Using Self-Organizing Maps for Comparing Students' Academic Performance in Online and Traditional Learning Environments

Zsuzsanna ONEȚ-MARIAN*, Gabriela CZIBULA, Mariana MAIER

Babeș-Bolyai University, 1 Mihail Kogălniceanu Street, 400084, Cluj-Napoca, Romania
zsuzsanna.onet@ubbcluj.ro (**Corresponding author*), gabriela.czibula@ubbcluj.ro,
mariana.dindelegan@ubbcluj.ro

Abstract: Nowadays, there is an increasing development of online learning and digitalisation in the field of education. In the current pandemic context, when teaching and evaluation activities, including lectures, assignments and examinations, are moved in online environments, there is an interest in analysing and understanding how online learning methodologies impact the students' learning outcomes as well as their evaluation process. This paper proposes the employment of a self-organizing map model trained using unsupervised learning for uncovering hidden patterns in academic data, with the aim of analysing the impact of online and traditional learning methods on the students' performance. By using real academic data sets collected from Babes-Bolyai University, from both online (academic year 2020-2021) and traditional learning environments (first semester of the academic year 2019-2020), the present analysis highlights that there are no significant differences between the students' academic performances in these environments. The unsupervised-learning based analysis is reinforced by the results obtained by applying logistic regression for predicting the students' performance.

Keywords: Unsupervised learning, Self-organizing maps, Students' performance prediction, Online learning.

1. Introduction

Nowadays there is an increasing development of *online learning* and digitalisation in the educational area. Due to the COVID-19 pandemic, normal activities were disrupted and the education moved online. In this context, when the teaching and evaluation activities, including lectures, assignments and examinations, are moved in online environments, there is an increasing interest in understanding the students' learning process and in improving their learning outcomes. There is an interest in comprehending how students react to online education, the importance of the online tools, the students' ability to access online courses, and the instructors' manner in conducting online learning activities (Butnaru et al., 2021).

The literature reveals two categories of online learning methods: *synchronous* and *asynchronous* (Tsipianitis & Groumpos, 2018). In a synchronous e-learning setting, both learners and instructors are simultaneously participating in the learning process, but at different locations. On the other hand, if the learning process does not take place in real time, i.e., there is no simultaneous participation of students and instructors, the learning process is called *asynchronous* (Tsipianitis & Groumpos, 2018).

Various machine learning approaches have been proposed in the literature for *students' performance prediction* (SPP), including support vector machines, decision trees (DTs), random forests (RFs), radial basis function networks, Bayesian learning, artificial neural networks, linear regression, relational association rules (Czibula et al., 2019a; Crivei et al., 2019). Recently, variational autoencoders have been used by Klingler et al. (2017) for learning feature embeddings that helped improve the performance of standard classifiers used in SPP. Tsiakmaki et al. (2020) investigated the transfer learning methods, by training a deep learning model on student data from one academic course and then applying it for other related courses. From an *unsupervised learning* (UL) perspective, clustering methods and *self-organizing maps* (SOMs) were applied for classifying students according to their study results (Kurdthongmee, 2008), (Khadir et al., 2015) and for detecting categories of students (very good, good, average, poor) according to their academic performance (Saxena et al., 2017).

Recent approaches have been developed in the educational data mining domain for analysing e-learning activities and online learning. A SOM clustering method has been employed in (Bara et al., 2018) for the analysis of e-learning activities with the goal of identifying clusters of students having similar learning behaviour. For detecting similar patterns in the way students are accessing the e-learning environment, the log files of their actions were used. Lee (2018) used SOMs in an exploratory study performed with the goal of investigating problem solving patterns in online

courses. The study suggested that the SOMs and hierarchical clustering can be useful for exploratory data analysis in helping instructors of *massive open online courses* to identify similar students according to various features.

Gopal et al. (2021) have recently analysed the impact of online classes on the satisfaction and performance of students during the COVID-19 pandemic. The data were collected from 544 students from Indian universities through online surveys. The authors evaluated different factors directly connected to students' satisfaction and performance with online classes: quality of instructors, design of the courses, prompt feedback, students' expectations. The study highlighted that the instructor's quality is the most important factor that affects a student's satisfaction during online classes, while the second factor is the student's expectations.

Until the second semester of the academic year 2019-2020, the instruction activities at the Babeș-Bolyai University were mostly traditional. Since the pandemic of COVID-19 started, these activities have been moved to online learning platforms, because physical meetings are not recommended.

With the broader goal of comparatively investigating online and traditional learning methods for assessing students' performance, this article studies the usefulness of *unsupervised learning* models for analysing the students' academic performance. The contribution of the paper is twofold. Firstly, the ability of *self-organizing maps* (Czibula et al., 2019b; Mihai et al., 2019) is investigated to uncover hidden patterns in academic data with the aim of analysing the impact of online and traditional learning methods on students' performance. SOM model is expected to be able to uncover the underlying structure of the students' performance related data and thus to discriminate, to a certain extent, between students belonging to different performance classes. Secondly, this paper aims to comparatively analyse the results of the UL analysis based on a real case study collected from Babeș-Bolyai University from both online and traditional learning environments. For a better evaluation of the results and for strengthening the UL-based analysis, Logistic regression (LR) (Subasi, 2020) will be applied as a supervised classification algorithm for estimating the students' performance.

A review of the existing approaches for students' performance analysis in online learning settings revealed that, despite some similarities, the purpose and perspective of the present paper differs from the related work. Unlike the works presented in (Bara et al., 2018) and (Lee, 2018) which focus exclusively on online learning, the present work considers and compares traditional and online learning. The courses considered in (Bara et al., 2018) and (Lee, 2018) were designed as online courses right from the beginning, using various tools and resources specific to such courses (the course considered in (Lee, 2018) is intended for asynchronous learning while in (Bara et al., 2018), this information is not provided, and the students, when enrolling in those courses, knew that they are going to study online). The course considered as case study was designed for traditional learning and had to be quickly adapted to synchronous online learning because of the pandemic, and the enrolled students did not expect to have online courses when they chose this university. From this perspective, the work described in (Gopal et al., 2021) is the closest to the present work, since the authors consider the same situation of moving to online learning because of the pandemic. The main difference compared to the present work resides in the goal of the analysis and the data used: Gopal et al. (2021) try to measure the students' satisfaction using the answers given by students for a questionnaire, which is quite subjective, while this paper analyses the students' performance considering the grades received for evaluations during the semester, which is more objective, even though there might be differences in the grading style of different teachers.

The data used in experiments are collected from two different sets of students, enrolled in two different academic years, so it is undoubtedly true that natural variations exist in their skills and openness for studying. These intrinsic differences between different generations of students are certainly reflected in their abilities and academic performances, but they do not directly impact the study performed in this paper. The purpose of this study is to investigate, through unsupervised and supervised learning methods, if certain patterns exist in the students' learning process, disregarding their learning capabilities. The main goal is to empirically assess, through the case studies used, to what extent these patterns (learned

in the form of relationships between the students' performance during the semester and their final examination grade) are preserved in traditional and online learning environments.

To summarize the contributions of the paper, the following research questions should be answered at: **RQ1** *Are SOMs able to detect hidden patterns in academic data sets based on unsupervised learning, patterns that would be relevant for predicting the students' performance?* **RQ2** *Are the results of the UL-based analysis correlated with the performance of a supervised classifier trained on the same case studies, for predicting the students' performance?* **RQ3** *To what extent do online learning and digitalisation influence the academic performance of students when compared to learning in traditional academic environments?*

The rest of the paper is organized as follows. The methodology applied for conducting the proposed unsupervised learning-based investigation on students' academic performance in online and traditional learning environments is introduced in Section 2. Section 3 presents the experimental results and discusses their relevance. The conclusions of the paper and directions to further improvements are outlined in Section 4.

2. Methodology

The methodology proposed for conducting the UL-based study on students' academic performance analysis in online and traditional learning environments is further introduced. Subsection 2.1 formalizes the students' classification problem and describes the two classification tasks used in the present study. Then, an overview of this approach is presented in subsection 2.2, detailing its main stages.

2.1 Theoretical Model

The students' performance classification problem is formalized as follows. Let $Stud = \{stud_1, stud_2, ..., stud_n\}$ denote a data set consisting of n instances, an instance $stud_i$ characterizing the performance of a student at a certain academic discipline $\mathcal{D}$, during an academic semester. Each instance (student) is characterized by a set of features $\mathcal{F} = \{f_1, f_2, ..., f_k\}$, which were identified as relevant for assessing the final performance of the students for the given course. For instance, in this study, the feature values are considered to be

the students' grades received during the semestrial evaluations. Accordingly, each student $stud_i$ is visualized as a k-dimensional vector $stud_i = (stud_{i1}, stud_{i2}, ..., stud_{ik})$, where $stud_{ij}$ represents the value of feature f_j for the student $stud_i$.

Given a set $\mathcal{C} = \{C_1, C_2, ..., C_m\}$ expressing classes of students' performance, the classification problem can be formalized as learning to approximate a function $f : Stud \rightarrow \mathcal{C}$ such that for each student $s \in Stud$ an estimation of his/her performance class $f(s) \in \mathcal{C}$ will be provided.

Predicting the final performance of students at a certain academic course based on their results obtained during the semestrial evaluations is a complex and difficult task, due to unpredictability in both the students' learning and instructors' evaluation processes.

It was decided to investigate two classification schemes, as follows: (1) **Grade-based classification**: in which there are seven classes (i.e. *m=7) corresponding to the following classes of grades:* ≤ 4 (the "fail" class), 5, 6, 7, 8, 9, 10. (2) **Category-based classification**: in which there are four categories (i.e. *m=4*) corresponding to the following: (1) *Excellent* (E - containing the grades 9 and 10); (2) *Good* (G - containing the grades 7 and 8); (3) *Satisfactory* (S - containing the grades 5 and 6); and (4) *Fail* (F - containing the grades 4 and below 4).

2.2 The Proposed Approach

As stated in Section 1, the objective of our study is to comparatively investigate online and traditional learning methods for assessing students' performance by means of *unsupervised learning*. The SOM unsupervised learning model will be further used for uncovering meaningful patterns in students' academic performance-related data sets with the aim of comparing the impact of online and traditional learning methods on students' performance.

The main stages of our research are as follows: *data collection and pre-processing, building the UL model, results interpretation and evaluation.* The next sections will detail these stages.

2.2.1 Data Collection and Pre-processing

The data sets used in this study are created for an undergraduate course at the Babeș-Bolyai University (BBU), Faculty of Mathematics and

Computer Science, held for second year students in the autumn semester. This course is taught both in Romanian and English and it has the same content and the same evaluation rules in both languages. Thus, the combined data sets for both the Romanian and English language will be used in this study. The students' performance is investigated for the academic year 2019-2020, when traditional face-to-face teaching was used and for the academic year 2020-2021, when both the teaching and evaluations happened online.

For the academic year 2019-2020, the data set, denoted by $D_{2019\text{-}2020}$, contains 10 features (i.e., $k=10$). The first seven features, f_i, $1 \leq i \leq 7$, are grades received by the students during the laboratory classes for homework assignments. Features f_8 and f_9 represent the grades received for two practical exams, when, during the laboratory, students had to solve a problem similar to their homework assignments. The last feature, f_{10}, represents the number of seminar activities, i.e., how many times, during the seminars, the student volunteered to solve a problem or answered a question. Since there were a total of seven seminars, the value of f_{10} is a number between 0 and 7, while the other features, since they are grades, have a value between 0 and 10.

For the academic year 2020-2021, the data set, denoted by $D_{2020\text{-}2021}$, contains 8 features (i.e., $k=8$). The first seven features represent the seven homework assignment grades, just like in case of $D_{2019\text{-}2020}$. The last feature, f_8, represents the seminar bonus, which is a value between 0 and 0.5, computed based on the number of seminar activities during the semester, and which was added directly to the final grade. The main difference compared to the previous year, regarding the features, is that, due to the online setting, practical exams were not organized.

Besides the above-mentioned features, for all the data sets, the final grade for every student (recorded after the retake session), is taken into consideration, which is the class label. But this label was not used for building the unsupervised model, it was used only for visualizing and evaluating the student's performance. If a student did not present a laboratory assignment or was absent from a practical exam, the value of the corresponding feature is 0, while students who had been absent from the final exam were removed from the data set.

The data set is publicly available at (Czibula et al., 2021). A more detailed analysis of the data set will be provided in subsection 3.1.

2.2.2 Building the UL Model

Self-organizing maps (Somervuo & Kohonen, 1999) are artificial neural networks used in unsupervised training to provide a *map*, which is a low- (in general two-), dimensional representation of the input instances. This map is made of an input layer (where neurons have the same dimension as the input data) and an output layer, where neurons are arranged in a grid, which is in general two-dimensional. During the training process, each input instance is mapped to a neuron on the output layer and this mapping is done in such a way that when training is done, the topological relations from the input space are preserved, meaning that instances that are near each other in the (high-dimensional) input space, will be mapped to neurons which are close to each other on the (two-dimensional) map. Thus, SOMs represent a good method for dimensionality reduction and data visualization.

In the present study, SOMs are applied for the unsupervised identification of some patterns relevant to estimating the students' final examination grade at a certain academic discipline, starting from their grades obtained during the semester. Experiments are conducted on the two data sets described in subsection 2.2.1. For both considered data sets, both classification schemes described in subsection 2.1 (i.e., *grade-based* and *category-based*) are considered. SOMs are used for encoding the input space $\mathbb{R}^k$ (i.e., the set of *k-dimensional* numerical vectors, as shown in subsection 2.1) into the 2D space. More specifically, the UL models will encode a mapping $f : \mathbb{R}^k \rightarrow \mathbb{R}^2$, by preserving the structure of the input data.

In this article the authors' own implementation of SOM was used, considering a torus topology for the map and the U-Matrix method (Lötsch & Ultsch, 2014) for visualizing the trained map: darker areas (i.e., low U-matrix values in this implementation) express clusters of similar instances, while whiter areas (i.e., high U-matrix values in this implementation) represent separation boundaries between the clusters. The following parameters were used in the proposed SOM implementation: an adaptive learning (0.01 for

initial value), a dimension of 50x50 for the map, 250 training epochs (during one training epoch all instances from the data set are randomly fed into the SOM). It is noted that the unnormalized data sets were used for training the SOMs.

2.2.3 Results Interpretation and Evaluation

The UL model's performance assessment is accomplished in two steps. The first step is an **interpretation** of the results obtained by the SOM models based on unsupervised learning.

As the second step, in order to sustain the previous analysis and interpretation of the results, an **evaluation** of the results is performed by applying a baseline supervised classification algorithm, Logistic Regression (LR) on the considered data sets. The results of the supervised classification process are measured using several evaluation measures employed in the supervised learning literature for measuring the performance of multi-class classifiers: **accuracy** (*Accuracy),* **precision** (*Precision),* **recall** (*Recall*) and **F-measure** (*F-measure)* (Picca et al., 2006). For computing the evaluation measures, the generalized confusion matrix *GM* is computed first, which is a square matrix of dimension *m* (as described in subsection 2.1, *m* is the number of classes to be predicted). An element gm_{ij} ($1 \leq i, j \leq m$) counts the number of instances (students) predicted as belonging to class C_i and whose actual class is C_j. Due to the imbalanced nature of the data set, for every class the *weight* of the class is defined as described in equation (1).

$$w_i = \frac{\sum_{j=1}^{m} gm_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{m} gm_{ij}} \qquad (1)$$

The precision of class C_i, denoted by $Precision_i$ is computed as the proportion of instances correctly assigned to C_i from all the instances that are predicted as belonging to class C_i (equation (2)).

$$Precision_i = \frac{gm_{ii}}{\sum_{j=1}^{m} gm_{ij}} \qquad (2)$$

The overall precision value (*Precision*) is defined as the weighted average of the obtained precision values for all the *m* classes (equation (3)).

$$Precision = \sum_{i=1}^{m} w_i \cdot Precision_i \qquad (3)$$

The *recall* of class C_i, denoted by $Recall_i$ is defined as the proportion of instances correctly assigned

to C_i from all the instances that should belong to class C_i, as presented in equation (4).

$$Recall_i = \frac{gm_{ii}}{\sum_{k=1}^{m} gm_{ki}} \qquad (4)$$

The overall recall (*Recall*) is defined as the weighted average of the recall values for the *m* classes.

The overall accuracy of the classification is computed as the proportion of the correctly classified instances from the total number of instances (equation (5)).

$$Accuracy = \frac{\sum_{k=1}^{m} gm_{kk}}{\sum_{i=1}^{m} \sum_{j=1}^{m} gm_{ij}} \qquad (5)$$

The overall *F-measure* is obtained as the harmonic mean between the weighted *precision* and *recall* values (equation (6)).

$$F-measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \qquad (6)$$

All measures take values from 0 to 1 and should be maximized for obtaining better classifiers.

3. Results and Discussion

The experimental results obtained following the proposed methodology on the data sets described in subsection 2.2 will be further presented, together with an analysis of the obtained results.

Firstly, the data sets used in the proposed experiments are analysed, then the results of the UL-based analysis are interpreted and evaluated.

3.1 Data Analysis

The distribution of grades and the categories of grades for each data set can be seen in Table 1. The third and fourth columns from Table 1 depict the *difficulty* of the data sets considering the *category-based* and *grade-based* classification. *Difficulty* is a measure introduced by Boetticher in (Boetticher, 2007), which measures how many instances are in a data set, whose nearest neighbour (excluding the class label when computing the distance between two instances) has a different label. The higher the value of the difficulty, the larger the number of instances whose nearest neighbour has a different label.

Since an instance can have several nearest neighbours with different labels, the difficulty computation proposed in (Boetticher, 2007) has been adapted in the following way: for every instance *inst*, the number of the nearest neighbour instances with the same label (which are called nn_{same}) and of the nearest neighbour instances with a different label (which are called nn_{diff}) are calculated and the difficulty of *inst* is computed according to equation (7):

$$diff(inst) = \frac{nn_{diff}}{nn_{diff} + nn_{same}} \qquad (7)$$

The difficulty of the data set containing *n* instances is computed as the average value of *diff* computed for all the instances. For computing the distance between two instances, the Euclidean distance measure was used.

Table 1 shows that both data sets have a very high difficulty. Obviously, when the categories are taken into consideration, since there are fewer possible values, the difficulty is lower, but it is still greater than 0.46. Moreover, an analysis of the data revealed that the proposed definition of the difficulty is useful, since there are several cases when an instance has more than one nearest neighbour. For example, in the $D_{2019-2020}$ data set, there is an instance with final grade 9, which has 14 nearest neighbours (with a distance of 0, which means that the features are equal for all 15 instances), 9 of them have the label 10, 4 of them have the label 9 and 1 has the label 7. The presence of such situations, together with the high difficulty values, indicate that the problem of finding a separation boundary between the classes is difficult.

To further analyse the difficulty of the data set, seeing that there might be many instances with the same (or very close) values for features, but different labels, another difficulty value has been computed, where nn_{same} denotes those instances where the absolute value of the difference of the label is greater than 1. For example, for the above-mentioned instance with label 9, the value of nn_{same} would be 13 (instances with label 10 and 9 are counted). Considering this new definition, the difficulty values are a lot lower, 0.258 for $D_{2019-2020}$ and 0.288 for $D_{2020-2021}$ when considering grades and 0.039 for both data sets when considering categories. This shows that the evaluations performed during the semester, denoted by the features, are relevant for predicting the final grade of a student, since most of the students with similar performance during the semester (i.e., close feature values) have similar, though not necessarily equal, final grades.

Figure 1 illustrates the Pearson correlation coefficients (Gijbels & Hubert, 2009) between the features and the final grade, for each data set. It is noted that for the academic year 2020-2021 (online learning) features 8 and 9 (representing the practical exams) are missing. One observes that there is a certain imbalance between the correlations, denoting an impact of the online learning environment. Good enough correlations between the practical exams (the green bars corresponding to features 8 and 9 – about 0.5 for feature 8 and 0.6 for feature 9) – are also noted. These may suggest that the lack of practical evaluations in the online learning environment may have an influence on the students' performance.
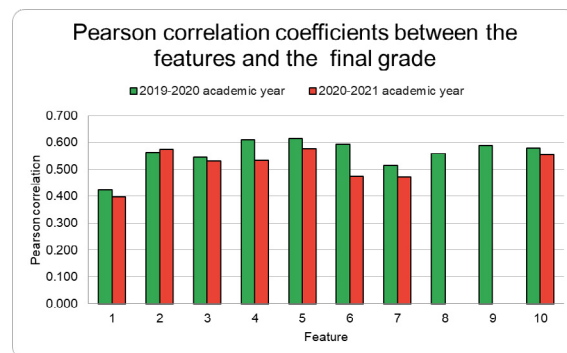

Figure 1. Correlations between the features and the final grade for the proposed case study

Table 1. Description of the data sets

| Data set | # of instances | Difficulty | | # of instances with category | | | | # of instances with grade | | | | | | |
		Category	Grades	F	S	G	E	4	5	6	7	8	9	10
$D_{2019-2020}$	352	0.463	0.667	23	91	135	103	23	21	70	65	70	57	46
$D_{2020-2021}$	413	0.463	0.663	18	75	185	135	18	18	57	93	92	64	71

3.2 Experimental Results

The results obtained by the SOM model will be further presented and discussed.

Figure 2 presents the U-matrix visualization of the SOMs trained using unsupervised learning, based on the data sets described in subsection 2.2.1. The two images from the first row represent the SOM visualizations for the data set collected in the 2019-2020 academic year (i.e., *traditional* learning): (a) the grade-based classification and (b) the category-based classification. The second row illustrates the 2D visualizations for the data

set collected in the 2020-2021 academic year (i.e., *online* learning): (c) the grade-based classification and (d) the category-based classification. On each of the SOMs from the figure, the instances (students) with the same label (i.e., **category** of the final grade - left side images - or **grade** - right side images) are marked with the same colour (the instances that failed to pass are coloured with blue). The darker regions express clusters of similar instances, while the whiter ones represent separation boundaries between the clusters.

The 2D visualisations of the SOMs from Figure 2 reveal about the same patterns on each data set.
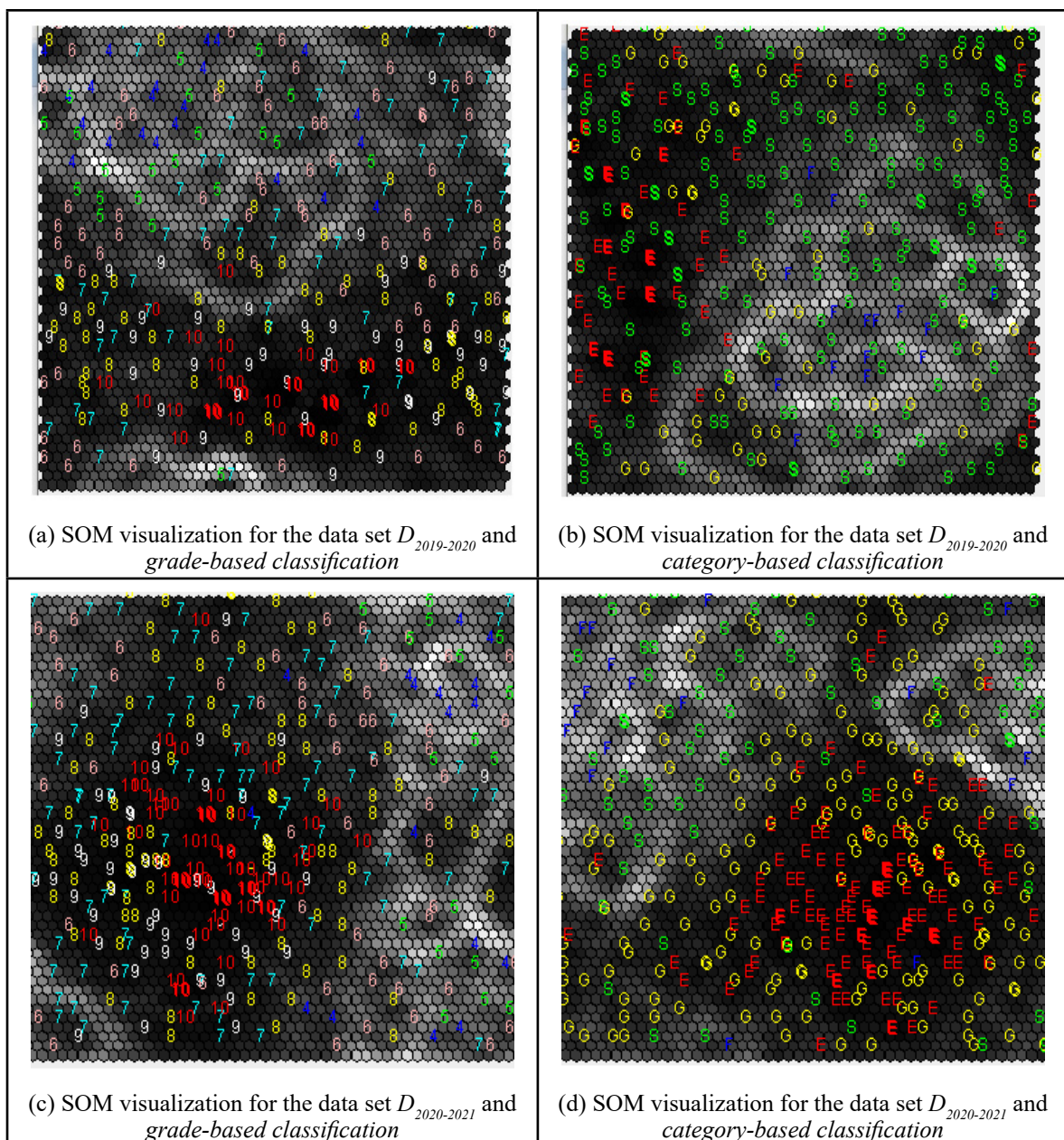

(a) SOM visualization for the data set $D_{2019-2020}$ and *grade-based classification*

(b) SOM visualization for the data set $D_{2019-2020}$ and *category-based classification*

(c) SOM visualization for the data set $D_{2020-2021}$ and *grade-based classification*

(d) SOM visualization for the data set $D_{2020-2021}$ and *category-based classification*

Figure 2. SOM visualizations for the first data set (images (a) and (b)), and second data set (images (c) and (d))

As expected, the first observation is that the problem of predicting the exact final grade for a student (images (a)/(c)) is more difficult than the problem of predicting the grade category (images (b)/(d)): a better grouping of the instances appears on the SOMs trained using the category-based classification schemes.

In all figures there is a good enough separation (two larger areas on the maps separated with a white boundary) between students with a good performance (i.e., categories E, G) and students with a poorer performance (i.e., categories S, F). Generally, there is no clear separation in classes inside these two larger areas, which is expectable due to the difficulty of estimating the final performance of students based only on their semester grades. Still, even if inside the area of students with better performance there is no separation in subclasses, inside the region of students with poorer performance one observes subclasses that may reveal similar patterns for the students inside these groups.

Overall, a main pattern observed is that neighbouring students on the map belong to near categories (i.e., categories F-S, G-E). But there are also outliers, i.e., neighbouring students on the maps, which belong to dissimilar categories (e.g., G and F).

Comparatively analysing the SOM visualizations for the traditional (academic year 2019-2020) and the online learning environments (academic year 2020-2021) one observes that the students with higher grades/categories of grades are better grouped for the data set corresponding to the online learning setting ($D_{2020-2021}$) than for the data set corresponding to traditional learning setting ($D_{2019-2020}$), where they are more dispersed.

A reason why the best grades and categories are better grouped online could be that, being in the comfort zone, students can customize their education, so that hardworking students can specialize in the desired direction more easily than in traditional learning. An explanation for why in traditional learning the best grades and categories are more widespread would be the fact that the activities are more diverse, and students have to adapt to each of them, so some find it more difficult to customize their training. On the other hand, students have physical interactions, which makes it easier for them to cooperate and compete with each other. Thus, in addition to learning techniques, they have the opportunity to observe each other and learn more easily from each other by sharing information and knowledge, thus the final exam can be a pleasant surprise for all.

A main pattern expressed by the SOMs, for all the data sets, is that they are able to discriminate between students with a good and those with a poor academic performance but are unable to make a clear distinction inside these large groups.

A possible cause may be the relatively small number of features used in the UL process, i.e., the number of grades measuring the performance of students during the semester. As pointed out above, outliers were noticed on the 2D visualisations provided by the SOMs. A possible solution for reducing the number of outliers and obtaining better mappings would be to increase the number of students' evaluations during the semester and not only during the seminar and laboratory activities, but also to monitor the students' performance during the lecture hours.

The main cause for such possible outliers is the intrinsic uncertainty of the educational processes, including both the students' learning process and the instructors' evaluation process. The analysis of the data sets used in the proposed experiments revealed students having a visible disparity between the marks obtained during the semester and their final performance (grade/category).

The reason for such anomalies may be a biased evaluation during the semester or some unpredictable occurrences in the process of students' learning.

3.3 Evaluation Results

As previously mentioned in subsection 2.2.3, the LR classification algorithm (Subasi, 2020) is applied on the considered data sets with the aim of reinforcing the results of the unsupervised analysis presented in subsection 3.2. For training the LR model 80% of the data was used, the remaining 20% was used for testing and the training/testing split was repeated 10 times. The results were averaged over these 10 runs of the algorithm. Table 2 presents the values of the *accuracy*,

precision, *recall* and *F-measure* for the considered data sets. The last column from Table 2 depicts the value of the *Area Under the Precision-Recall Curve* (AUPRC) measure. The Precision-Recall curve represents a two-dimensional plot (of both recall and precision). In case of classifiers which return the class directly, the points at (0,1) and (1,0) are linked to the point (recall, precision), and the area under the resulting curve will be computed as given in equation (8).

$$AUPRC = \frac{Precision + Recall}{2} \qquad (8)$$

AUPRC is considered a good measure for evaluating the performance of the classifiers applied on imbalanced data sets, higher values for AUPRC indicating better classifiers.

The results from Table 2 highlight the same patterns observed in subsection 3.2, for both data sets:

- the category-based prediction is easier than the grade-based prediction, as better performances (in terms of all the evaluation measures) were obtained by LR;

- the category-based classification provided slightly better F-measure value (with about 6%) for the online learning-related data set compared to the traditional learning-related data set;

- the grade-based classification provided very similar F-measure values for the traditional and online learning-related data sets, with a slight improvement (of 0.1%) for the academic year 2020-2021.

Students have adapted quickly to online learning, given that they belong to generation Z (so-called "digital natives"). A reason for which the prediction by categories is more accurate in online than in the traditional learning could be that, in traditional learning, students go through more varied experiences in different environments, like traveling to university, physical meetings with others, returning home, study or research at library, relaxation, etc. In online learning, there is a monotony in instruction, given that the physical environment for training is the same with that for relaxation and family relationships. In this monotony, "digital natives" can focus on their area of interest. In traditional learning, the preparation for instruction consists in preparing the necessary materials, leaving home, going to university, finding the classroom, choosing a place in the classroom, physical interaction with colleagues and teachers, while the preparation for online learning consists in preparing the necessary materials (mostly in electronic format) and the proper settings of the device which connects the student to the online activities. Having the internet at hand, they can anytime search for the information of their interest, without disturbing colleagues or teachers. Thus, in online learning, students stay in their comfort zone throughout the instructional process. Online virtual interaction allows digital natives to pay more attention to themselves and assimilate more easily the information from their interest area and targeted category. The synchronous learning in online was studied, where the teacher's role is similar to his/her role in traditional learning. In time, the teacher's role will change, because of the use of new technologies and asynchronous learning. As shown by Tsipianitis & Groumpos (2018), there is a risk for educators to be replaced by actual technologies. Another concern for them is the loneliness, which results from the deficit of teamwork in distance learning programs. It is considered that this could be a challenge for teachers to update their activities for engaging students to work together.
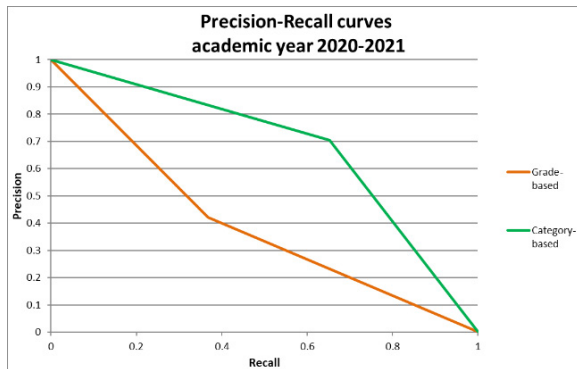
Figure 3 illustrates the Precision-Recall curves for the data sets $D_{2019\text{-}2020}$ and $D_{2020\text{-}2021}$ for both the *grade-based* and *category-based* classification. One observes that the curves have very similar shapes for both the *traditional* (academic year 2019-2020) and *online* (academic year 2020-2021) learning.

Table 2. Results obtained using the LR classifier on the considered data sets

Data set	Classification type	*Accuracy*	*Precision*	*Recall*	*F-measure*	*AUPRC*
$D_{2019\text{-}2020}$	Grade-based	0.379	0.405	0.379	0.392	0.392
	Category-based	0.556	0.701	0.556	0.620	0.629
$D_{2020\text{-}2021}$	Grade-based	0.368	0.421	0.368	0.393	0.394
	Category-based	0.653	0.705	0.653	0.678	0.679
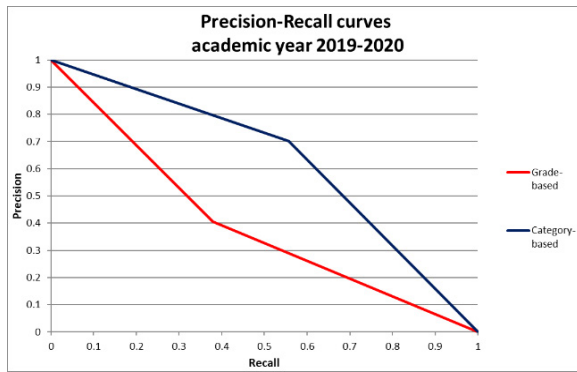
(a) Precision-Recall curves for the data set $D_{2019\text{-}2020}$, grade-based and category-based classification

(b) Precision-Recall curves for the data set $D_{2020\text{-}2021}$, grade-based and category-based classification

Figure 3. Precision-Recall curves for the data set $D_{2019\text{-}2020}$ (a) and the data set $D_{2020\text{-}2021}$ (b)

The AUPRC values obtained for the two academic years are close, being slightly better for 2020-2021 (with 0.25% for the grade-based classification and with 5% for the category-based classification).

Considering the *category-based* classification results, which are more precise than the *grade-based* ones, one observes that the *recall* and *accuracy* values are slightly better (with about 1%) in the traditional setting, suggesting that the practical exams might have had a certain significance in the traditional learning environment. Still, there is no significant difference observed between the overall performances in the traditional and online setting, and this is proven by very similar AUPRC values for both academic years (2019-2020 and 2020-2021).

4. Conclusion and Future Work

This paper introduced and analyzed *self-organizing map* models for unsupervised students' performance classification. The unsupervised

SOM model is used to encode a non-linear mapping for reducing the dimensionality of the input instances. For the experiments, real academic data sets collected from BBU from both online and traditional learning environments were used.

The research questions stated as the starting point of the present study were answered. The proposed experiments empirically confirmed the hypothesis that there are hidden patterns in academic data sets, patterns that would be relevant in students' performance prediction and the SOM model is able to detect these patterns based on unsupervised learning. Moreover, it is emphasized that there exists a high correlation between the results of the UL-based analysis and the efficiency of a supervised classifier trained for predicting the students' performance. A major goal of this study was to conduct an analytic comparison of students' performance, between learning influenced by online environment or digitalisation and learning in traditional academic environment. The obtained results highlight that there are no significant differences between the evaluations in the traditional and online learning environments. Still, the traditional evaluations seem to be a little more accurate than the evaluations in the online learning in terms of F-measure for the *grade-based* classification.

The similar performances of the supervised classifier trained on the two data sets revealed that the students from both the traditional and the online learning setting had similar learning patterns, independent of their learning competencies and skills. The relevance of these findings and the results of this study are not influenced by the variations in data due to the natural differences in the collectivities of subjects, from one year to the other. The data used in the present experiments are, still, of small-scale and will be further enlarged to obtain a better validation of the conclusions of this study.

Future work will be carried out in order to enlarge the data sets used in the experiments, to extend this study on data collected from grown-up education, as well as to include other unsupervised learning models in the present analysis e.g. *autoencoders*, *t-Distributed Stochastic Neighbor Embedding* (van

der Maaten & Hinton, 2008) or *gradual relational association rules* (Miholca & Czibula, 2019). In addition, the aim is to analyse the importance of progressive evaluation, like practical exams during the semesters or the assignments for each lecture/seminar.

REFERENCES

Bara, M. W., Ahmad, N. B., Modu, M. M. & Ali, H. A. (2018). Self-organizing map clustering method for the analysis of e-learning activities. In *2018 Majan International Conference (MIC)*, (pp. 1-5).

Boetticher, G. D. (2007). Improving Credibility of Machine Learner Models in Software Engineering, *Advances in Machine Learning Applications in Software Engineering,* 52-72. IGI Global.

Butnaru, G. I., Niță, V., Anichiti, A. & Brînză, G. (2021). The Effectiveness of Online Education during Covid 19 Pandemic - A Comparative Analysis between the Perceptions of Academic Students and High School Students from Romania, *Sustainability, 13*(9), 5311.

Crivei, L. M., Mihai, A. & Czibula, G. (2019). A study on applying relational association rule mining based classification for predicting the academic performance of students. In *KSEM 2019: The 12th International Conference on Knowledge Science, Engineering and Management, LNAI 11775*, (pp. 287-300).

Czibula, G., Maier, M. & Oneț-Marian, Z. (2021). *Students performance prediction data set - traditional vs. online learning* [Online]. Available at: <<https://figshare.com/articles/dataset/Educationaldatamining-traditionalvsonlinelearning/14330447/5>>, last accesed: 14 October 2021.

Czibula, G., Mihai, A. & Crivei, L. (2019a). SPRAR: A novel relational association rule mining classification model applied for academic performance prediction, *Procedia Computer Science, 159,* 20-29.

Czibula, G., Mihai, A., Mihuleț, E. & Teodorovici, D. (2019b). Using self-organizing maps for unsupervised analysis of radar data for nowcasting purposes. *Procedia Computer Science, 159,* 48-57.

Gijbels, I. & Hubert, M. (2009). Robust and Nonparametric Statistical Methods, *Comprehensive Chemometrics*, 189-211. Elsevier.

Gopal, R., Singh, V. & Aggarwal, A. (2021). Impact of online classes on the satisfaction and performance of students during the pandemic period of COVID 19, *Education and Information Technologies, 1-25.*

Khadir, S., Amanullah, K. & Shankar, P. (2015). Student's Academic Performance Analysis using SOM, *International Journal for Scientific Research and Development, 3,* 1037-1039.

Klingler, S., Wampfler, R., Käser, T., Solenthaler, B. & Gross, M. H. (2017). Efficient Feature Embeddings for Student Classification with Variational Auto-encoders. In *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017*, Wuhan, Hubei, China, June 25-28 (pp. 1-8).

Kurdthongmee, W. (2008). Utilization of a Self Organizing Map as a Tool to Study and Predict the Success of Engineering Students at Walailak University, *Walailak Journal of Science and Technology, 5,* 111-123.

Lee, Y. (2018). Using Self-Organizing Map and Clustering to Investigate Problem-Solving Patterns in the Massive Open Online Course: An Exploratory Study, *Journal of Educational Computing Research, 57*(2), 471-490.

Lötsch, J. & Ultsch, A. (2014). Exploiting the Structures of the U-Matrix, *Advances in Self-Organizing Maps and Learning Vector Quantization,* 249-257. Springer International Publishing.

Mihai, A., Czibula, G. & Mihuleț, E. (2019). Analyzing Meteorological Data Using Unsupervised Learning Techniques. In *ICCP 2019: IEEE 15th International Conference on Intelligent Computer Communication and Processing* (pp. 1-8). IEEE Computer Society.

Miholca, D. & Czibula, G. (2019). DynGRAR: A dynamic approach to mining gradual relational association rules, *Procedia Computer Science, 159,* 10-19.

Picca, D., Curdy, B. & Bavaud, F. (2006). Non-linear correspondence analysis in text retrieval: a kernel view. In *Proceedings of JADT 2006: 8es Journées Internationales d'Analyse Statistique des Données Textuelles* (pp. 741-747).

Saxena, K., Jaloree, S., Thakur, R. & Kamley, S. (2017). Self Organizing Map (SOM) based Modelling

Technique for Student Academic Performance Prediction, *International Journal on Future Revolution in Computer Science and Communication Engineering, 3*(9), 115-120.

Somervuo, P. & Kohonen, T. (1999). Self-organizing maps and learning vector quantization for feature sequences, *Neural Processing Letters, 10*, 151-159.

Subasi, A. (2020). Machine learning techniques, *Practical Machine Learning for Data Analysis Using Python*, 91-202. Academic Press.

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S. & Ragos, O. (2020). Transfer Learning from Deep Neural Networks for Predicting Student Performance, *Applied Sciences, 10*, 2145.

Tsipianitis, D. & Groumpos, P. (2018). University Asynchronous Distance Learning Programs to Enhance Interregional Sustainable Development, *IFAC-PapersOnLine, 51*(30), 346-351.

van der Maaten, L. & Hinton, G. (2008). Visualizing Data using t-SNE, *Journal of Machine Learning Research, 9*(2605), 2579-2605.