None; 2)	10	linear
dense_5 (Dense)	(None; 4)	12	ReLU
dense_6 (Dense)	(None; 6)	30	ReLU
dense_7 (Dense)	(None; 8)	56	ReLU
dense_8 (Dense)	(None; 10)	90	linear

Table 3. Architecture of the AE used for the data sets from the 2019–2020 academic year.

Table 4. Architecture of the AE used for the data sets from the 2020–2021 academic year.

Layer (Type)	Output Shape	Param #	Activation
dense_1 (Dense)	(None; 8)	88	ReLU
dense_2 (Dense)	(None; 6)	54	ReLU
dense_3 (Dense)	(None; 4)	28	ReLU
encoding (Dense)	(None; 2)	10	linear
dense_4 (Dense)	(None; 4)	12	ReLU
dense_5 (Dense)	(None; 6)	30	ReLU
dense_6 (Dense)	(None; 8)	56	linear

Regarding the hyperparameters used for training the AEs as well as their architecture, multiple experiments (using a grid search strategy) were performed using various combinations of hyperparameter values (optimizer, train batch size, number of epochs, loss function). The variation of the loss function was monitored during the AE training, and the best combination of values was preserved. The final values for the hyperparameters, used for training both autoencoders (Tables 3 and 4), are presented in Table 5. For training the AEs, we have used the data sets without any preprocessing. AEs were used for reducing the dimensionality of the input data sets by using a dimensionality of 2 for the latent (hidden) space. We have chosen the size 2 for the latent space in order to be able to visually represent the instances. After training the autoencoders, we used only the encoder part to provide a 2-dimensional representation for every instance, which was visualized in the 2D space for revealing the underlying structure of the input data.

Table 5. Hyperparameters used for the AE.

Hyperparameter	Value
optimizer	adam
train batch size	16
number of epochs	5000
loss	mean squared error

3.2.3. Results Evaluation

The evaluation of the results will be made in two steps. We will start with an interpretation of the results provided by the AE model. Experiments are directed on both case studies presented in Section 3.2.1. For both case studies, the *grade-based* and *category-based* classification schemes (described in Section 3.1) are considered.

As the second step, with the intention of sustaining the interpretation and analysis of the results obtained through unsupervised learning from the first step, a performance evaluation is conducted by applying linear discriminant analysis (LDA) [44] as a supervised classification algorithm that is trained (on the data sets described in Section 3.2.1) to predict the final grade received by a student, considering his/her semester grades as input. For measuring the performance of the supervised classification task, several evaluation measures from the supervised learning literature are used. Since we have a multi-class classification problem, we chose the following evaluation measures suitable for this task: *accuracy, precision, recall* and *F-measure* [45].

For specific test data (set of students) and for each of the *m* classes to be predicted (see Section 3.1), their *precision, recall* and *F1* are computed.

The precision  $P_i$  of a class i ( $\forall 1 \le i \le m$ ) is computed as the percentage of students from the test data correctly assigned to class  $Cl_i$  from all students that are predicted as belonging to  $Cl_i$ . The recall  $R_i$  of the class i is computed as the proportion of students correctly assigned to  $Cl_i$  from all students that should belong to  $Cl_i$ . The harmonic mean between  $P_i$  and  $R_i$  represents the F1 score for the *i*-th class, denoted by  $F_i$ .

From the  $P_i$ ,  $R_i$ ,  $F1_i$  values, the overall *precision* (*Prec*), *recall* (*Recall*) and *F-measure* (F1) are then computed as the weighted average of  $P_i$ ,  $R_i$  and  $F1_i$  values, respectively. Due to the fact that classes are imbalanced, the weight for a class *i* is computed as the proportion of students from the test data whose label is class  $C_i$ . The overall *accuracy* (*Acc*) is the ratio of the number of correctly classified students to the total number of students from the test data.

The values of all performance measures (*Prec, Recall, F1, Acc*) belong to [0, 1], with larger values corresponding to better classifiers.

#### 4. Experimental Results

The experimental results obtained by applying the methodology introduced in Section 3.2 are further presented. An analysis of the data sets used is first provided in Section 4.1; then, Section 4.2 presents the results of the unsupervised learning-based analysis.

#### 4.1. Data Analysis

Table 6 presents the distribution of the grades for each data set used in our experiments, while a histogram of the grades for each case study is presented on Figure 1. For measuring the imbalancement (i.e., impurity degree) of each data set, we computed its *entropy* [46]: lower entropy values indicate a higher degree of imbalancement. The third column from Table 6 depicts the entropy of the data set, while the fourth column gives the "difficulty" of the data set. The "difficulty" [47] was introduced in the literature as an estimate of how difficult it is to classify the instances from a data set. It is computed as the number of instances whose closest neighbour (considering the Euclidean distance computed for the features without the class label) has a different label. The larger the number of instances with a label not matching the label of the nearest neighbour is, the higher the value of the difficulty measure will be and, consequently, the more difficult the classification task is.

The highest and the lowest values for both the entropy and the difficulty are highlighted in the table. We note that all data sets have about the same degree of impurity.  $D'_t$ and  $D'_o$  are slightly more impure than the data sets  $D_t$  and  $D_o$ . The difficulty of the data sets shows that more than 60% of the instances have a nearest neighbour with a different class label, indicating that finding a good separation between the classes is not trivial at all. While all difficulties seem to be close to each other,  $D_o$  seems to be the most difficult data set, and  $D_t$  is the least difficult, but the difference between them is less than 0.06.

	Data Cat	<b>F</b> actoria	D'(Cult	# of Instances wi							ith Grade			
Case Study	Data Set	Entropy	Difficulty	# of Feature	s # of Instances	$\leq$ 4	5	6	7	8	9	10		
Finat	$D_t$	0.81	0.628	10	183	15	10	26	29	38	34	31		
riist	Do	0.79	0.684	8	209	11	9	24	43	46	36	40		
Facand	$D'_t$	0.78	0.657	10	169	8	11	44	36	32	23	15		
Second	$D'_o$	0.78	0.632	8	204	7	9	33	50	46	28	31		

Table 6. Description of the data sets.

Figure 2 illustrates the Pearson correlation coefficients [48] between the features and the final grade for each case study and each data set. We mention that for online learning (the 2020–2021 academic year), the two features representing the practical exams ( $f_8$  and  $f_9$ ) are missing. One observes that for the first case study, the common features (1–7 and 10) have similar correlations with the final grade in both academic years. The situation is not the same for the second case study, where there is a certain imbalance between the correlations, denoting the impact of the online learning environment. We can also see that there is a good enough correlation between the practical exams (the green coloured bars corresponding to features 7 and 8)—about 0.6 for the first case study and 0.5 for the second. These may suggest that the lack of practical evaluations in the online learning environment may have an influence on the students' performance.



Figure 1. Histogram of grades for the first case study (a) and second case study (b).



Figure 2. Correlations between the features and the final grade for the first case study (a) and second case study (b).

# 4.2. Analysis of the Results Obtained Using AEs

Figures 3 and 4 represent the visualizations of the two-dimensional encodings learnt during the unsupervised training of the autoencoder model presented in Section 3.2.2 for the two case studies. In each figure, images (a) and (b) represent the 2019–2020 academic year, and images (c) and (d) represent the 2020–2021 academic year. For both academic

years, images (a) and (c) represent the category-based classification, while (b) and (d) represent grade-based classification. In order to facilitate the interpretation of the results, we have colored the instances according to their grade/category, but information about category or grade was not used during the training process, only for the visualization. As mentioned in Section 3.2.2, the autoencoders were trained for 5000 epochs, and for all data sets, by the end of the training, the loss on the training data was between 1.1 and 2.17.



**Figure 3.** AE visualizations for the first case study: data set  $D_t$  and the *category* (**a**) and *grade*-based classifications (**b**); data set  $D_o$  and the *category* (**c**) and *grade*-based classifications (**d**).



**Figure 4.** AE visualizations for the second case study: data set  $D'_t$  and the *category* (**a**) and *grade*-based classifications (**b**); data set  $D'_o$  and the *category* (**c**) and *grade*-based classifications (**d**).

In all images, we can see a very dense region where many instances are grouped, and the other instances are scattered around in the rest of the plot. The dense regions in general contain the instances with a better academic performance (E and G in the case of categories, respectively 7, 8, 9, 10 in the case of grades), while the instances scattered around are mainly those with a poorer academic performance (categories S, F and grades 4, 5, 6).

Analysing the plots based on the categories (images (a) and (c) from Figures 3 and 4), we can see that instances from category E (marked with red circles on the scatter plots) are very close to each other. This makes sense, since in order to have a high final grade, students need to have high grades for their laboratory assignments and practical exams (if they exist), so the feature vectors for these instances should be close to each other. Interestingly, both in data set  $D_t$  and  $D_0$ , there is an outlier, one red circle far away from the others,

but which seems to have its own surrounding of yellow and green instances. Instances from category G (marked with yellow squares) are more spread over the plots; many of them are closely grouped with instances of category E, but we can find them close to instances of S and F as well. This, again, makes sense, since you can have a *good* performance in many different ways: maybe almost *excellent* performance during the semester, but poorer performance for the final exam (these being the instances close to the red circles), or poor performance during the semester, but a good performance for the final exam, or something in between. Instances of category S (the green triangles) are mainly mixed with the yellow and blue ones, further away from the red instances, although for every data set we have a few green triangles close to red circles. In the case of data set  $D'_t$ , however, there seem to be quite a lot of instances of S close to those of E and G. Finally, instances of category F (the blue diamonds) are almost always scattered around far away from instances with category E, with very few exceptions.

If we analyse the plots based on the grades (images (b) and (d) from Figures 3 and 4), we can see the same behaviour; instances with higher grades are grouped together, and as the grade gets lower, the instances seem to be more dispersed throughout the plot. Additionally, it can be seen that since there are more classes than in the case of category-based visualization, the separation between the instances from different classes is not that clear and instances belonging to different, but close, classes are mixed on the plots. We note slightly better groupings for the first case study than for the second one, for both the *traditional* and *online* setting.

Comparatively analysing the plots corresponding to *traditional* learning (2019–2020 academic year) and *online* learning (2020–2021 academic year), there are no significant differences observed between the mappings. For both case studies, the same pattern is noticeable: for the *category-based* classification, the instances are slightly better grouped for the online learning-related data sets (image (c) vs. (a)), while for the *grade-based* classification, the instances are slightly better grouped for the traditional learning-related data sets (image (b) vs. (d)).

## 5. Discussion

For these data sets, AE models indicate a good distinction between students with good results and those with poor results, but they are not able to separate them very clearly internally. A potential reason is a relatively low quantity of features involved in the UL process, i.e., the number of grades for the students' activities over a semester. As suggested in Section 4.2, exceptions were pointed out on the 2D visualisations offered by the AEs. A way to reduce the number of outliers and obtain improved mappings could be increasing the number of evaluations for students throughout the semester, e.g., to supervise students' activity throughout lectures, not only their activity during the seminars or laboratories.

A possible explanation for the detected outliers is the inherent ambivalence of the educational processes considering both the students' learning and the instructors' evaluation processes. The analysis of the data sets exploited in our work uncovered students with a noticeable discrepancy between the results obtained from evaluations during the semester and the outcome (grade/category) at the final exam. These anomalies could result from biased semester evaluations or from unexpected situations that may affect the students' learning process.

It is worth mentioning that the unsupervised learning-based analysis previously conducted is not specific to the case studies described in Section 3.2.1, but the interpretation of the results is data-driven. Our study may be applied for student performance data sets enriched with various student course features such as the students' historical learning performance, background information, course attendance, and students' motivation. These additional features might be more helpful for better understanding student learning and will be further investigated.

#### 5.1. t-SNE Visualization of the Data Sets

To support the interpretation of the results obtained using AEs (Section 4.2) and to provide insight about how the data is organized, a 3D t-SNE [26] is applied for reducing the dimensionality of the high-dimensional instances from the data sets used in our case studies. t-SNE is used in data mining as an exploratory data analysis tool for revealing patterns in data useful for clustering. *Student t-distribution* [49] is used by the model to better disperse the clusters.

For t-SNE visualizations, the implementation from scikit-learn [50] was employed, with the following values for the hyper-parameters: 20 for *perplexity*, 3 for the learning rate (for obtaining a superior learning curve) and 1000 iterations.

Figures 5 and 6 represent the 3D t-SNE visualizations for the two case studies. In each figure, images (a) and (b) represent the 2019–2020 academic year and images (c) and (d) represent the 2020–2021 academic year. For both academic years, images (a) and (c) represent the category-based classification, while (b) and (d) represent grade-based classification. In all figures, the instances are colored according to their grade/category, without using category/grade for building the model, but only for the visualization. Before applying the t-SNE algorithm, the data was normalized with the *inverse hyperbolic sine (asinh)* for increasing sensitivity to outliers in data (e.g., values which are very small or very large).



**Figure 5.** 3D t-SNE visualizations for the first case study: data set  $D_t$  and the *category* (**a**) and *grade*-based classifications (**b**); data set  $D_o$  and the *category* (**c**) and *grade*-based classifications (**d**).



**Figure 6.** 3D t-SNE visualizations for the second case study: data set  $D'_t$  and the *category* (**a**) and *grade*-based classifications (**b**); data set  $D'_o$  and the *category* (**c**) and *grade*-based classifications (**d**).

The 3D t-SNE visualizations depicted in Figures 5 and 6 reveal the same patterns as those revealed by the AE visualizations presented in Section 4.2, confirming the conclusions exposed previously. In summary, the learned patterns are:

- The plots for the grade-based classification (images (b) and (d) from the figures) provide better groupings than the plots for the grade-based classification (images (a) and (c));
- Students with poor academic performance (e.g., category F and grades 4, 5) are well separated from the students with a good performance (e.g., category E and grades 9, 10 respectively). Inside these two larger classes (good vs poor performance) there is no clear separation in subclasses. However, the subclasses observed inside these two regions may express similar patterns for students with similar academic performance;
- Instances belonging to near classes of performance (e.g., categories F-S, G-E or grades 10-9, 8-7-6, 5-4) are neighbors on the t-SNE graph well;
- Students with higher grades/categories of grades are slightly better grouped for the data sets corresponding to the traditional learning setting ( $D_t$  and  $D'_t$ ) than for the data sets corresponding to the online setting ( $D_o$  and  $D'_o$ ). Still, no significant differences are noticed in the patterns learned in the traditional and online settings.

To better quantify how well the students are separated in clusters after applying the 3D t-SNE on each of the data set from our case studies, we introduce an evaluation measure Acc, expressing the *accuracy* of the t-SNE mapping. The *Acc* value for a t-SNE graph is defined as the percentage of data points belonging to the same class (i.e., colored the same) as their nearest neighbor (closest point on the t-SNE chart). Denoting by  $\mathcal{X}$  the set of 3D points outputed by the t-SNE algorithm, the *accuracy* is defined as  $Acc = \frac{\sum_{p \in \mathcal{X}} \omega(c(p), c(nn(p)))}{|v|}$ where  $\omega(r_{p}) = \int_{-\infty}^{\infty} 1 \quad if n = c$ 

where  $\omega(p,q) = \begin{cases} 1 & if \ p = q \\ 0 & otherwise \end{cases}$  (*nn*(*p*) represents the nearest neighbor of *p*, and *c*(*p*) represents the class corresponding to p). Acc takes values from 0 to 1 and expresses how "close" the students are who belong to the same class (i.e., grade for the grade-based classification and category for the category-based classification) on the 3D t-SNE space, as well as how well the clusters are separated. Higher values for the *accuracy* measure highlight a better clustering of the data points.

The accuracy (Acc) values computed for the t-SNE plots from Figures 5 and 6 are depicted in Table 7. The best Acc values for each case study and classification scheme are highlighted.

Classification	First Ca	se Study	Second Case Study			
Category –	D <sub>t</sub> 0.519	D <sub>o</sub> 0.574	$\frac{D'_t}{0.488}$	D'o 0.514		
Grade	0.371	0.311	0.279	0.357		

Table 7. Acc values computed for the t-SNE plots from Figures 5 and 6.

One observes that the results from Table 7 are consistent with the AE visualizations from Figures 3 and 4.

- For the *category-based* classification scheme (first row from Table 7) the *accuracy* of the t-1. SNE mapping is slightly higher for the data sets collected in the online learning setting  $(D_{\rho} \text{ and } D'_{\rho}, \text{ respectively})$ , for both case studies. Additionally, the t-SNE clustering for the first case study (for both traditional and online learning) is slightly more accurate than for the second case study;
- 2. For the grade-based classification scheme (second row from Table 7) the patterns are slightly different than for the *category-based* classification. More specifically, the Acc value for the first case study denotes slightly better partitioning for the data set collected in the traditional setting  $(D_o)$ . For the second case study, the accuracy is higher in the online learning setting, as for the *category-based* classification;
- 3. For both case studies and all data sets, more accurate t-SNE clustering is obtained for the *category-based* classification scheme.

## 5.2. Supervised Learning-Based Analysis

With the aim of reinforcing the interpretation of the results presented in Section 4.2, a supervised classification method was chosen and applied to the examined case studies and data sets. We decided to select linear discriminant analysis (LDA) because it is a popular and powerful linear classification algorithm which tries to maximize the separation between classes. Additionally, it has been previously applied in the EDM literature for SPP. 80% of the data was used for training; the remaining 20% was used for model testing. The data split was repeated 10 times, and the values for the performance measures were averaged over 10 runs of the algorithm. Table 8 presents the values for the accuracy (A), precision (P), recall (R) and F-measure (F1) computed as described in Section 3.2.3 for all data sets and both *category* and *grade*-based classification schemes. The results are averaged over 10 runs of the algorithm.

	First Case Study									Second Case Study						
Classification		I	$D_t$			Ľ	) <sub>o</sub>		$D'_t$ $D'_o$			$\overline{b'_o}$				
	A	Р	R	F1	A	Р	R	F1	Α	Р	R	F1	A	Р	R	F1
Category	0.635	0.648	0.635	0.633	0.638	0.699	0.638	0.648	0.638	0.642	0.638	0.640	0.668	0.693	0.668	0.672
Grade	0.368	0.422	0.368	0.500	0.383	0.502	0.383	0.426	0.377	0.435	0.377	0.392	0.366	0.413	0.366	0.375

Table 8. Results obtained using the LDA classifier.

The results from Table 8 reveal that the unsupervised learning analysis (AE, t-SNE visualizations and results from Table 7) are well correlated with the results of the supervised classification. The performance of the LDA classifier highlights the same patterns that were observed after the analysis of the mappings provided by the AEs (Section 4): (1) the *category-based* prediction is easier than the *grade-based* prediction, as higher values for the performance measures were obtained by LDA; (2) the *category-based* classification provided slightly better *F*1 values (with about 1.5% for the first case study and 3% for the second) for the online learning-related data sets compared to the traditional learning-related data sets compared to the online learning-related data sets.

A reason why category-based prediction was easier in online than in traditional learning could be that students orient their learning for a range of grades. In online settings, they might have noticed some patterns in the evaluation of assignments, which they applied during the examination for the range they were interested in. In traditional learning, grade-based prediction is easier because face-to-face evaluation is more transparent than online evaluation; therefore, the grades from examination are more correlated with the grades from the activity during the semester.

Considering the results of the grade-based classification, one may conclude that the semester evaluations in the traditional learning setting were slightly better correlated with the final performance of students than those in the online learning environment. This difference may be due to the practical exams that were not organized in the online setting. Another possible reason is that in the case of traditional learning, after receiving the results of the final exam, students had the possibility to meet and discuss with the teacher their paper, and grades might have been adjusted to better fit the students' performance. In the case of online learning, students had the chance to request the re-correction of their written exam, but the dialogue between the teacher and student was not possible, so the final grades might not have been so well adjusted to the actual knowledge level. However, a considerable discrepancy between the global performance in the traditional and online setting is not noticeable. For verifying this hypothesis from a statistical viewpoint, a one tailed paired Wilcoxon signed-rank test [51,52] was applied. The sample of values (accuracy, precision, recall, F1) for the category and grade-based classifications for both case studies in the traditional learning setting was tested against the sample of values obtained for the online learning setting (the performances were described in Table 8). A p-value of 0.06681 was obtained, showing that there is no notable difference between the results from online and traditional learning at a significance level of alpha = 0.05.

Certainly, there are some natural variations in the students' skills or interests for learning, because the discussed experiments used data from different sets of students attending in two different academic years. The discrepancy between these generations of students can be seen in their competencies and performances, but it does not influence the presented study. The goal of this paper was to inspect, through unsupervised and supervised learning methods, whether there are specific patterns in the students' learning process, regardless of their skills for study. We aimed to make an empirical evaluation as to what degree patterns (which were extracted from our case studies) are maintained in traditional and online learning.

#### 5.3. Comparison to Related Work

An evaluation of the published approaches for students performance analysis in online learning settings [16–18] revealed that, despite some similarities, the aim and perspective of our study stand apart from the related work.

The approaches presented in [16,17] are based only on online learning, while our research studies and likens traditional and online learning. The course considered in our study was sketched for traditional learning but, due to the pandemic, had to be promptly adjusted to synchronous online learning, so the attending students were surprised by online courses, being registered in traditional learning when they enrolled at our university. The courses described in [16,17] were built for online learning, with specific instruments and resources ([17] designed for asynchronous learning. There is not information in this sense for [16]), so the enrolled students knew from the beginning that they would study online. Thus, this study is compatible with the work presented in [18], since both of them describe similar contexts of switching learning to an online environment because of the pandemic. Unlike this study, Gopal et al. have another goal and data set [18]: they use the answers received from students in a questionnaire (a rather subjective tool) to evaluate their satisfaction. Our data sets consist of grades earned by students from evaluations over a semester (which is more objective than a questionnaire) with the aim of analysing students' performance.

# 6. Conclusions and Future Work

A study on using autoencoders as an *unsupervised learning* model for students' performance analysis was introduced in this paper. The AE model was introduced for encoding, through the latent representation, relevant features from the input data. Real academic data sets collected from Babeş-Bolyai University on two academic years in both traditional (2019–2020) and online (2020–2021) learning settings were used in the experiments.

The research questions on which our study relied on have been answered. The experiments empirically supported the hypothesis that AEs are able, in an unsupervised manner, to uncover learning patterns in student performance data that would be important to forecast the performance of students. Moreover, we highlighted that the results of the unsupervised learning-based analysis are highly correlated with the performance of a supervised classifier trained for predicting the students' performance. As an important target of our paper, we aimed to inspect the way that online learning influenced the academic performance of students when compared to learning in traditional academic environments. The experiments also emphasized that there were no substantial discrepancies between the students' learning patterns in the traditional and synchronous online learning contexts, even if we noticed that the semester evaluations in the traditional setting seemed to be a little more correlated with the students' final performance.

Future work will be oriented to extend our study by including other student course features like the students' history learning performance, background information, course attendance, and students' motivation. In addition, we aim to investigate if extending the feature set with grades received by students at other courses preceding the analysed course in the curricula would help increase the accuracy of the students' performance analysis. We further intend to use data collected from gymnasium and high schools, as well as to use other types of AEs and to include other unsupervised learning models in our analysis, such as *gradual relational association rule* mining [53].

Author Contributions: Conceptualization, M.-I.M., G.C. and Z.-E.O.-M.; methodology, M.-I.M., G.C. and Z.-E.O.-M.; software, M.-I.M., G.C. and Z.-E.O.-M.; validation, M.-I.M., G.C. and Z.-E.O.-M.; formal analysis, M.-I.M., G.C. and Z.-E.O.-M.; investigation, M.-I.M., G.C. and Z.-E.O.-M.; resources, M.-I.M., G.C. and Z.-E.O.-M.; data curation, M.-I.M., G.C. and Z.-E.O.-M.; writing–original draft preparation, M.-I.M., G.C. and Z.-E.O.-M.; writing–review and editing, M.-I.M., G.C. and Z.-E.O.-M.; visualization, M.-I.M., G.C. and Z.-E.O.-M.; funding acquisition, G.C. and Z.-E.O.-M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI – UEFISCDI, project number PN-III-P4-ID-PCE-2020-0800, within PNCDI III.

Institutional Review Board Statement: Not applicable.

**Informed Consent Statement:** Not applicable.

Data Availability Statement: Not applicable.

**Acknowledgments:** The authors would like to thank the editor and the anonymous reviewers for their useful suggestions and comments that helped to improve the paper and the presentation. This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI – UEFISCDI, project number PN-III-P4-ID-PCE-2020-0800, within PNCDI III.

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

# References

- 1. Bogarín, A.; Cerezo, R.; Romero, C. A survey on educational process mining. *Wiley Interdisc. Rew. Data Min. Knowl. Discov.* 2018, *8*, e1230. [CrossRef]
- 2. Mohamad, S.K.; Tasir, Z. Educational Data Mining: A Review. Procedia Soc. Behav. Sci. 2013, 97, 320–324. [CrossRef]
- 3. Jishan, S.T.; Rashu, R.I.; Haque, N.; Rahman, R.M. Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. *Decis. Anal.* **2015**, *2*, 1. [CrossRef]
- Nasution, S.M. Online-Learning and Students' Motivation: A Research Study on the Effect of Online Learning on students' motivation in IAIN Padangsidimpuan. Asian Soc. Sci. Humanit. Res. J. (ASHREJ) 2020, 2, 9–16. [CrossRef]
- 5. Tsipianitis, D.; Groumpos, P. University Asynchronous Distance Learning Programs to Enhance Interregional Sustainable Development. *IFAC-PapersOnLine* **2018**, *51*, 346–351.
- 6. Beaulac, C.; Rosenthal, J.S. Predicting University Students' Academic Success and Major using Random Forests. *arXiv* 2018, arXiv:1802.03418.
- 7. Czibula, G.; Mihai, A.; Crivei, L. A novel relational association rule mining classification model applied for academic performance prediction. *Procedia Comput. Sci.* 2019, 159, 20–29. [CrossRef]
- Crivei, L.M.; Andrei, M.; Gabriela, C. A study on applying relational association rule mining based classification for predicting the academic performance of students. In Proceedings of the KSEM 2019: The 12th International Conference on Knowledge Science, Engineering and Management, LNAI 11775, Athens, Greece, 28–30 August 2019; pp. 287–300.
- 9. Tsiakmaki, M.; Kostopoulos, G.; Kotsiantis, S.; Ragos, O. Transfer Learning from Deep Neural Networks for Predicting Student Performance. *Appl. Sci.* 2020, *10*, 2145. [CrossRef]
- 10. Ayers, E.; Nugent, R.; Dean, N. A Comparison of Student Skill Knowledge Estimates. In Proceedings of the 2nd International Conference on Educational Data Mining, Cordoba, Spain, 1–3 July 2009; pp. 1–10.
- 11. Dutt, A.; Aghabozrgi, S.; Ismail, M.A.B.; Mahroeian, H. Clustering Algorithms Applied in Educational Data Mining. *Int. J. Inf. Electron. Eng.* **2015**, *5*, 112–116. [CrossRef]
- 12. Klingler, S.; Wampfler, R.; Käser, T.; Solenthaler, B.; Gross, M.H. Efficient Feature Embeddings for Student Classification with Variational Auto-encoders. In Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017, Wuhan, China, 25–28 June 2017; pp. 1–8.
- Tamhane, A.; Ikbal, S.; Sengupta, B.; Duggirala, M.; Appleton, J. Predicting student risks through longitudinal analysis. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 1543–1552. [CrossRef]
- 14. Poudyal, S.; Mohammadi-Aragh, M.J.; Ball, J.E. Data Mining Approach for Determining Student Attention Pattern. In Proceedings of the 2020 IEEE Frontiers in Education Conference (FIE), Uppsala, Sweden, 21–24 October 2020; pp. 1–8. [CrossRef]
- McDonald, M.; Dorn, B.; McDonald, G. A statistical analysis of student performance in online computer science courses. In Proceedings of the Thirty-Fifth SIGCSE Technical Symposium on Computer Science Education, SIGCSE 2004, Norfolk, VA, USA, 3–7 March 2004; pp. 71–74.
- 16. Bara, M.W.; Ahmad, N.B.; Modu, M.M.; Ali, H.A. Self-organizing map clustering method for the analysis of e-learning activities. In Proceedings of the 2018 Majan International Conference (MIC), Tandil, Argentina, 8–12 October 2018; pp. 1–5. [CrossRef]
- 17. Lee, Y. Using Self-Organizing Map and Clustering to Investigate Problem-Solving Patterns in the Massive Open Online Course: An Exploratory Study. *J. Educ. Comput. Res.* **2019**, *57*, 471–490. [CrossRef]
- Gopal, R.; Singh, V.; Aggarwal, A. Impact of online classes on the satisfaction and performance of students during the pandemic period of COVID 19. *Educ. Inf. Technol.* 2021, 26, 6923–6947. [CrossRef] [PubMed]
- Farghally, M.F.; Mohammed, M.; Mahmoud, H.F.F.; Ellis, M.; Haqq, D.; Domino, M.; Jones, B.; Shaffer, C.A. The Online Transition of Two CS Courses in Response to COVID-19. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*; Association for Computing Machinery: New York, NY, USA, 2021; p. 1251.

- 20. Nguyen, T.V.; Wong, R.K.; Hegde, C. On the dynamics of gradient descent for autoencoders. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, Naha, Okinawa, Japan, 16–18 April 2019; pp. 2858–2867.
- Le, Q. Building high-level features using large scale unsupervised learning. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013, 2013; pp. 8595–8598.
- Deng, J.; Zhang, Z.; Marchi, E.; Schuller, B. Sparse autoencoder-based feature transfer learning for speech emotion recognition. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2–5 September 2013; pp. 511–516.
- 23. Czibula, G.; Codre, C.; Teletin, M. AnomalP: An approach for detecting anomalous protein conformations using deep autoencoders. *Expert Syst. Appl.* **2021**, *166*, 114070. [CrossRef]
- Teletin, M.; Czibula, G.; Bocicor, M.I.; Albert, S.; Pandini, A. Deep Autoencoders for Additional Insight into Protein Dynamics. In *Artificial Neural Networks and Machine Learning—ICANN 2018*; Springer International Publishing: Cham, Switzerland, 2018; pp. 79–89.
- 25. Kong, H.; Kim, W. Generating summary sentences using Adversarially Regularized Autoencoders with conditional context. *Expert Syst. Appl.* **2019**, *130*, 1–11. doi:10.1016/j.eswa.2019.04.014. [CrossRef]
- 26. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. J. Mach. Learn. Res. 2008, 9, 2579–2605.
- 27. Durfee, A.; Schneberger, S.; Amoroso, D.L. Evaluating students computer-based learning using a visual data mining approach. *J. Inform. Educ. Res.* **2007**, *9*, 1–28.
- 28. Lee, C.S. Diagnostic, predictive and compositional modeling with data mining in integrated learning environments. *Comput. Educ.* **2007**, *49*, 562–580. [CrossRef]
- Aguiar, E.; Lakkaraju, H.; Bhanpuri, N.; Miller, D.; Yuhas, B.; Addison, K. Who, when, and why: A machine learning approach to prioritizing students at risk of not graduating high school on time. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, ACM, New York, NY, USA, 16–20 March 2015; pp. 93–102.
- Kizilcec, R.F.; Piech, C.; Schneider, E. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In Proceedings of the Third International Conference on Learning Analytics and Knowledge, Leuven, Belgium, 8–12 April 2013; pp. 170–179. http://doi.acm.org/10.1145/2460296.2460330.
- 31. Khalil, M.; Ebner, M. Clustering patterns of engagement in Massive Open Online Courses (MOOCs): The use of learning analytics to reveal student categories. *J. Comput. High. Educ.* **2017**, *29*, 114–132. [CrossRef]
- 32. Ezen-Can, A.; Boyer, K.E.; Kellogg, S.; Booth, S. Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, Poughkeepsie, NY, USA, 16–20 March 2015. http://doi.acm.org/10.1145/2723576.2723589. [CrossRef]
- Rodrigues, R.L.; Gomes, A.S.; Souza, F.F.; Ramos, J.L.C.; Silva, J.C.S.; Maciel, A.M.A. Discovering level of participation in MOOCs through clusters analysis. In Proceedings of the 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT), Austin, TX, USA, 25–28 July 2016; pp. 232–233. [CrossRef]
- Bosch, N. Unsupervised Deep Autoencoders for Feature Extraction with Educational Data. In Proceedings of the Workshop on Deep Learning with Educational Data at the 10th International Conference on Educational Data Mining, Wuhan, China, June 25–28 2017; pp. 1–8.
- 35. Du, X.; Yang, J.; Hung, J.L. An Integrated Framework Based on Latent Variational Autoencoder for Providing Early Warning of At-Risk Students. *IEEE Access* 2020, *8*, 10110–10122. [CrossRef]
- 36. Li, X.; Zhu, X.; Zhu, X.; Ji, Y.; Tang, X. Student Academic Performance Prediction Using Deep Multi-source Behavior Sequential Network. In *Advances in Knowledge Discovery and Data Mining*; Springer: Cham, Switzerland, 2020; Volume 12084, pp. 567–579.
- Mardesci, H. The Effect of Online Learning on University Students' Learning Motivation. Educ. Learn. J. 2020, 27, 42–47. [CrossRef]
- 38. Patra, S.; Sahu, K.K. Digitalisation, Online Learning and Virtual World. Horiz. J. Hum. Soc. Sci. Res. 2020, 2, 45–52. [CrossRef]
- 39. Dhawan, S. Online Learning: A Panacea in the Time of COVID-19 Crisis. *J. Educ. Technol. Syst.* 2020, 49, 5–22. doi: 10.1177/00472395209340181. [CrossRef]
- 40. Coman, C.; Ţîru, L.G.; Meseşan-Schmitz, L.; Stanciu, C.; Bularca, M.C. Online Teaching and Learning in Higher Education during the Coronavirus Pandemic: Students' Perspective. *Sustainability* **2020**, *12*, 10367. doi:10.3390/su122410367. [CrossRef]
- Czibula, G.; Maier, M.; Onet-Marian, Z. Students Performance Prediction Data Sets-Traditional vs. Online learning. 2021. https://figshare.com/articles/dataset/Educational\_data\_mining\_-\_traditional\_vs\_online\_learning/14330447/5 (accessed on 15 February 2021).
- 42. Lopez Pinaya, W.H.; Vieira, S.; Garcia-Dias, R.; Mechelli, A. Chapter 11—Autoencoders. In *Machine Learning*; Mechelli, A., Vieira, S., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 193–208.
- Chollet, F.; Haifeng, J.; Kento, N.; Eder, S.; Qianli, S.Z.; Oleg, S.; Taehoon, L.; Matt, W.; Kian, M.A.; Kashif, R.; et al. Keras. 2015. Available online: https://github.com/fchollet/keras (accessed on 15 February 2021).
- 44. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A.E. Linear discriminant analysis: A detailed tutorial. *AI Commun.* 2017, 30, 169–190. [CrossRef]
- Picca, D.; Curdy, B.; Bavaud, F. Non-linear correspondence analysis in text retrieval: A kernel view. In Proceedings of the JADT 2006: 8es Journees internationales d'Analyse statistique des Donnees Textuelles, Besançon, France, 19–21 April 2006; pp. 741–747.

- 46. Sekerka, R.F. 15—Entropy and Information Theory. In *Thermal Physics*; Sekerka, R.F., Ed.; Elsevier: Amsterdam, The Netherlands, 2015; pp. 247–256.
- 47. Boetticher, G.D. Improving the Credibility of Machine Learner Models in Software Engineering. In *Advances in Machine Learning Applications in Software Engineering;* IGI Global: Hershey, PA, USA, 2007.
- Gijbels, I.; Hubert, M. 1.07—Robust and Nonparametric Statistical Methods. In *Comprehensive Chemometrics*; Brown, S.D., Tauler, R., Walczak, B., Eds.; Elsevier: Oxford, UK, 2009; pp. 189–211.
- 49. Martin, B. Chapter 6—Sampling Distributions Associated with the Normal Distribution. In *Statistics for Physical Science*; Martin, B., Ed.; Academic Press: Boston, MA, USA, 2012; pp. 105–122.
- 50. Scikit-learn. Machine learning in Python. 2021. Available online: http://scikit-learn.org/stable/ (accessed on 15 February 2021).
- 51. Siegel, S.; Castellan, N. Nonparametric Statistics for the Behavioral Sciences, 2nd ed.; McGraw–Hill, Inc.: New York, NY, 1988.
- 52. Social Science Statistics. Available online: http://www.socscistatistics.com/tests/ (accessed on 15 February 2021).
- 53. Miholca, D.L.; Czibula, G. DynGRAR: A dynamic approach to mining gradual relational association rules. *Procedia Comput. Sci.* **2019**, *159*, 10–19. [CrossRef]