Received 9 June 2022, accepted 25 July 2022, date of publication 1 August 2022, date of current version 5 August 2022.
Digital Object Identifier 10.1109/ACCESS.2022.3195531

RESEARCH ARTICLE

IntelliDaM: A Machine Learning-Based Framework for Enhancing the Performance of Decision-Making Processes. A Case Study for Educational Data Mining

GABRIELA CZIBULA[1], GEORGE CIUBOTARIU[1], MARIANA-IOANA MAIER[1], AND HANNELORE LISEI[2]
[1]Department of Computer Science, Babeş-Bolyai University, 400084 Cluj-Napoca, Romania
[2]Department of Mathematics, Babeş-Bolyai University, 400084 Cluj-Napoca, Romania
Corresponding author: Gabriela Czibula (gabriela.czibula@ubbcluj.ro)

This work was supported by the Ministry of Research, Innovation and Digitization, CNCS/CCCDI–UEFISCDI, within PNCDI III, under Project PN-III-P4-ID-PCE-2020-0800.

ABSTRACT Nowadays, both predictive and descriptive modelling play a key role in decision-making processes in almost every branch of activity. In this article we are introducing *IntelliDaM*, a generic machine learning-based framework useful for improving the performance of data mining tasks and subsequently enhancing decision-making processes. Through its components designed for feature analysis, unsupervised and supervised learning-based data mining, *IntelliDaM* facilitates hidden knowledge discovery from data. Intensive research has been conducted in the field of *educational data mining*, as education institutions are interested in constantly adapting their educational programs to the needs of society by improving the quality of managerial decisions, course instructors' decision-making, or information gathering for course design. The present work conducts a longitudinal educational data mining study by applying *IntelliDaM* to real data collected at Babeş-Bolyai University, Romania, for a Computer Science course. The problem of mining educational data has been thoroughly examined using the proposed framework, with the goal of analysing students' performance. A very good performance has been achieved for the classification task (an $F1$ score of around 92%), and the results also highlighted a statistically significant performance improvement by using a technique for selecting discriminative data features. The performed study confirmed that *IntelliDaM* could be a useful instrument in educational environments, particularly for improving decision-making processes, like designing courses, the setup of efficient examinations, avoiding plagiarism, or offering support regarding stress management.

INDEX TERMS Data mining, educational data mining, machine learning, students' performance analysis, prediction.

I. INTRODUCTION

Applying *data mining* [1] techniques in order to extract meaningful knowledge from various data types is of great interest, being used for improving decision-making processes in various domains. *Machine learning* (ML) [2] offers a wide range of models and techniques for uncovering hidden patterns in data from numerous practical domains, such as *bioinformatics* (for protein dynamics analysis [3], [4]), *meteorology* (for precipitation nowcasting and radar data analysis [5], [6]), *software engineering* (for software structure analysis [7] and restructuring [8], aspect mining [9]),

The associate editor coordinating the review of this manuscript and approving it for publication was Vijay Mago.

medicine (for clinical decision support [10] and medical data analysis [11]), *computer vision* (for image analysis [12]), *educational data mining* (for academic data analysis [13], [14]), etc. *Educational data mining* (EDM) is a domain of research that applies data mining, ML, and statistics to data obtained from educational contexts. Thus, EDM evaluates educational data using computational techniques to investigate education-related issues [19]. For a better understanding of students, and the situations in which they learn in educational settings, EDM focuses on the development of techniques for examining unique data types.

In educational environments, data mining offers methods to support decision-making and thus provide decision support. Uncovering meaningful patterns and extracting knowledge from education-related data sets is a challenging and intensively investigated topic in the *educational data mining* (EDM) literature [15], particularly in the context of the COVID-19 pandemic [16], [17]. The main goals of EDM are to understand the students' learning process, predict students' learning outcomes, provide a better comprehension of the education-related phenomena [18] and help education institutions to understand and improve their education-related processes. Nowadays, academic institutions are more frequently interested in improving their teaching methodologies, learning processes [19] and the academic performance of their students and instructors [20]. EDM addresses techniques to understand the learning processes and identify patterns in data for supporting academic institutions in decision-making (regarding university admission [21] or the influence of students' performance during their years of college [22]).

Every education provider, or, generally speaking, every service provider, tries to offer suitable products to its beneficiaries. In this regard, providers must have an appropriate image of the clients' performance so that the offered products or services may be adapted according to these performances. Thus, an important direction in EDM is *student performance prediction* (SPP). SPP's target is to predict the grade of a student before attending a course or having an examination [23]. SPP problems require techniques from different domains: data mining, sociology, psychology, pedagogy, etc. [23]. There are some specific directions in the literature related to SPP: SPP of students at risk [24], [25], students' dropout prediction [26], [27], evaluation of students' performance [28], and remedial action plans [29]. According to the students' level or the educational field where SPP is needed, different features are considered to be "good predictors": grades, historical performance data [23], [30], students' demographic data [31], or students' behaviour [32] and engagement [33].

Given the rapid evolution of society, a paradigm shift in education is necessary. Thus, the education systems must consider the available tools so that this shift can bring benefits to students, instructors, and educational institutions. For instance, in the context of the COVID-19 pandemic, education shifted to *online environments*, and the education providers needed to change their *traditional learning* approach. The achievement of learning (in both traditional and online contexts) depends on the quality of teaching and the motivation of students. The teaching quality does not guarantee the students' motivation or vice versa because the latter depends on other factors, intrinsic or extrinsic [34]. In this context, there is an increasing interest in understanding how students are learning and how to improve their academic performance.

In order to enhance the performance of decision-making processes in educational environments there is an interest in designing ML-based frameworks for EDM purposes, including SPP. Such software tools would help uncover students' learning patterns from academic data sets, and finding their relevance and correlation with the students' performance. Therefore, we are proposing *IntelliDaM*, a machine learning-based framework for mining students' performance data. *IntelliDaM* offers three types of data analysis components designed for: (1) feature analysis and selection; (2) unsupervised learning-based data analysis; and (3) supervised learning-based predictive models. To evaluate the performance of *IntelliDaM*, we use real data collected at Babeş-Bolyai University, Romania, over three academic years, for a Computer Science (CS) discipline. Besides the proposed framework, the additional contributions envisaged by our study are: (1) to emphasise the effectiveness of *IntelliDaM* in analysing students' performance-related data; (2) to analyse and interpret, for the considered case study, the relevance of the patterns unsupervisedly mined from academic data and to show how these patterns are correlated with students' academic performance; and (3) to test whether or not the prediction of the students' final performance in a certain academic discipline is enhanced by their results achieved in previous CS courses from the curriculum. Even if it is empirically evaluated on academic data, the proposed *IntelliDaM* framework is a general one, and it may be applied to any data analysis task. To the best of our knowledge, the ML-based framework introduced in this paper, and the study conducted using *IntelliDaM* for students' performance analysis are new in the educational data mining literature.

To summarise the contributions of the study, we aim to answer the following research questions:

- **RQ1** How we can design a machine learning-based framework useful for enhancing the performance of data mining tasks and how effective is such a framework for mining students' performance-related information in a real case study? In this respect, the *IntelliDaM* framework will be introduced and analysed.
- **RQ2** How relevant are the students' learning patterns that were uncovered, through unsupervised learning, from the academic data sets, and to what extent are these patterns correlated with the students' academic performance?
- **RQ3** How important are the features used in learning and how do they impact the predictive performance? To what extent could the prediction of students' performance in a certain academic discipline be improved

by considering the students' results obtained in prerequisite courses from the academic curriculum as additional features in learning?

The rest of the paper is organised as follows. Section II is a brief incursion into the literature regarding the prediction and analysis of students' performance. Section III describes the methodology utilised for developing the *IntelliDaM* framework and presents its main components. Section IV presents the experimental evaluation of the proposed framework for an EDM case study, describing the data sets, the experiments, and their results. Section IV-D discusses the results obtained, whilst Section V presents the conclusions of the research and some ideas for future work.

II. RELATED WORK ON STUDENTS' PERFORMANCE ANALYSIS AND PREDICTION

Various contributions have been proposed in the EDM literature on students' academic performance, both from *supervised* and *unsupervised* learning perspectives. On the one hand, from a supervised learning-based viewpoint, *students' performance prediction* focuses on developing predictive models that are trained on historical data for estimating the students' future academic performance. *Students' performance analysis* (SPA), on the other hand, deals with developing, through unsupervised learning, descriptive models for determining how data are organised and for uncovering meaningful patterns in academic data sets.

Early research in SPP analysed various supervised learning classifiers: decision trees (DT) [35], Naïve Bayes, artificial neural networks [20], radial basis function networks [36], linear regression, support vector machines (SVMs) [37], and random forests (RF) [38]. For classification, a maximum F-score of 0.888 has been reported using a DT classifier [35], while for regression, a *root mean squared error* (RMSE) of 1.705 was obtained using SVMs [37]. More recently, supervised classifiers based on mining *relational association rules* (RARs) have been introduced [13], [39]. In terms of the *Area under the ROC curve* (AUC) measure, the maximum performance of 0.85 was reported for the RAR-based models.

Recent approaches in SPP reported values such as 0.94 for the F-score when using a DT classifier [33], and accuracies of 93.67% with RF [40] or 90.1% with *linear support vector machines* (LSVM) [41]. *Logistic Regression* (LR) obtains good results, with the AUC of 0.9541 and the accuracy values of 88.8% for predicting the students' status (passed/fail) and 68.7% for predicting the final grade of students [42]. Collaborative filtering, matrix factorization and *Restricted Boltzmann Machines* (RBM) are compared in a study related to SPP [43], where RBM achieves the best values for RMSE (0.3) and *mean absolute error* (0.23).

Besides the supervised classification and regression models, unsupervised learning methods have been of great interest in the analysis of the students' academic performance.

Clustering methods such as hierarchical agglomerative and partitional clustering were applied by Ayers *et al.* [44] for grouping students according to their skill sets. Parack *et al.* [45] applied the *k-means* clustering algorithm to identify groups of students having similar learning patterns with the aim of identifying related cognitive styles for each group. Other clustering methods such as Expectation-Maximization (EM) and Particle Swarm Optimization-based clustering [46] have been applied to discover students' profiles and patterns connected to their academic performance. *Self-organizing maps* (SOMs) were applied to classify students according to their study results [47] and to group students into categories (very good, good, average, poor) according to their academic performance [48].

Recent approaches in SPA addressed unsupervised RAR mining and SOMs [49], [50], to extract from academic data sets patterns and rules relevant for analysing students' academic performance. Other studies compare the ability of autoencoders and SOMs to find learning patterns in data sets related to students' performance in traditional and online environments [51], [52].

The recent EDM literature abounds in studies regarding online learning environments, such as the students' performance and satisfaction in online courses [16], their performance and behavior in Massive Online Open Courses (MOOC) [53], [54], and their learning profile in online learning environments [55].

III. METHODOLOGY

For the purpose of answering RQ1, this section introduces the methodology for developing the *IntelliDaM* framework and conducting our study on mining students' performance-related data. The main components of *IntelliDaM* are: (1) a feature analysis and selection component; (2) a component for unsupervised learning-based data analysis providing data visualisations and performance evaluation metrics; and (3) a component for supervised learning-based data mining providing performance evaluation for the predictive models.

An overview of the *IntelliDaM* framework is illustrated in Figure 1.

A. FORMALISATION

The theoretical model used in the paper for students' performance analysis is further detailed. We start by formalising the learning problem, then Section III-A1 and Section III-A2 detail the unsupervised and supervised learning approaches, respectively.

Let us consider a data set $\mathcal{S}t = \{stud_1, stud_2, \ldots, stud_k\}$ consisting of k instances, each instance $stud_i$ being represented as a vector characterising a student's performance in a certain academic course (discipline) $\mathcal{C}$, during an academic semester. As in a vector space model, the instances (students) are characterised by a list of attributes (features) $\mathcal{A} = \{a_1, a_2, \ldots, a_l\}$ available at the end of the teaching period and considered to be relevant for expressing the students' performance in the course $\mathcal{C}$ (e.g., attributes values may represent the grade received during the semester evaluations or the number of course/seminar/laboratory attendances, etc.). Thus, a student $stud_i$ is visualised as an l-dimensional vector

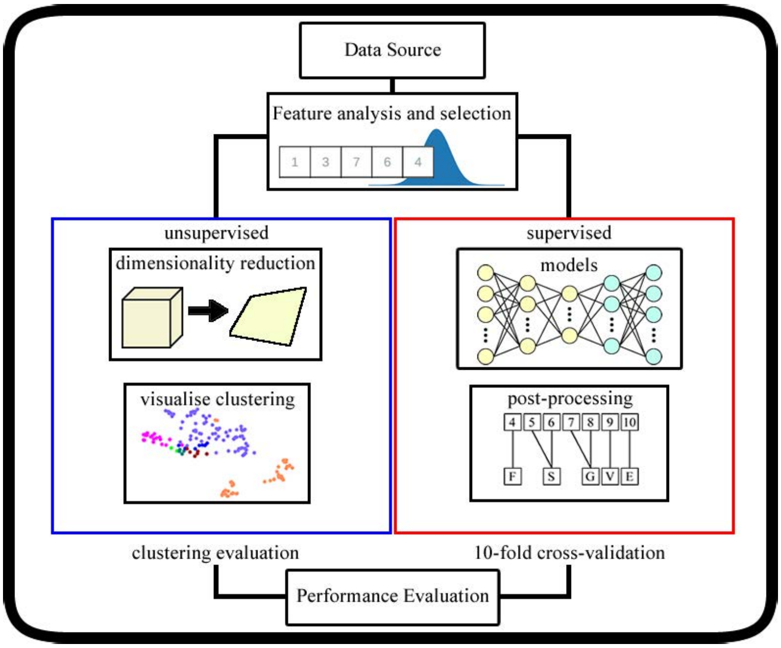


FIGURE 1. **Overview of *IntelliDaM* framework.**

$stud_i = (stud_{i1}, stud_{i2}, \ldots, stud_{il})$, where $stud_{ij}$ represents the value of attribute a_j for the student $stud_i$.

We note that for each student $stud_i$ from the data set $\mathcal{S}t$ the ground truth is available. The students are labelled with the *final grade* g_i they received in the course $\mathcal{C}$, but this label is not included in the feature vector representing the data instances. Let us denote by $\mathcal{G} = \{g_1, g_2 \ldots, g_p\}$ the set of labels available for the students from $\mathcal{S}t$, where p=7, ($\mathcal{G}$ contains grades from 4 to 10). As previously mentioned, these labels will be used only for evaluation purposes without being included in the learning process.

Usually, the *final grade* received by the students in a discipline is computed by considering their semester grades and other additional features, but also the grade received in a written or oral examination that takes place in the examination session, at the end of the academic semester. The SPP task is a difficult one, as we are trying to approximate a student's final performance (exact grade or a category of grades) based only on the features available at the end of the teaching period, but without knowing the student's future examination grade. The learner has to estimate the grade that the student will receive at the future exam, the learning hypothesis being difficult to search, due to uncertainties in both the students' learning and instructors' evaluation processes.

1) UNSUPERVISED LEARNING APPROACH

From an unsupervised learning perspective, the SPA problem may be formalised as searching for a hard partition $\{P_1, P_2, \ldots, P_n\}$ of the set $\mathcal{S}t$ of students (represented as l-dimensional vectors, as previously shown), such that $\mathcal{S}t = \bigcup_{i=1}^{n} P_i, \quad P_i \cap P_j = \emptyset \quad \forall 1 \leq i,j \leq n, \quad i \neq j$. A good partitioning is achieved when the students from each cluster (group) are highly similar to each other (with respect to their academic performance), while students from different clusters are dissimilar with respect to their academic performance.

The unsupervised learning task is specified as learning to approximate a target function $t : \mathcal{S}t \rightarrow \mathcal{C}at$, where $\mathcal{C}at = \{cat_1, cat_2, \ldots, cat_m\}$ represents a set of students' performance categories (classes). The learned hypothesis ($\hat{t} \approx t$) will assign to each student $stud \in \mathcal{S}t$ a category $\hat{t}(stud) \in \mathcal{C}at$. In the unsupervised learning setting, the unsupervised classification problem consists of searching for an approximation $\hat{t}$ of the target function such that $\hat{t}(stud)$ is as close as possible to the student's real performance category $t(stud)$, i.e. $\hat{t}(stud) \approx t(stud)$.

There are multiple ways to define the set $\mathcal{C}at$ of categories, the number of categories being specific to the education system from different countries [56]:

(i) **7 categories** ($m = 7$) The categories are expressing classes of grades: 10, 9, 8, 7, 6, 5, and ≤ 4 (this category corresponds to the failing students).

(ii) **5 categories** ($m = 5$) These categories, also used in the Romanian education system, are: *excellent* (E) (the class containing the students with grade 10); *very good* (V) (the class containing the students with grade 9); *good* (G) (the class including the students with the grades 7 and 8); *satisfactory* (S) (the class including the students with grades 5 and 6); and *fail* (F) (the class which includes the students with grades 4 and below 4).

Theoretically, the SPP performance should be increased when partitioning into larger categories (i.e., by considering a lower number of categories). This assumption will be empirically tested in Sections IV-C3 and IV-C4.

2) SUPERVISED LEARNING APPROACH

Let us denote by $\mathcal{G}$ the set of possible grades used for evaluating the students' performance, i.e., $\mathcal{G} = \bigcup_{i=4}^{10}\{i\}$. From a supervised learning perspective, the target function in our learning task is the mapping $t : \mathcal{S}t \rightarrow \mathcal{G}$ that assigns a grade $t(stud) \in \mathcal{G}$ to each student $stud \in \mathcal{S}t$.

The SPP problem considered in this paper is formalised as a regression problem, more specifically as the problem of learning the hypothesis $\hat{t}$ (approximation of the target function t) such that $\hat{t}(stud) \approx t(stud)$ ($\forall stud \in \mathcal{S}t$) with a certain degree of confidence. As an alternative, the SPP problem may be formalised as a supervised classification task, similar to the definition from the unsupervised learning setting. In the supervised classification setting the goal is to assign to each student not an exact grade, but a category (class) corresponding to the student's performance. Theoretically, the task of supervised classification of students in classes/categories should be easier than the regression task, in which the aim is to estimate the exact value of a student's grade.

In this paper, we do not investigate the SPP directly as a supervised classification task, but as a regression one. However, for evaluation purposes, the continuous values predicted by the regressor (the estimations for the final grades) will be discretised into categories (as it will be further shown in Section III-D1)), and thus supervised classification metrics may be employed, besides regression metrics, for performance evaluation.

B. FEATURE ANALYSIS AND SELECTION

As in any machine learning task, independent of the type of learning (supervised or unsupervised), the relevance of the features used for characterising the input instances is crucial for obtaining high performance. Theoretically, for a machine learning task, we would need features that are independent, and additionally, for supervised learning tasks, features highly correlated with the target output (in our case the final grade).

There are three main functionalities provided by the feature analysis component of *IntelliDaM*.

1) STATISTICAL-BASED FEATURE ANALYSIS

In order to study the correlation between each pair of features (a_i and a_j, for all $i \neq j$) and between each feature a_i and the target output ($g \in \mathcal{G}$), we will use Pearson and Spearman rank correlation coefficients [57]. Pearson correlation coefficient is used for measuring the degree of linear relationship between two features, while the Spearman rank correlation coefficient describes the strength and direction of the monotonic relationship between two variables.

Following [58], the strength of association of two variables by using one of these two correlation coefficients can be interpreted as shown in Table 1.

TABLE 1. Strength of association of two variables by using correlation coefficients [58].

Range of the absolute value of the correlation coefficient	Strength of association
[0, 0.2)	Very weak
[0.2, 0.4)	Weak
[0.4, 0.6)	Moderate
[0.6, 0.8)	Strong
[0.8, 1]	Very strong

2) FEATURE SELECTION

For selecting the most relevant (discriminative) features from a given feature set, we employ an extension of the Relief algorithm [59], namely ReliefF [60], which supports multi-class classification. Solving complex SPP tasks is expected to make use of features that interact with each other, due to consistent patterns observable through students' learning periods. Since the ReliefF algorithm is unaffected by such interactions, we decided to apply it as a statistical pre-processing step in our pipeline. Moreover, while Relief selects one positive and one negative closest neighbour of an instance for computing the features' relevance level, ReliefF chooses k such instances, which offers us a robust mean of selecting features. Eventually, the *skrebate* [61] implementation allows us to choose the number of desired features, rather than a relevance threshold.

3) FEATURE SETS QUALITY ANALYSIS

For expressing the relevance (predictive performance) of a feature/attribute set $\mathcal{A} = \{a_1, a_2, \ldots, a_l\}$, we have to estimate how "good" these features are for learning to differentiate the target output for the input instances. More specifically, for our case study, we have to measure how difficult it is to discriminate between the academic performance of students that are characterised by a specific feature set.

The **quality** of a feature set $\mathcal{A}$ characterising a data set $\mathcal{S}t$, denoted by $QF(\mathcal{A}, \mathcal{S}t)$, will be introduced for measuring how easy it is to predict the label (final grade) of an input instance (student) in a learning task in which the instances are described by the attributes from $\mathcal{A}$.

For a grade $g \in \bigcup_{i=4}^{10}\{i\}$, let us denote by $diff(g, \mathcal{S}t, \mathcal{A})$ the *difficulty* of predicting the grade g in the data set $\mathcal{S}t$ in which the instances are characterised by the attribute set $\mathcal{A}$. The value $diff(g, \mathcal{S}t, \mathcal{A})$ is defined as the ratio of students from the data set $\mathcal{S}t$ that are labelled with g and that have a nearest neighbour (1-NN, computed using the Euclidean distance) whose label differs from g. We note that $0 \leq diff(g, \mathcal{S}t, \mathcal{A}) \leq 1$ and lower values for the difficulty measure indicate that the prediction of grade g is easier for the data set $\mathcal{S}t$, given the feature set $\mathcal{A}$. A minimum value of 0 is obtained for the $diff(g, \mathcal{S}t, \mathcal{A})$ measure when all the students from $\mathcal{S}t$ that are labelled with a grade g have as 1-NN students with the same final grade g, while the maximum value of 1 is obtained when all students labelled with g have as 1-NN students with a grade $g' \neq g$.

Based on the above introduced *difficulty* values, the quality $QF(\mathcal{A}, \mathcal{S}t)$ is defined as in Formula (1)

$$QF(\mathcal{A}, \mathcal{S}t) = 1 - \frac{\sum_{g=4}^{10} \mathit{diff}(g, \mathcal{S}t, \mathcal{A})}{7}. \tag{1}$$

We note that $QF(\mathcal{S}t, \mathcal{A})$ ranges in [0, 1], with higher values suggesting that the problem of predicting the final grades for the students from the data set $\mathcal{S}t$ characterised by the feature set $\mathcal{A}$ is less difficult, i.e., the attribute set $\mathcal{A}$ is more "qualitative".

C. UNSUPERVISED LEARNING-BASED ANALYSIS

For better understanding the underlying patterns in the analysed data and for extracting meaningful knowledge from it, our framework uses two competitive dimensionality reduction methods, *t-Distributed Stochastic Neighbour Embedding* (t-SNE) [62] and *Uniform Manifold Approximation and Projection* (UMAP) [63]. Both of them perform very well and can be considered state-of-the-art methods for dimensionality reduction. They are based on neighbour graphs, which is typical for algorithms that focus more on the local structure, rather than on the global one. However, UMAP uses the cross-entropy loss, which preserves some of the global inter-cluster distances, while t-SNE's Kullback–Leibler (KL) divergence [64] focuses only on local distances. We consider that t-SNE's and UMAP's different approaches offer us a consistent comparison for data clustering and visualisation and will offer us a better comprehension of the data and the SPP task.

To further help visualise homogeneous structures, we are also going to make use of a *k-means* labelling step, which shall let us label the instances and compact the data into groups having similar characteristics, therefore decreasing the variance of such formations.

1) PERFORMANCE EVALUATION

External evaluation metrics are used in our experiments to evaluate the performance of the clustering models: *homogeneity*, *completeness*, and *V-measure* [65]. They are external evaluation metrics, as they require the ground truth of the instances, i.e. their correct labelling. All these measures range in [0, 1]. The homogeneity is maximised when all the clusters contain instances of a single class. The maximum value for completeness is obtained when all the instances of a class belong to the same cluster. Similar to the *F-score* metric used for evaluating the performance of supervised classifiers, the *V-measure* is defined as the harmonic mean between homogeneity and completeness.

D. SUPERVISED LEARNING-BASED ANALYSIS

IntelliDaM framework uses three regression models that implement different approaches:

1) The first model is the **Tweedie** regressor. Tweedie distributions are mainly used to generate a *Generalized Linear Model* (GLM). The GLMs include any distributions from the exponential family, hence they are able to fit relatively complex data sets.
2) The second regressor uses **Stochastic Gradient Descent** (SGD). It is a linear model fitted by minimizing an L2 loss [66] with SGD. This time, the stochasticity implied in the data fitting is expected to yield better results, since the model should be particularised relying on the data, rather than statistically fit a distribution to the data.
3) The last model that we label as **Poly** uses *polynomial and interaction features* [67] processed by a linear regressor. We expect that a linear model should have a lot more information now and result in better performance.

1) PERFORMANCE EVALUATION

As discussed in Section III-A2, from a supervised learning viewpoint, the SPP problem is approached as a regression problem, with the goal of learning a hypothesis $h : \mathcal{S}t \rightarrow \mathcal{G}$ that will provide for each student $stud \in \mathcal{S}t$ a grade $h(stud) \in \mathcal{G}$ (an estimation of the student's final performance in the course $\mathcal{C}$).

For evaluating the performance of the regression models, two evaluation measures will be used: RMSE and *normalized root mean squared error* (NRMSE) [68]. The RMSE on a given test set is defined as the square root of the obtained mean squared error (the error for a testing instance/student s is defined as the difference between the predicted grade and the real one). The NRMSE is the normalized RMSE obtained by dividing the value of RMSE by the range of the target output, in our case 6 (10-4), and is usually expressed as a percentage [69].

For a more comprehensive performance evaluation, the SPP regression task is transformed into a multi-class classification problem, by converting the real-valued outputs provided by the regressor (grades) into classes/categories (i.e., the set $\mathcal{C}at = \{cat_1, cat_2, \ldots, cat_m\}$ described in Section III-A1). Thus, evaluation metrics used for assessing the performance of multi-class classification may be used for the SPP regressor evaluation: **accuracy** (Acc), **precision** ($Prec$), **recall** ($Recall$), and **F-measure** ($F1$) [70]. Due to the imbalanced nature of the classification problem (i.e., there is an unequal distribution of classes/categories), $Prec$ is computed as the weighted mean of the precision values calculated for the classes cat_i ($\forall 1 \leq i \leq m$). Similarly, the overall $Recall$ and $F1$ values are determined as the weighted average of the metrics ($Recall$ and $F1$, respectively) computed for the classes.

The Pseudocode for the high-level implementation of the *IntelliDaM* framework is given in Algorithm 1.

IV. EXPERIMENTAL RESULTS

This section presents the experiments performed and the results obtained for evaluating the performance of the

Algorithm 1 Pseudocode for the High-Level Implementation of *IntelliDaM*

```
Input: St - the data set of students; A - the set of attributes describing the instances from St; G - the set of students' labels
       (grades)
F ← FeatureAnalysisAndSelection(St, A, G)            /* the feature analysis and selection component of
  IntelliDaM; F is the set of relevant features provided by the RelieFF algorithm */
if the chosen approach is unsupervised learning then
    embedding ← dimensionalityReduction(St)  /* the dimensionality reduction component using UMAP and
     t-SNE */
    partition ← clusteringVisualisation(embedding)          /* the clustering visualisation component for
     visualising the clusters from the low-dimensional input space; partition is the
     set of clusters */
    evaluatePerformance(partition)                             /* evaluate the clustering quality */
else
                                                  /* the chosen approach is supervised learning */
    for r ∈ {Tweedie, SGD, Poly } do
                                                           /* the supervised learning − based component */
        model_r ← buildRegressionModel(r, A, G)                      /* train the regression models */
        evaluatePerformance(model_r)        /* evaluate the performance of regression model */
    end
end
```

IntelliDaM framework on a case study for students' performance mining. The experimental evaluation, conducted following the methodology introduced in Section III, is aimed to answer the research questions RQ2 and RQ3.

A. DATA SETS

The experiments use real data sets, gathered from an undergraduate course, "*Logic and functional programming*" (LFP), held in the third semester for second-year students of the Faculty of Mathematics and Computer Science, Babeş-Bolyai University. The main objective of the LFP course is to introduce the declarative programming paradigm, specifically the logic and functional programming paradigms, and two programming languages specific to each of the paradigms: Prolog and Lisp, respectively.

Two data sets will be considered as case studies:

- $D_{2018-2020}$ - data collected for two academic years (2018-2019 and 2019-2020), when all teaching and evaluation activities were performed face-to-face.
- $D_{2020-2021}$ - data collected for the 2020-2021 academic year when all activities were moved online, due to the COVID-19 pandemic. We mention that the LFP course was designed for traditional learning, but was adapted in 2020-2021 to synchronous online learning because of the pandemic.

The data sets used to support the findings of this study are available at [71].

1) DESCRIPTION OF THE DATA SET $D_{2018-2020}$

For the $2018-2019$ and $2019-2020$ academic years there are 13 features characterising the students, i.e. $l = 13$. The features are as follows:

- $a_1, \ldots, a_7$ (labelled as L1-L7) are students' homework assignments prepared at home and presented during the *laboratory* classes, as shown in Table 2. The values for these features are grades between 0 and 10 (the students received the grade 0 if they did not turn in a lab assignment).

TABLE 2. Description of the features $a_1, \ldots, a_7$ for the academic years 2018-2020.

Feature	Label	Topic of the laboratory assignment
a_1	L1	"Recursive programming in C++"
a_2	L2	"Lists in Prolog"
a_3	L3	"Processing of heterogeneous lists in Prolog"
a_4	L4	"Backtracking in Prolog"
a_5	L5	"Recursive programming in Lisp (1)"
a_6	L6	"Recursive programming in Lisp (2)"
a_7	L7	"Using MAP functions in Lisp"

- a_8, a_9 (labelled as PE_1, PE_2) are *practical exams* (in Prolog and Lisp, respectively). These exams were held during the laboratories, when students had to solve, without using any help, a problem similar to their homework assignments. The values for the features are grades between 0 and 10 (the grade was 0 if a student did not participate in the practical exam).
- a_{10} (labelled as SA) represents the *seminar activity*. The students received from the seminar instructor a score for their activity during the seminars. The value for this feature ranges from 1 to 17.5.
- a_{11} (labelled as NA) represents the *number of attendances at the seminar* activity. The LFP course has 7 seminars, thus this feature has a value between 0 and 7.
- a_{12} (labelled as DSA) represents the "*Data structures and algorithms*" course, a prerequisite course for LFP.

DSA is a CS course held in the second semester and addresses the fundamental abstract data types, data structures and algorithms for manipulating them used in designing software applications. The values for this feature are the final grades (between 4 and 10) received by the students in the DSA course.

- a_{13} (labelled as FP) represents the "*Fundamentals of programming*" course, a prerequisite course for LFP. FP is a CS course held in the first semester and addresses the algorithms' design as well as fundamental concepts used in developing software applications. The Python programming language is used as a technical instrument for implementing the theoretical concepts in the FP course. The values for this feature are the final grades (between 4 and 10) received by the students in the FP course.

2) DESCRIPTION OF THE DATA SET $D_{2020-2021}$

For the 2020 − 2021 academic year, there are only 11 features characterising the instances ($l = 11$). Unlike the previous academic years, the practical exams (features a_8 and a_9 described in Section IV-A1) were not organised in the online setting. All the other features are the same as described in Section IV-A1.

As discussed in Section III-A, the data sets include, besides the features previously described, the label for each instance (target output). The labels are not included in the feature vectors characterising the instances and thus they are not taken into consideration when training the machine learning models. For our case studies, the label is the final grade received by the student for the LFP course (after the retake session). The final grade for the LFP course, according to the course syllabus, is 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. So, while the final grade depends on the values of the features, it is not determined by them.

Descriptive statistics about the data sets $D_{2018-2020}$ and $D_{2020-2021}$ are given in Table 3.

In the following, we will use the following notations:

- **LFP** denotes the set of features specific to the *Logic and functional programming* course, more specifically **LFP**$=\bigcup_{i=1}^{11}\{a_i\}$ for the data set $D_{2018-2020}$ and **LFP**$=\bigcup_{i=1}^{9}\{a_i\}$ for the data set $D_{2020-2021}$.
- **DSA** denotes the *Data structures and algorithms* course: attribute labelled as DSA (attribute a_{12} for the data set $D_{2018-2020}$ and attribute a_{10} the data set $D_{2020-2021}$).
- **FP** denotes the *Fundamental of programming* course: attribute labelled as FP (attribute a_{13} for the data set $D_{2018-2020}$ and attribute a_{11} the data set $D_{2020-2021}$).

B. EXPERIMENTS

With the goal of addressing research question RQ3 and investigating how the learning performance is impacted by the data features, experiments will be performed (on both data sets described in Section IV-A) using the following feature sets:

1) **LFP+DSA+FP**. This feature set consists of all the features described in Section IV-A characterising the students' performance: 13 features are used for the data set $D_{2018-2020}$ (Section IV-A1), while for the data set $D_{2020-2021}$ 11 features will be considered.
2) **LFP+FP**. This feature set consists of all features described in Section IV-A, except from the DSA feature (describing the grade in the DSA course, feature denoted by a_{12} for the data set $D_{2018-2020}$ and by a_{10} for the data set $D_{2020-2021}$).
3) **LFP+DSA**. This feature set consists of all the features described in Section IV-A, except from the FP feature (describing the grade in the FP course, feature denoted by a_{13} for the data set $D_{2018-2020}$ and by a_{11} for the data set $D_{2020-2021}$).
4) **LFP**. This feature set consists of all the features described in Section IV-A, excepting the DSA and FP features (describing the grades in the DSA and FP courses, feature denoted by a_{12} and a_{13} for the data set $D_{2018-2020}$ and by a_{10} and a_{11} for the data set $D_{2020-2021}$). Thus, the LFP set of features includes 11 features for the data set $D_{2018-2020}$ and 9 features for the data set $D_{2020-2021}$.
5) **Relief-based**. This feature set consists of the first k most relevant features obtained after applying the ReliefF algorithm [60]. The optimal value for k will be automatically determined using a grid search procedure, for each particular regressor. Details are provided in the experimental part.

For both data sets and all the feature sets previously described, the sets of grade categories presented in Section III-A1 will be used for the *unsupervised learning-based* analysis and for evaluating the performance of the *supervised regressors*.

C. RESULTS

This section presents the experimental results obtained by applying all the components of the *IntelliDaM* framework to the data sets described in Section IV-A. The experiments will be conducted according to the description from Section IV-B. The Section IV-C1 starts by detailing the experimental setup for the unsupervised and supervised learning experiments. Afterwards, the experimental results provided by the three components of *IntelliDaM* will be detailed.

1) EXPERIMENTAL SETUP

The *IntelliDaM* framework is implemented with the help of *scikit* 0.24.2 [67] machine learning library and *skrebate* 0.62 [61] library for feature selection. While experimenting with ReliefF, multiple values for the *n_neighbours* parameter

TABLE 3. Description of the used data sets.

Data set	# of features	# of instances	# of instances with grade						
			4	5	6	7	8	9	10
$D_{2018-2020}$	13	339	10	12	41	48	69	58	101
$D_{2020-2021}$	11	167	5	5	12	32	40	33	40

that decides the feature importance scores have been used. We are going to list the results for *n_neighbours* = 150, which yielded better performances.

For the unsupervised learning component, both methods used for dimensionality reduction output 2-dimensional data. On the one hand, t-SNE uses a perplexity of 15.0, an early exaggeration of 10.0, a learning rate of 150 and runs for 2000 iterations with random initialisation. On the other hand, UMAP uses the same number of neighbours as t-SNE, namely 15, and is initialised with a fuzzy spectral embedding.

For solving the SPP task using supervised regression, the models' fine-tuned parameters for our problem are as follows. The SGD regressor uses the squared error, penalizes the learning with an L2 norm, and has a *regularisation strength* (α) of 10^{-4}. The Tweedie regressor uses the normal distribution and an α of 1.0. Eventually, the *Poly regressor* uses a polynomial feature extraction of degree 5 and the linear regressor does not normalise the data, since it has been normalised beforehand. For assessing the performance of each regressor, a 10-fold cross-validation methodology is used.

2) FEATURE ANALYSIS

As discussed in Section III-B, the feature analysis stage starts with a statistical-based analysis, then a feature selection method is applied to determine the subset of relevant features. In the end, the quality of different feature (sub)sets is evaluated.

a: STATISTICAL-BASED FEATURE ANALYSIS

Figure 2 illustrates the Pearson and Spearman rank correlation coefficients between the features and the final grade, for each data set.

From this analysis, one can observe a moderate correlation between features and final grade. In the first data set, the minimum values of the correlations are obtained for the feature L1 (Pearson: 0.362 and Spearman: 0.341). This feature has a moderate/weak correlation in the second data set as well (Pearson: 0.411 and Spearman: 0.293). A possible reason for these correlations could be the fact that the feature L1 represents the first laboratory assignment where the requirement is to solve problems in a recursive manner in an imperative programming language (C++) and not in the declarative paradigm. In addition, unlike the other laboratory assignments, L1 has to be completed during the lab class and this may increase the difficulty of the assignment.

In $D_{2018-2020}$, the highest correlations are obtained for the DSA feature (Pearson: 0.665 and Spearman: 0.683) while in $D_{2020-2021}$, the maximum Pearson correlation value is 0.66, given by the feature L2 (the second laboratory) and the maximum Spearman value is 0.588 for the DSA feature. One can notice that this feature has a moderate Pearson correlation, too (0.583). A possible reason for these correlations between the LFP final grade and the DSA grade could be the fact that DSA is taught by the same professor, so students were familiarized to the teaching style and expectations from this course. Also, in the first plot, we can see two extra features, which represent the practical exams in $D_{2018-2020}$. We notice moderate correlations of these features with the final grade (Pearson: 0.58 and Spearman: 0.571 for the first practical exam; and Pearson: 0.526 and Spearman: 0.492 for the second practical exam). This fact highlights the importance of the practical exams during the semester.

Another observation is related to the correlations between the NA feature (number of seminar attendances) and the final grade. In the traditional learning environment, the correlation values are moderate (Pearson: 0.507 and Spearman: 0.477), while in the online environment they are weak (Pearson: 0.244 and Spearman: 0.248). A possible reason for this difference could be the social aspect of the traditional environment, where students develop social habits, which help them stay focused on the presented and discussed information. Unlike the face-to-face activities, the online environment is more focused on individual aspects (e.g., students may stop their microphone and webcam whenever they want), so many personal distractors may appear during the activities.

Pearson correlation and Spearman rank correlation coefficients between all pairs of features have been calculated for both data sets and most of the correlations were found very weak or weak. In the data set $D_{2018-2020}$, 66.667% of pairs of features have very weak or weak Pearson correlation (more exactly, 2.564% have very weak and 64.103% weak correlation); 66.667% of pairs of features have very weak or weak Spearman rank correlation (more exactly, 6.41% have very weak and 60.257% weak rank correlation). Moreover, in the data set $D_{2020-2021}$ 70.909% of pairs of features have very weak or weak Pearson correlation (more exactly, 18.182% have very weak and 52.727% weak correlation); 76.363% of pairs of features have very weak or weak Spearman rank correlation (more exactly, 12.727% have very weak and 63.636% weak rank correlation).

The weakest and strongest Pearson correlations between the features were computed. In the data set $D_{2018-2020}$, the lowest Pearson correlation between features is 0.196 between the features L7 (the last Lisp laboratory) and SA (the seminar activity) and the highest correlation of 0.627 is between the

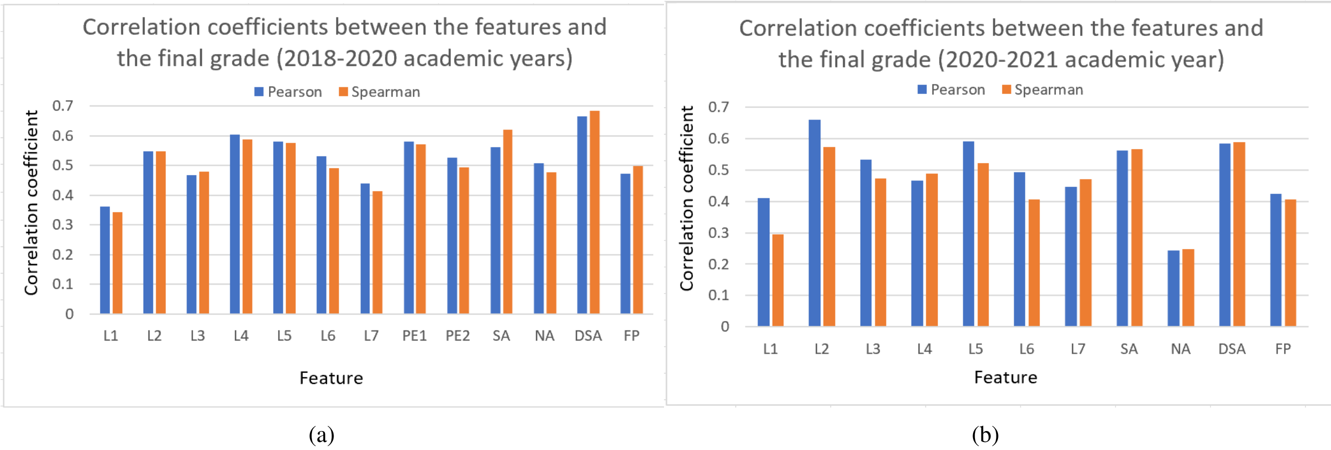


FIGURE 2. **Correlations between the features and the final grade for the data sets $D_{2018-2020}$ (a) and $D_{2020-2021}$ (b).**

features L5 and L6 (Lisp laboratories). The lowest Pearson correlation is 0.154 in $D_{2020-2021}$ between the features L4 (the last Prolog laboratory) and L6 (the second Lisp laboratory), while the highest Pearson correlation is 0.712 between the features L2 and L3 (Prolog laboratories). The weakest and strongest Spearman rank correlations were calculated: 0.166 (between features L1 and PE_2) and 0.603 (between features L4 and L5) in the data set $D_{2018-2020}$, 0.094 (between features L1 and PE_1) and 0.551 (between features SA and NA) in the data set $D_{2020-2021}$. We observe that in the online teaching environment the correlations are weaker than in the traditional case; there is less correlation between the features.

These correlations between the features highlight the connection between Prolog laboratories (L2-L4), respectively between Lisp laboratories (L5-L7). A higher correlation between laboratory assignments focused on the same declarative paradigm (logic - Prolog laboratories, or functional - Lisp laboratories) was expected. Also, a weak correlation between laboratory requirements on different declarative languages (as observed in the online case, between features L4 and L6) is not unusual. A possible reason for the weak correlation found between L7 and SA in the traditional learning environment could be the fact that L7 is the last homework, and it can't be recovered by students who missed it, even if they had good results during the entire semester at the seminar activity.

b: FEATURE SELECTION

The list of features, in a decreasing order of their relevance, as provided by the ReliefF algorithm [60] are given in Table 4.

From Table 4 we observe the following:

- For both data sets, the first three most relevant features are the grades received at the DSA and FP courses and the seminar activity (SA). In addition, the number of seminar attendances (NA) is the fifth most relevant feature in the traditional learning setting ($D_{2018-2020}$), while for the online setting ($D_{2020-2021}$) it is the sixth most relevant feature.

TABLE 4. **The list of features, in a decreasing order of their relevance, provided by the ReliefF algorithm [60].**

Data set	List of features (decreasing order of their relevance)
$D_{2018-2020}$	DSA, FP, SA, L4, NA, PE_1, PE_2, L5, L6, L2, L1, L3, L7
$D_{2020-2021}$	SA, DSA, FP, L2, L1, NA, L5, L3, L4, L6, L7

- In the traditional learning setting, the practical exams (PE_1, PE_2) are more relevant than the laboratory grades (L1-L7).
- The least relevant feature, for all data sets, is L7.

c: FEATURE SETS QUALITY ANALYSIS

The **quality** (QF measure introduced in Section III-B) of the feature sets described in Section IV-B is illustrated in Table 5. The quality of the **Relief-based** feature set (Table 4) was computed for 11 features for data set $D_{2018-2020}$, and 9 features for the data set $D_{2020-2021}$.

For the data set $D_{2018-2020}$, Table 5 reveals very similar QF values (around 0.3) for all feature sets (the **Relief-based** feature set is slightly outperformed, in terms of QF, by the other feature sets). For the data set collected in the online learning environment ($D_{2020-2021}$), the **Relief-based** feature set appears to be the most relevant (has the maximum QF value).

3) UNSUPERVISED LEARNING-BASED ANALYSIS

We further present the results of the unsupervised learning-based analysis following the methodology introduced in Section III-A1. Figures 3 and 4 illustrate the t-SNE and UMAP visualisations of the data sets $D_{2018-2020}$ and $D_{2020-2021}$, respectively. For increasing the readability, each instance from the plots is labelled with its ground truth label (final grade, ranging between 4 and 10).

According to the methodology introduced in Section III-C, in order to analyse the data sets, we are going to investigate the unsupervised classification into seven classes

TABLE 5. Values for the quality measure QF computed for the considered data sets and for various feature sets.

Data set	Feature set				
	LFP+DSA+FP	LFP+FP	LFP+DSA	LFP	Relief-based
$D_{2018-2020}$	0.32	0.32	0.31	0.31	0.30
$D_{2020-2021}$	0.27	0.25	0.26	0.26	0.33

(**7 categories**, using 7 ground truth classes corresponding to grades from 4 to 10) and five classes (**5 categories**, using 5 compacted classes: 4, 5-6, 7-8, 9, 10), as described in Section III-A1. The use of 5 instead of 7 classes slightly reduced the difficulty of the classification problem, since the results provided us heterogeneous manifolds and few well-delimited clusters.

For computing the metrics for external clustering evaluation, a k-means labelling step is used in which all instances partitioned into the same cluster are labelled with the same class/category. The k-means labelling method tends to preserve most of the classes' distribution, to correctly label the easy instances, to make the inter-class transition smoother while underlining the complex and noisy nature of the data. The external evaluation measures (described in Section III-C1) computed for the k-means partitioning (into 7 categories) of the 2D data points provided by t-SNE and UMAP methods are summarized in Table 6.

From Table 6 we notice that for both data sets, all the external evaluation measures have greater values for t-SNE than for UMAP. The performance of the partitioning on $D_{2020-2021}$ is slightly greater than the one on $D_{2018-2019}$, and this is visible in Figure 4, where we observe a slightly better grouping of students with similar grades than in Figure 3.

4) SUPERVISED LEARNING-BASED ANALYSIS

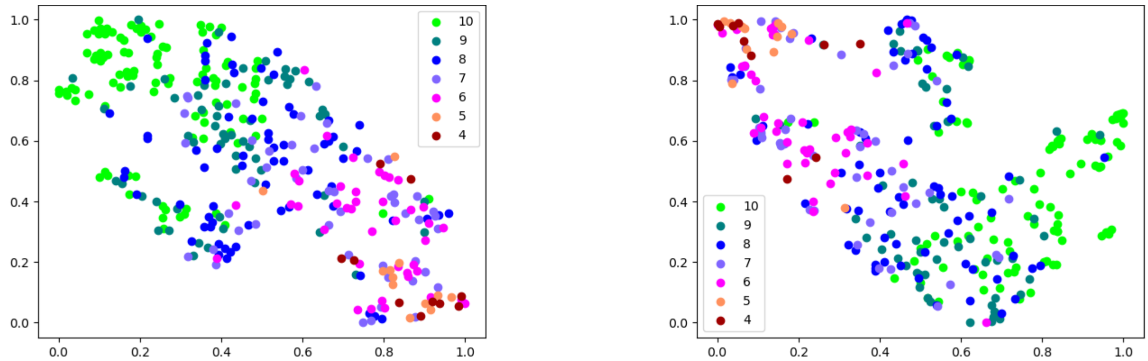
Table 8 illustrates the results of applying the supervised learning-based methodology introduced in Section III-D on the data sets described in Section IV-A. The experiments are conducted as shown in Section IV-B. The average values for the performance metrics obtained during the cross-validation (Section IV-C1) are provided. For each data set, the best values for the performance metrics and the best feature set are highlighted.

For determining the **Relief-based** set of features for a data set and a specific regressor (i.e., the first k most relevant features provided by the ReliefF algorithm [60]), a grid search has been applied for various k. For a specific subset of features (the first k features provided by the **ReliefF** algorithm), a 10-fold cross-validation was applied to evaluate the performance of the classification (into 5 and 7 categories). The optimal value for k is determined as the one that maximizes the performance during the 10-fold validation process. Table 7 presents the values for k determined after a grid search applied on both data sets, using the considered regressors.

Analysing the results of the supervised learning experiments presented in Table 8, one observes the following:

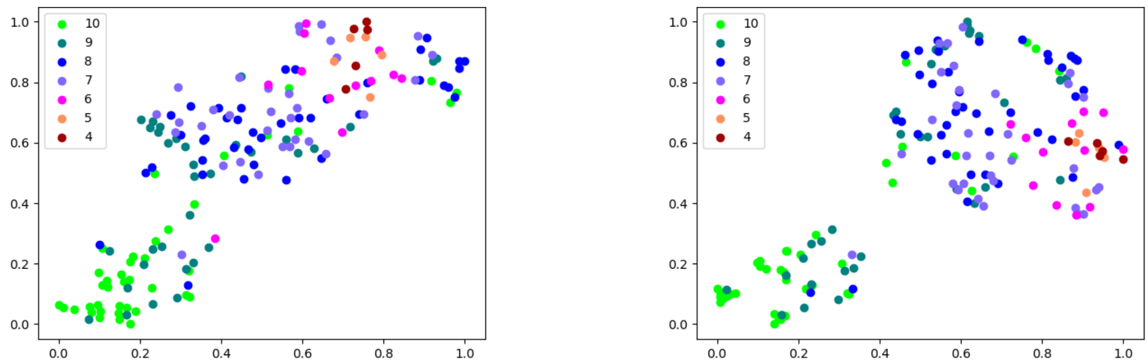
- The performance of the classification into 5 categories (in terms of all evaluation metrics: *Acc*, *Prec*, *Recall*, $F1$) is better than the one of classification into 7 categories. This empirically confirms the hypothesis discussed in Section III-A1, namely that it is "easier" to classify the students in larger categories/classes.
- The best performances on both data sets and for all feature sets are provided by the **Poly** regressor.
- The **Relief-based** feature set provides the best performance for the classification tasks, for both data sets. Only for the data set $D_{2020-2021}$, the regression performance, in terms of RMSE, for the Relief-based feature set is slightly outperformed by the feature set **LFP+DSA+FP**. A high performance for the Relief-based feature set was expected, as the quality QF of this feature set exceeded the quality of the other feature sets. However, only a slight increase in performance (less than 1% for $F1$ measure) is observed for the **Relief-based** set of features, compared to the original set of features (**LFP+DSA+FP**).
- Around 4% improvement of $F1$ measure is noticed when using the enhanced feature set **LFP+DSA+FP** instead of the **LFP** feature set. This suggests that the students' performance in the LFP course may be increased by enlarging the feature set with the grades received in the DSA and FP courses.
- High classification performances are obtained for both data sets and **Relief-based** feature selection ($F1$ values higher than 92%). The performances for the classification tasks applied on the data set collected from *traditional learning* environment ($D_{2018-2020}$) is comparable with that of classifying students' performance on the *online learning* data set ($D_{2020-2021}$). This suggests that the students' learning patterns in both learning environments were highly similar. The $F1$ values on $D_{2018-2020}$ slightly exceeded (with less than 1%, for both classification tasks) the values obtained on $D_{2020-2021}$. A possible explanation may be the fact that the feature set for traditional learning included the practical exams that were not organized in the online setting.

We remark the following regarding the performance of the **Poly** regressor on the data sets characterised by the **Relief-based** feature set. On both data sets ($D_{2018-2020}$ and $D_{2020-2021}$), the regressor succeeded to accurately predict the classes of grades 5-10 (with a maximum of 95.65% for grade 6 in the online setting and a minimum of 91% for the grade 5 in the traditional setting). For both data sets, the hardest to predict is the grade 4 (an $F1$ of only 53% is obtained for the data set $D_{2018-2020}$, while for the data set $D_{2020-2021}$ the $F1$ value increases to 77%). A possible



(a) (b)

FIGURE 3. t-SNE (a) and UMAP (b) visualisations of the data set $D_{2018-2020}$.



(a) (b)

FIGURE 4. t-SNE (a) and UMAP (b) visualisations of the data set $D_{2020-2021}$.

TABLE 6. Performance metrics for the 7 classes k-means clustering applied on the 2D t-SNE and UMAP data points. The data sets are characterised by the entire feature set (LFP+DSA+FP).

Data set	Method	External evaluation		
		Homogeneity	*Completeness*	*V-measure*
$D_{2018-2020}$	**t-SNE**	**0.291**	**0.314**	**0.302**
	UMAP	0.267	0.285	0.276
$D_{2020-2021}$	**t-SNE**	**0.309**	**0.328**	**0.318**
	UMAP	0.307	0.316	0.311

TABLE 7. Optimal number of features from the Relief-based feature set determined on the data sets using various regressors.

Data set	Regressor	Classification	
		7 categories	5 categories
$D_{2018-2020}$	**Tweedie**	10	10
	SGD	10	10
	Poly	11	11
$D_{2020-2021}$	**Tweedie**	9	7
	SGD	10	10
	Poly	9	10

explanation for the difficulty to predict the "fail" class could be the unpredictable circumstances in the exam day, the weak stress management during the exam, or the students' plagiarism during the semester.

D. DISCUSSION

The experimental results of the unsupervised and supervised learning experiments performed in Sections IV-C3 and IV-C4 are further discussed with the goal of answering the research questions RQ2 and RQ3.

The performance of the unsupervised learning models presented in Section IV-C3 highlighted that students' learning patterns that are correlated with their academic performance may be unsupervisedly uncovered from the academic data sets. From the clustering visualisations, we can observe that students with similar performances are grouped together, thus it is very likely that they might have similar learning patterns. We also remark (from Table 6 and the plots from Figures 3 and 4) that the t-SNE dimensionality reduction provided a better partitioning of the students.

TABLE 8. Experimental results. The average for the evaluation measures obtained during 10-fold cross-validation are provided.

Regressor	Data set	Feature set	Regression		Classification: 7 categories				Classification: 5 categories			
			RMSE	NRMSE (%)	*Acc* (%)	*Prec* (%)	*Recall* (%)	*F*1 (%)	*Acc* (%)	*Prec* (%)	*Recall* (%)	*F*1 (%)
Tweedie	$D_{2018-2020}$	**LFP+DSA+FP**	0.938	15.6	37.75	56.37	37.75	36.49	48.96	68.03	48.96	47.86
		LFP+FP	0.942	15.7	35.98	55.75	35.98	34.50	47.19	67.41	47.19	45.86
		LFP+DSA	0.985	16.4	33.62	55.96	33.62	30.33	44.83	67.38	44.83	41.68
		LFP	1.000	16.6	32.74	52.03	32.74	28.63	43.36	63.45	43.36	39.71
		Relief-based	0.926	15.4	38.64	56.80	38.64	37.76	49.55	68.03	49.55	48.81
	$D_{2020-2021}$	**LFP+DSA+FP**	0.982	16.3	29.94	20.69	29.94	22.45	50.29	41.65	50.29	43.87
		LFP+FP	0.972	16.2	32.33	23.19	32.33	24.80	52.09	42.72	52.09	45.18
		LFP+DSA	0.987	16.4	32.33	23.74	32.33	24.27	51.49	40.99	51.49	44.00
		LFP	1.001	16.6	31.13	23.10	31.13	23.45	49.70	39.93	49.70	42.65
		Relief-based	0.989	16.4	33.53	25.12	33.53	26.30	52.09	39.60	52.09	44.10
SGD	$D_{2018-2020}$	**LFP+DSA+FP**	0.882	14.7	46.31	56.04	46.31	47.60	57.52	66.87	57.52	58.91
		LFP+FP	0.870	14.5	43.95	54.15	43.95	45.19	55.75	65.85	55.75	57.19
		LFP+DSA	0.903	15	44.83	53.67	44.83	46.20	56.93	65.65	56.93	58.69
		LFP	0.917	15.2	40.70	49.25	40.70	42.28	53.68	62.43	53.68	55.56
		Relief-based	0.871	14.5	46.60	56.01	46.60	47.82	58.11	67.41	58.11	59.55
	$D_{2020-2021}$	**LFP+DSA+FP**	0.864	14.4	49.10	52.61	49.10	49.26	63.47	65.19	63.47	63.36
		LFP+FP	0.893	14.8	47.90	50.58	47.90	47.68	62.87	63.30	62.87	61.92
		LFP+DSA	0.928	15.4	43.71	47.35	43.71	42.07	62.27	63.29	62.27	60.74
		LFP	0.932	15.5	43.11	44.56	43.11	40.67	62.27	61.28	62.27	60.61
		Relief-based	0.846	14.1	49.10	53.21	49.10	49.44	63.47	65.46	63.47	63.40
Poly	$D_{2018-2020}$	**LFP+DSA+FP**	0.361	6	92.03	95.23	92.03	93.11	92.03	95.23	92.03	93.12
		LFP+FP	0.483	8	89.38	91.08	89.38	89.84	89.67	91.41	89.67	90.15
		LFP+DSA	0.476	7.9	88.79	90.62	88.79	89.38	89.38	91.24	89.38	89.99
		LFP	0.495	8.2	87.61	89.25	87.61	87.94	88.20	89.94	88.20	88.57
		Relief-based	**0.330**	**5.5**	**92.92**	**95.65**	**92.92**	**93.79**	**92.92**	**95.65**	**92.92**	**93.80**
	$D_{2020-2021}$	**LFP+DSA+FP**	**0.212**	**3.5**	92.21	93.10	92.21	92.35	92.81	93.61	92.81	92.93
		LFP+FP	0.215	3.5	92.21	92.87	92.21	92.21	92.21	92.90	92.21	92.24
		LFP+DSA	0.382	6.3	90.41	91.30	90.41	90.68	90.41	91.30	90.41	90.70
		LFP	0.340	5.6	88.62	90.16	88.62	89.00	88.62	90.13	88.62	89.01
		Relief-based	0.264	4.4	**92.81**	**93.88**	**92.81**	**92.95**	**92.81**	**93.89**	**92.81**	**93.04**

The *V-measure* values (around 0.3) suggest that unsupervised learners, using the **LFP+DSA+FP** feature set, are not able to detect well-separated clusters of students. However, a statistically significant improvement (at a significance level of $\alpha = 0.05$, using a two-tailed paired Wilcoxon signed-rank test [72], [73]) was observed in the clustering of students from the data set collected in the online learning environment.

The results of the unsupervised learning-based analysis from Section IV-C3 have been strengthened in Section IV-C4 by the supervised learning experiments conducted for students' performance prediction. As shown in Section IV-C4, the supervised regressors, unlike the unsupervised learning models, succeeded to detect better separation boundaries between the students' categories. This was expectable, as the learning task is easier in a supervised learning scenario when feedback is provided to the learner in the form of training examples. The prediction performance was slightly better (an average below 1%) on the first data set ($D_{2018-2020}$) for the **Relief-based** feature set and the **Poly** regressor.

The importance of the feature sets used in the learning process was highlighted by evaluating the performance of various regression models using various feature sets characterising the students. The results depicted in Table 8 emphasize improvements in performance when using the **Relief-based** feature set compared to the **LFP** and **LFP+DSA+FP** feature sets. The statistical significance of the observed differences has been tested by using a two-tailed paired Wilcoxon signed-rank test [72], [73]. The sample of performance measures values obtained for all experiments presented in Table 8 when using the **Relief-based** feature set has been tested against the sample of performance values obtained for the **LFP** and **LFP+DSA+FP** features, respectively. *p*-values less than 0.01 were obtained, confirming that between the two performance samples there is a statistically significant difference at a significance level of $\alpha = 0.01$. Hence, the obtained results are statistically significant. Thus, RQ3 has been answered, due to the statistically significant performance improvement achieved by the **Relief-based** feature set that includes both the DSA and FP features. RQ2 may be answered as well: the good performance achieved in the SPP task confirms that the patterns uncovered by the unsupervised learning models are strongly correlated with the students' academic performance.

The DSA and FP features proved to be very relevant for predicting the students' performance in the LFP course. Thus, it is empirically confirmed that the background knowledge provided by the DSA and FP courses is essential for a good understanding of the concepts from LFP. The FP course offers the fundamental concepts of programming like modularisation, recursion, backtracking, testing, etc. In the FP course, these concepts are implemented in Python, which is a multi-paradigm programming language close to the functional paradigm. In the DSA course, students learn about various data structures (arrays, linked lists, hash tables, trees) used for implementing data containers and algorithms for manipulating these data structures. Even if in the DSA course the concepts are implemented in C++ programming language, which is an imperative programming language, it seems that the algorithmic skills acquired in this course are fundamental for comprehending and thoroughly understanding the declarative programming paradigm discussed in the LFP course.

The experiments performed revealed that the features related to practical exams (PE1 and PE2 in the data set $D_{2018-2020}$) have more relevance than the features regarding laboratory grades. The explanation could be that practical exams represent a type of long-term evaluation, so they reveal the consolidated knowledge of students, being at the end of a learning unit. The laboratory grades are short-term evaluations and indicate to what extent a student has solved the homework correctly. Thus, the students can get help in their work (e.g. internet or fellowship resources), but in the practical exams, they have to work on their own.

In what concerns the comparison between the students' performance in the *traditional* learning environment ($D_{2018-2020}$) and the *online* one, we observe that most of the results indicate better performance metrics values for $D_{2018-2020}$ than for $D_{2020-2021}$. Some reasons could be:

- $D_{2018-2020}$ data set has more instances than $D_{2020-2021}$ data set;
- in the 2020-2021 academic year, the LFP course was taught for the first time in an online environment, and it was a challenge for all, instructors and students, to adapt to that new situation;
- students' monitoring is more difficult in online environments, and this was proved by the differences between the relevance of the NA feature (seminar attendances) and its moderate/weak correlation with final grade;
- in traditional learning, practical exams were held, so students had to learn for evaluations during the semester, not only before the final exam;
- when homework is checked in online, the students can cheat with their answers at screening questions.

V. CONCLUSION AND FUTURE WORK

The research previously presented introduced a machine learning-based framework, *IntelliDaM*, which includes components for feature analysis, unsupervised and supervised learning-based mining, and is useful for enhancing the performance of data mining tasks. The effectiveness of the framework is evaluated and analysed in an EDM case study consisting of real data collected at the Babeş-Bolyai University, Romania, over three academic years, for a Computer Science (CS) discipline.

The proposed *IntelliDaM* framework is easily configurable and it may be applied to data mining in various application domains (e.g., bioinformatics, software engineering, medicine, meteorology, etc.). It provides automatised data exporting processes, well-determined pipeline steps, multiple entry points to reduce the experiments' duration and favours traceable and reproducible results.

The research questions stated in Section I were answered. The experimental evaluation highlighted that the results provided by the unsupervised and supervised learning-based components of the *IntelliDaM* in the considered case study are strongly correlated. The results revealed that the students' learning patterns may be uncovered through machine learning models, and that these patterns are useful for predicting their academic performance. Additionally, it has been shown that the students' performance prediction for a certain discipline may be improved by including in the learning process additional features representing their results from previous CS courses.

The study carried out in this paper using *IntelliDaM* confirmed that the proposed framework could be a useful instrument in educational environments, especially in decision-making processes, such as the choice of the topics in designing or updating courses, the setup of efficient examinations, the prevention of plagiarism, or the solutions regarding stress management.

Future work will be conducted to analyse SPP from middle and high schools in connection with to the scholars' results in the national exams. We aim to further apply the *IntelliDaM* framework to data mining tasks from other application domains such as meteorology, software defects detection, marketing and advertising, gaming, or sports.

ACKNOWLEDGMENT

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. The publication of this article was partially supported by the 2021 Development Fund of the Babeş-Bolyai University, Cluj-Napoca (Romania).

REFERENCES

[1] I. Matloob, S. A. Khan, R. Rukaiya, M. A. K. Khattak, and A. Munir, "A sequence mining-based novel architecture for detecting fraudulent transactions in healthcare systems," *IEEE Access*, vol. 10, pp. 48447–48463, 2022.

[2] P. Hegedus and R. Ferenc, "Static code analysis alarms filtering reloaded: A new real-world dataset and its ML-based utilization," *IEEE Access*, vol. 10, pp. 55090–55101, 2022.

[3] G. Czibula, C. Codre, and M. Teletin, "AnomalP: An approach for detecting anomalous protein conformations using deep autoencoders," *Expert Syst. Appl.*, vol. 166, Mar. 2021, Art. no. 114070.

[4] M. Teletin, G. Czibula, M.-I. Bocicor, S. Albert, and A. Pandini, "Deep autoencoders for additional insight into protein dynamics," in *Artificial Neural Networks and Machine Learning—ICANN 2018*. Cham, Switzerland: Springer, 2018, pp. 79–89.

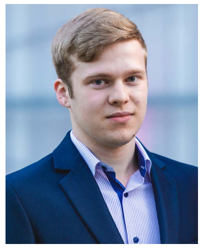
[5] A. Mihai, G. Czibula, and E. Mihuletc, "Analyzing meteorological data using unsupervised learning techniques," in *Proc. IEEE 15th Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, Sep. 2019, pp. 1–8.

[6] G. Czibula, A. Mihai, A.-I. Albu, I.-G. Czibula, S. Burcea, and A. Mezghani, "AutoNowP: An approach using deep autoencoders for precipitation nowcasting based on weather radar reflectivity prediction," *Mathematics*, vol. 9, no. 14, p. 1653, Jul. 2021. [Online]. Available: https://www.mdpi.com/2227-7390/9/14/1653

[7] G. Czibula, I. G. Czibula, and R. D. Gaceanu, ''A support vector machine model for intelligent selection of data representations,'' *Appl. Soft Comput.*, vol. 18, pp. 70–81, May 2014.
[8] Z. Marian, I.-G. Czibula, and G. Czibula, ''A hierarchical clustering-based approach for software restructuring at the package level,'' in *Proc. 19th Int. Symp. Symbolic Numeric Algorithms Sci. Comput. (SYNASC)*, Sep. 2017, pp. 239–246.
[9] G. S. Cojocar and G. Czibula, ''On clustering based aspect mining,'' in *Proc. 4th Int. Conf. Intell. Comput. Commun. Process.*, Aug. 2008, pp. 129–136.
[10] G. Czibula, I. G. Czibula, G. S. Cojocar, and A. M. Guran, ''IMASC—An intelligent MultiAgent system for clinical decision support,'' in *Proc. 1st Int. Conf. Complex. Intell. Artif. Natural Complex Syst. Med. Appl. Complex Syst. Biomed. Comput.*, Nov. 2008, pp. 185–190.
[11] S. Nitica, G. Czibula, and V.-I. Tomescu, ''A comparative study on using unsupervised learning based data analysis techniques for breast cancer detection,'' in *Proc. IEEE 14th Int. Symp. Appl. Comput. Intell. Informat. (SACI)*, May 2020, pp. 99–104.
[12] G. Ciubotariu, V.-I. Tomescu, and G. Czibula, ''Enhancing the performance of image classification through features automatically learned from depth-maps,'' in *Computer Vision Systems*, M. Vincze, T. Patten, H. I. Christensen, L. Nalpantidis, and M. Liu, Eds. Cham, Switzerland: Springer, 2021, pp. 68–81.
[13] G. Czibula, A. Mihai, and L. M. Crivei, ''S PRAR: A novel relational association rule mining classification model applied for academic performance prediction,'' *Proc. Comput. Sci.*, vol. 159, pp. 20–29, Jan. 2019.
[14] H. Mardesci, ''The effect of online learning on university students' learning motivation,'' *JPP (Jurnal Pendidikan dan Pembelajaran)*, vol. 27, no. 1, pp. 42–47, Sep. 2020.
[15] L. Chen, P. Chen, and Z. Lin, ''Artificial intelligence in education: A review,'' *IEEE Access*, vol. 8, pp. 75264–75278, 2020.
[16] R. Gopal, V. Singh, and A. Aggarwal, ''Impact of online classes on the satisfaction and performance of students during the pandemic period of COVID 19,'' *Educ. Inf. Technol.*, vol. 26, no. 6, pp. 6923–6947, 2021.
[17] C. Coman, L. G. Tîru, L. Meseşan-Schmitz, C. Stanciu, and M. C. Bularca, ''Online teaching and learning in higher education during the coronavirus pandemic: Students' perspective,'' *Sustainability*, vol. 12, no. 24, p. 10367, Dec. 2020.
[18] A. Bogarín, R. Cerezo, and C. Romero, ''A survey on educational process mining,'' *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 1, p. e1230, Jan. 2018, doi: 10.1002/widm.1230.
[19] S. K. Mohamad and Z. Tasir, ''Educational data mining: A review,'' *Proc. Social Behav. Sci.*, vol. 97, pp. 320–324, Nov. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042813036859
[20] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, ''Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique,'' *Decis. Anal.*, vol. 2, no. 1, pp. 1–25, Mar. 2015.
[21] H. A. Mengash, ''Using data mining techniques to predict student performance to support decision making in university admission systems,'' *IEEE Access*, vol. 8, pp. 55462–55470, 2020.
[22] A. I. Adekitan and O. Salau, ''The impact of engineering students' performance in the first three years on their graduation result using educational data mining,'' *Heliyon*, vol. 5, no. 2, Feb. 2019, Art. no. e01250. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S240584401836924X
[23] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, ''Educational data mining techniques for Student performance prediction: Method review and comparison analysis,'' *Frontiers Psychol.*, vol. 12, Dec. 2021, Art. no. 842357, doi: 10.3389/fpsyg.2021.698490.
[24] K. T. Chui, D. C. L. Fung, M. D. Lytras, and T. M. Lam, ''Predicting at-risk university students in a virtual learning environment via a machine learning algorithm,'' *Comput. Hum. Behav.*, vol. 107, Jun. 2020, Art. no. 105584. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563218303091
[25] H. Zeineddine, U. Braendle, and A. Farah, ''Enhancing prediction of Student success: Automated machine learning approach,'' *Comput. Electr. Eng.*, vol. 89, Jan. 2021, Art. no. 106903. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790620307552
[26] A. O. Oyedeji, A. M. Salami, O. Folorunsho, and O. R. Abolade, ''Analysis and prediction of Student academic performance using machine learning,'' *JITCE J. Inf. Technol. Comput. Eng.*, vol. 4, no. 1, pp. 10–15, Mar. 2020.
[27] N. Walia, M. Kumar, N. M. Nayar, and G. Mehta, ''Student's academic performance prediction in academic using data mining techniques,'' in *Proc. 1st Int. Conf. Intell. Commun. Comput. Res. (ICICCR)*, Rajpura, India, 2020, pp. 1–5.
[28] K. Thaker, Y. Huang, P. Brusilovsky, and H. Daqing, ''Dynamic knowledge modeling with heterogeneous activities for adaptive textbooks,'' in *Proc. 11th Int. Conf. Educ. Data Mining*, Jul. 2018, pp. 592–595.
[29] A. Ahadi, R. Lister, H. Haapala, and A. Vihavainen, ''Exploring machine learning methods to automatically identify students in need of assistance,'' in *Proc. 11th Annu. Int. Conf. Int. Comput. Educ. Res.*, Aug. 2015, pp. 121–130.
[30] N. Tomasevic, N. Gvozdenovic, and S. Vranes, ''An overview and comparison of supervised data mining techniques for student exam performance prediction,'' *Comput. Educ.*, vol. 143, Jan. 2020, Art. no. 103676. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360131519302295
[31] A. Siddique, A. Jan, F. Majeed, A. I. Qahmash, N. N. Quadri, and M. O. A. Wahab, ''Predicting academic performance using an efficient model based on fusion of classifiers,'' *Appl. Sci.*, vol. 11, no. 24, Dec. 2021, Art. no. 11845. [Online]. Available: https://www.mdpi.com/2076-3417/11/24/11845
[32] Y. Lee, ''Using self-organizing map and clustering to investigate problem-solving patterns in the massive open online course: An exploratory study,'' *J. Educ. Comput. Res.*, vol. 57, no. 2, pp. 471–490, Apr. 2019.
[33] J. Dhilipan, N. Vijayalakshmi, S. Suriya, and A. Christopher, ''Prediction of students performance using machine learning,'' *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1055, no. 1, Feb. 2021, Art. no. 012122, doi: 10.1088/1757-899x/1055/1/012122.
[34] S. Minda, ''Online-learning and students' motivation: A research study on the effect of online learning on students' motivation in IAIN padangsidimpuan,'' *Asian Social Sci. Humanities Res. J. (ASHREJ)*, vol. 2, no. 2, pp. 9–16, 2020.
[35] A. K. Pal and S. Pal, ''Analysis and mining of educational data for predicting the performance of students,'' *Int. J. Electron. Commun. Comput. Eng.*, vol. 4, no. 5, pp. 2278–4209, 2013.
[36] O. K. Oyedotun, S. Tackie, E. Olaniyi, and A. Khashman, ''Data mining of Students' performance: Turkish students as a case study,'' *Int. J. Intell. Syst. Appl.*, vol. 7, no. 9, pp. 20–27, Sep. 2015.
[37] T.-O. Tran, H.-T. Dang, V.-T. Dinh, T.-M.-N. Truong, T.-P.-T. Vuong, and X.-H. Phan, ''Performance prediction for students: A multi-strategy approach,'' *Cybern. Inf. Technol.*, vol. 17, no. 2, pp. 164–182, Jun. 2017.
[38] C. Beaulac and J. S. Rosenthal, ''Predicting university Students' academic success and major using random forests,'' 2018, *arXiv:1802.03418*.
[39] L.-M. Crivei, M. Andrei, and G. Czibula, ''A study on applying relational association rule mining based classification for predicting the academic performance of students,'' in *Proc. KSEM 12th Int. Conf. Knowl. Sci., Eng. Manage.*, vol. 11775, 2019, pp. 287–300.
[40] F. Ünal. (2020). *Data Mining for Student Performance Prediction in Education*. [Online]. Available: https://www.intechopen.com/chapters/71573
[41] N. Naicker, T. Adeliyi, and J. Wing, ''Linear support vector machines for prediction of student performance in school-based education,'' *Math. Problems Eng.*, vol. 2020, pp. 1–7, Oct. 2020. [Online]. Available: https://EconPapers.repec.org/RePEc:hin:jnlmpe:4761468
[42] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, ''Student performance prediction model based on supervised machine learning algorithms,'' *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 928, no. 3, Nov. 2020, Art. no. 032019, doi: 10.1088/1757-899x/928/3/032019.
[43] Z. Iqbal, J. Qadir, A. N. Mian, and F. Kamiran, ''Machine learning based student grade prediction: A case study,'' 2017, *arXiv:1708.08744*.
[44] E. Ayers, R. Nugent, and N. Dean, ''A comparison of student skill knowledge estimates,'' in *Proc. 2nd Int. Conf. Educ. Data Mining Educ. Data Mining (EDM)*, Cordoba, Spain, Jul. 2009, pp. 1–10.
[45] S. Parack, Z. Zahid, and F. Merchant, ''Application of data mining in educational databases for predicting academic trends and patterns,'' in *Proc. IEEE Int. Conf. Technol. Enhanced Educ. (ICTEE)*, Jan. 2012, pp. 1–4.
[46] A. Dutt, ''Clustering algorithms applied in educational data mining,'' *Int. J. Inf. Electron. Eng.*, vol. 5, no. 2, pp. 112–116, 2015.
[47] W. Kurdthongmee, ''Utilization of a self organizing map as a tool to study and predict the success of engineering students at Walailak university,'' *Walailak J. Sci. Technol.*, vol. 5, no. 1, pp. 111–123, 2008.
[48] K. Saxena, S. Jaloree, R. Thakur, and S. Kamley, ''Self organizing map (SOM) based modelling technique for student academic performance prediction,'' *Int. J. Future Revolution Comput. Sci. Commun. Eng.*, vol. 3, no. 9, pp. 115–120, 2017.
[49] G. Ciubotariu and L. M. Crivei, ''Analysing the academic performance of students using unsupervised data mining,'' *Studia Universitatis Babes-Bolyai Ser. Inf.*, vol. 64, no. 2, pp. 1–14, 2019.

[50] L. M. Crivei, G. Czibula, G. Ciubotariu, and M. Dindelegan, ''Unsupervised learning based mining of academic data sets for students' performance analysis,'' in *Proc. IEEE 14th Int. Symp. Appl. Comput. Intell. Informat. (SACI)*, May 2020, pp. 1–6.
[51] Z. Onet-Marian, G. Czibula, and M. Maier, ''Using self-organizing maps for comparing students' academic performance in online and traditional learning environment,'' *Stud. Informat. Control*, vol. 30, no. 4, pp. 1–11, 2021.
[52] M.-I. Maier, G. Czibula, and Z.-E. Onet-Marian, ''Towards using unsupervised learning for comparing traditional and synchronous online learning in assessing students' academic performance,'' *Mathematics*, vol. 9, no. 22, p. 2870, Nov. 2021.
[53] L.-Q. Chen, M.-T. Wu, L.-F. Pan, and R.-B. Zheng, ''Grade prediction in blended learning using multisource data,'' *Scientific Program.*, vol. 2021, pp. 1–15, Sep. 2021, doi: 10.1155/2021/4513610.
[54] O. M. Gushchina and A. V. Ochepovsky, ''Data mining of students' behavior in E-learning system,'' *J. Phys., Conf. Ser.*, vol. 1553, no. 1, May 2020, Art. no. 012027, doi: 10.1088/1742-6596/1553/1/012027.
[55] K. Liang, Y. Zhang, Y. He, Y. Zhou, W. Tan, and X. Li, ''Online behavior analysis-based student profile for intelligent E-learning,'' *J. Electr. Comput. Eng.*, vol. 2017, pp. 1–7, Mar. 2017. [Online]. Available: https://www.hindawi.com/journals/jece/2017/9720396/
[56] T. C. Dublin. (2017). *European Erasmus Conversion Table for Trinity Students Returning From Study Abroad*. Accessed: Feb. 7, 2022. [Online]. Available: https://www.tcd.ie/study/assets/PDF/Grade%20Conversion%20Tables_November%202017.pdf
[57] Y. Dodge, *The Concise Encyclopedia of Statistics*. Cham, Switzerland: Springer, 2008.
[58] M. J. Campbell and T. D. V. Swinscow, *Statistics at Square One*, 10th ed. London, U.K.: BMJ Books, 2002.
[59] K. Kira and L. A. Rendell, ''The feature selection problem: Traditional methods and a new algorithm,'' in *Proc. 10th Nat. Conf. Artif. Intell.*, W. R. Swartout, Ed. San Jose, CA, USA: AAAI Press, Jul. 1992, pp. 129–134.
[60] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, ''Overcoming the myopia of inductive learning algorithms with RELIEFF,'' *Appl. Intell.*, vol. 7, no. 1, pp. 39–55, Jan. 1997, doi: 10.1023/A:1008280620621.
[61] R. J. Urbanowicz, R. S. Olson, P. Schmitt, M. Meeker, and J. H. Moore, ''Benchmarking relief-based feature selection methods for bioinformatics data mining,'' 2017, *arXiv:1711.08477*.
[62] L. van der Maaten and G. Hinton, ''Visualizing data using t-SNE,'' *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[63] L. McInnes, J. Healy, and J. Melville, ''UMAP: Uniform manifold approximation and projection for dimension reduction,'' 2018, *arXiv:1802.03426*.
[64] Y. Qi, ''A very brief introduction to nonnegative tensors from the geometric viewpoint,'' *Mathematics*, vol. 6, no. 11, p. 230, Oct. 2018.
[65] A. Rosenberg and J. Hirschberg, ''V-measure: A conditional entropy-based external cluster evaluation measure,'' in *Proc. EMNLP-CoNLL Joint Conf. Empirical Methods Natural Language Process. Comput. Natural Language Learn.*, J. Eisner, Ed. Prague, Czech Republic: ACL, Jan. 2007, pp. 410–420.
[66] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[67] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, ''Scikit-learn: Machine learning in Python,'' *J. Mach. Learn. Res.*, vol. 12 no. 10, pp. 2825–2830, 2012.
[68] R. J. Hyndman and A. B. Koehler, ''Another look at measures of forecast accuracy,'' *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
[69] G. Czibula, A. Mihai, and E. Mihuleţ, ''NowDeepN: An ensemble of deep learning models for weather nowcasting based on radar products' values prediction,'' *Appl. Sci.*, vol. 11, no. 1, p. 125, Dec. 2020.
[70] D. Picca, B. Curdy, and F. Bavaud, ''Non-linear correspondence analysis in text retrieval: A kernel view,'' in *Proc. JADT 8es Journees Internationales d'Analyse Statistique des Donnees Textuelles*, 2006, pp. 741–747.
[71] G. Ciubotariu. (2022). *Intelli-DaM*. Accessed: Feb. 8, 2022. [Online]. Available: https://github.com/george200150/Intelli-DaM
[72] A. P. King and R. J. Eckersley, Eds., ''Inferential statistics III: Nonparametric hypothesis testing,'' in *Statistics for Biomedical Engineers and Scientists*. New York, NY, USA: Academic, 2019, pp. 119–145.
[73] Google. *Online Web Statistical Calculators*. Accessed: Feb. 1, 2022. [Online]. Available: https://astatsa.com/WilcoxonTest/

GABRIELA CZIBULA currently works as a Professor at the Department of Computer Science, Faculty of Mathematics and Computer Science, Babeş-Bolyai University, Cluj-Napoca, Romania. She has published more than 200 papers in prestigious journals and conference proceedings. Her research interests include machine learning, distributed artificial intelligence and multiagent systems, and bioinformatics.



GEORGE CIUBOTARIU is currently pursuing the M.Sc. degree in applied computational intelligence specialization with the Faculty of Mathematics and Computer Science, Babeş-Bolyai University, Cluj-Napoca, Romania. His research interests include computer vision and machine learning.

MARIANA-IOANA MAIER is currently pursuing the Ph.D. degree with the Department of Computer Science, Faculty of Mathematics and Computer Science, Babeş-Bolyai University, Cluj-Napoca, Romania. Her research interests include computational intelligence, machine learning, and educational data mining.

HANNELORE LISEI received the Ph.D. degree in mathematics from Martin-Luther-University Halle-Wittenberg, Germany, and the Habilitation degree in mathematics from Babeş-Bolyai University, Cluj-Napoca, Romania. She is currently an Associate Professor at the Faculty of Mathematics and Computer Science, Babeş-Bolyai University. She is working in the field of probability theory, stochastic analysis, and statistics.