EANN 2017

CrossMark

Improving the evaluation process of students' performance utilizing a decision support software

I. E. Livieris[1] · T. Kotsilieris[2] · V. Tampakas[1] · P. Pintelas[3]

Abstract
During the last decades, educational data mining constitutes a significant tool, offering a first step and a helping hand in the prediction of students' progress and performance. In this work, we present a user-friendly decision support software, for accurately predicting the students' performance at the final examinations of the academic year. The proposed software incorporates a classification scheme which has two major features. Firstly, it identifies with high accuracy the students at-risk of failing the final examinations; secondly, it classifies the students based on their predicted grades. Our numerical experiments show that it achieves better performance than any examined single learning algorithm. The proposed software was developed to provide assistance to students' evaluation and mostly to the early identification of students' at-risk in order to take proper actions for improving their performance.

Keywords Educational data mining · Machine learning · Student evaluation system · Decision support software

1 Introduction

Nowadays, where knowledge and quality constitute a critical factor in global economy, education plays a significant role as knowledge center and human resource developer [6]. The main objective of the educational institutes and one of their biggest challenges is to provide quality education to their students. One way to accomplish the higher level of quality is by predicting students' academic performance and consequently taking early actions by providing appropriate support to students with learning difficulties in order to improve their performance. During the last decades, one of the most important innovations in the educational systems was the extraction of relevant knowledge hidden in the educational dataset utilizing data mining techniques.

Educational data mining (EDM) is an essential process where intelligent methods are applied to extract data patterns from students' databases in order to discover key characteristics and hidden knowledge. This new research field has grown exponentially and gained popularity in the modern educational era because of its potential to improve the quality of the educational institutions and system. The application of EDM is mainly concentrated on improving the learning process by the development of accurate models that predict students' characteristics and performance. The importance of EDM is founded on the fact that it allows educators and researchers to extract useful conclusions from sophisticated and complicated questions such as "find the students who are at-risk in failing the examinations" or "find the students who will exhibit excellent performance" in which traditional database queries cannot be applied [23].

✉ I. E. Livieris
livieris@teiwest.gr

T. Kotsilieris
tkots@teikal.gr

V. Tampakas
vtampakas@teimes.gr

P. Pintelas
pintelas@upatras.gr

1 Department of Computer and Informatics Engineering (DISK Lab), Technological Educational Institute of Western Greece, 263-34 Antirrion, Greece

2 Department of Business Administration (LAIQDA Lab), Technological Educational Institute of Peloponnese, 241-00 Kalamata, Greece

3 Department of Mathematics, University of Patras, 265-00 Patra, Greece

In Greece, like in most countries, secondary education is a two-tied system which comprises two main stages: Gymnasium and Lyceum. Gymnasium covers the first 3 years with the purpose to enrich students' knowledge in all fields of learning while Lyceum covers the next 3 years which further cultivates the students' personalities while at the same time prepares them for admission in higher education. Essentially, Lyceum acts like a bridge between school education and higher learning specializations that are offered by universities [23]. Thus, the ability to monitor students' academic performance and progression is considered essential since the early identification of possible low performers could lead the academic staff to develop personalized learning strategies (extra learning material, exercises, seminars, training tests) aiming to improve students' performance.

During the last decade, the application of data mining on educational data for the development of accurate and efficient decision support systems (DSS) for monitoring students' performance is becoming very popular [7, 10, 14, 23, 27, 28]. More analytically, an academic DSS is a knowledge-based information system that captures, handles and analyzes information which affects or is intended to affect decision making performed by people in the scope of a professional task appointed by a user [5]. Through the use of a predictive DSS, it is possible to forecast students' success in a course and identify those at-risk. Therefore, the development of an academic DSS is significant to students, educators and educational organizations and it will be more valuable if knowledge mined from the students' performance is available for educational managers in their decision-making process.

More comprehensively, students will be properly informed about their academic standing and teachers can identify slow learners, improve their teaching methods and design better strategies for early intervention. Furthermore, analyzing students' learning and making predictions regarding further aspects of their performance is essential for an educational system in order to provide personalized learning activities tailored to each student's special needs or even guiding them to follow technical education.

Nevertheless, the development of such prediction model is a very attractive and challenging task (see [2, 34–36] and references therein). Generally, educational datasets have skewed class distribution in which most cases are usually located in one class [16, 22, 23]. Therefore, a classifier induced from an imbalanced dataset has typically a low error rate at the majority class and an unacceptable error rate for the minority classes. Moreover, the difficulty to distinguish between noise and rare cases is also responsible for poor performance on the minority class.

The objective of this research is to contribute on the prediction of students' performance with major emphasis on the detection of students who may fail to meet the course requirements. Therefore, we are dealing with the following two main tasks:

(a) prediction of students' success or failure (pass/fail), and
(b) prediction of the passed students' final grades (good/ very good/ excellent).

In this work, we present DSS-PSP (Decision Support Software for Predicting Students' Performance) which consists of an integrated software application and provides decision support for evaluating students' performance in the final examinations. The proposed software identifies the students at-risk of failing the final examinations and classifies the students based on their predicted passing grades.

To this end, the DSS-PSP incorporates a two-level classifier [21] which achieves better performance than any single learning algorithm. Moreover, significant advantages of the presented tool are the employment of a simple and user-friendly interface, its scalability due to its modular nature of design and implementation and its operating system neutrality.

Our primary goal is to support the academic task of successfully predicting the students' performance in the final examinations of the school year. Furthermore, decision-makers are able to evaluate various educational strategies and generate forecasts by utilizing several input data.

The remainder of this paper is organized as follows: The next section presents a survey on machine learning algorithms that have been successfully used for predicting students' performance. Section 3 presents a description of the educational dataset utilized in our study and our proposed 2-level machine learning classifier. Finally, Sect. 4 presents the main features of our decision support software and Sect. 5 presents our conclusions.

2 Related studies

During the last decade, the application of data mining techniques for the development of accurate and efficient (DSS) provided useful outcomes and results that assist in addressing many issues and problems in the educational domain. Romero and Ventura [34, 35] and Baker and Yacef [2] have provided some extensive reviews of different types of educational systems and how data mining can be successfully applied to each of them. More specifically, they described in detail the process of mining learning data, as well as how to apply the data mining techniques, such as statistics, visualization, classification, clustering and association rule mining. Along this line, recently Dutt et al. [12] presented a review of how EDM

seeks to discover new insights into learning with new tools and techniques, so that those insights impact the activity of practitioners in all levels of education.

Deniz and Ersan [10] demonstrated the usefulness of an academic decision support system in evaluating huge amounts of student-course related data. Moreover, they presented the basic concepts used in the analysis and design of a new DSS software package, called "Academic Decision Support System" and presented various ways in which student performance data can be analyzed and presented for academic decision making.

Kotsiantis [16] compared some state-of-the-art regression algorithms to find out which algorithm is more appropriate not only for the accurate prediction of student's performance but also to be used as an educational supporting tool for tutors. Additionally, he presented a prototype decision support system for predicting students' academic progress in a distance learning system using key demographic characteristics, attendance and their marks in written assignments.

In their study, Chau and Phung [7] highlighted the importance of educational decision-making support to students, educators and educational institutes and pointed out that this support will be more valuable if lots of relevant data and knowledge mined from data are available for educational managers in their decision-making process. Additionally, they proposed a knowledge-driven DSS for education with a semester credit system by taking advantage of educational data mining. Their proposed educational DSS is helpful for educational managers to make more appropriate and reasonable decisions about students' study and further give support to students for their graduation.

Romero et al. [33] studied how web usage mining can be applied in e-learning systems in order to predict the marks that university students will obtain in the final examination of a course. Instead of traditional classification algorithms, they propose a classification via clustering to improve the prediction of first-year students' performance. In addition, they developed a specific mining tool which takes into account the student's active involvement and daily usage in a Moodle forum.

Nagy et al. [27] proposed a "Student Advisory Framework" that integrates educational data mining and knowledge discovery to build an intelligent system. The system can be used to provide pieces of consultations to a first-year university student to pursue a certain education track where he/she will likely succeed in, aiming to decrease the high rate of academic failure among these students. The framework acquires information from the datasets which stores the academic achievements of students before enrolling to higher education together with their first-year grade after enrolling in a certain department. After acquiring all the relevant information, the intelligent system utilizes both classification and clustering techniques to provide recommendations for a certain department for a new student. Additionally, they presented a case study to prove the efficiency of the proposed framework. Students' data were collected from Cairo Higher Institute for Engineering, Computer Science and Management during the period from 2000 to 2012.

Grivokostopoulou et al. [14] presented a data mining methodology to discover relationships between the students' learning performance data and predict their final performance in the "Artificial Intelligence" course. More specifically, they analyzed the students' performance at six interim examinational tests during the semester using decision trees and extract semantic rules to make predictions. Their primary goal was to trace students that are in edge to fail the examinations and in danger to drop off the course. Moreover, their proposed methodology has been integrated in an educational system used to assist students in learning the course and enhance the quality of the educational content.

Mishra et al. [25] focused on the early identification of secondary school students who are at high risk of failure, thereby helping the educators to take timely actions in order to improve the students' performance and success rate through extra coaching and counseling. Moreover, the authors classified the important attributes that influenced students' third semester performance and established the effects of emotional quotient parameters (i.e., assertion, empathy, decision-making ability, leadership ability, drive and stress management skills) that influenced placement.

Paz et al. [29] developed a DSS based on a clustering algorithm for college completion model. Their proposed system utilized data from students' registration and grades databases while the client front-end ensures adequate presentation so as to reveal significant details and dependencies. The system can be used to not only for supplying information to the user but also to aid the decision-making process aiming to decrease the high rate of academic failure among students.

In more recent works, Livieris et al. [22] introduced a software tool for predicting the students' performance in the course of "Mathematics" of the first year of Lyceum. They conducted an experimental analysis utilizing a variety of classification algorithms which revealed that the neural network classifier achieved the best accuracy and exhibited more consistent behavior. Along this line, in [23] the authors presented a user-friendly decision support software for predicting students' performance, together with a case study concerning the final examinations in Mathematics. Their proposed tool is based on a hybrid prediction system which combines four learning algorithms utilizing a simple voting scheme. Their experimental results revealed that the

application of data mining can offer significant insights in student progress and performance.

Marquez-Vera et al. [24] studied the serious problem of early prediction of high school dropout and proposed a methodology to discover comprehensible prediction models of student dropout as soon as possible. Additionally, they presented a case study using data from 419 first-year high school Mexican students. The authors illustrated that their proposed method is possible of successfully predicting student dropout within the first 4–6 weeks of the course and trustworthy enough to be used in an early warning system.

Recently, Kužnar and Gams [18] presented "Metis", a novel system which predicts students' failure and provides tools in form of smartphone application to apply preventive measures aiming to mitigate a negative outcome. Metis utilizes machine learning techniques on educational data stored in a Slovenian school information system to identify students with an increased risk of failing a course. The identified students are then referred to an education advisor who will construct an action plan to support the student through individual consultations with the student. Furthermore, the action plan can be followed by a smartphone application which serves as an interface for praising progress and achievements of the objectives of the action plan. Finally, the authors presented that one of the most important advantages of Metis is the ability to perform according to the educator's needs with respect to the required minimum precision detection.

3 Methodology

The aim of this study is to develop a decision support tool for predicting students' performance at the final examinations. For this purpose, we have adopted the following methodology which consists of three stages.

The first stage of the proposed methodology concerns data collection and data preparation and in the next stage, we introduce our proposed 2-level classification scheme. In the final stage, we evaluate the classification performance of our proposed 2-level classification algorithm with that of the most popular and frequently utilized algorithms by conducting a series of experiments.

3.1 Dataset

For the purpose of this study, we have utilized a dataset concerning the performance of 2260 students in courses of "Algebra" and "Geometry" of the first 2 years of Lyceum. The data have been collected by the *Microsoft showcase school "Avgoulea-Linardatou"* during the years 2007–2016. Extensive research has been conducted to determine the factors that have a material impact in students' success in the examinations. The majority of these studies focused on predicting students' grade in a course based on a variety of assignment methods [9, 22, 23, 32]. Table 1 reports the set of attributes used in our study which concern information about the students' performance such as oral grades, tests grades, final examination grades and semester grades which assess students' understanding of important mathematical concepts and topics daily.

Each instance is characterized by ten (10) time-variant attributes which refer to the students' performance utilizing a 20-point grading scale, where 0 is the lowest grade and 20 is the perfect score. The assessment of students during each semester consists of oral examination, two 15-min pre-warned tests, a 1-h examination and the overall semester performance of each student. The 15-min tests include multiple choice questions and short answer problems while the 1-h examinations include several theory questions and a variety of difficult mathematical problems requiring solving techniques and critical analysis. Finally, the overall semester grade of each student addresses the personal engagement of the student in the lesson and his progress. The final grade of each semester is dependent on the student's performance on: written assignments/homeworks ($\sim 20\%$), student's behavior and active participation in the classroom ($\sim 30\%$), written tests and the 1-h examination ($\sim 50\%$). The final examination is run at the end of the course and encompasses the material taught in both semesters. This assessment system is followed by all Lyceums across the country and is prescribed by the Ministry of Education.

The students were classified utilizing a four-level classification scheme according to students' performance evaluation in the Greek schools:

- "Fail" stands for student's performance between 0 and 9 (212 instances).
- "Good" stands for student's performance between 10 and 14 (690 instances).
- "Very good" stands for student's performance between 15 and 17 (659 instances).
- "Excellent" stands for student's performance between 18 and 20 (699 instances).

Moreover, similar to [22, 23], since it is of great importance for an educator to recognize weak students in the middle of the academic period, two datasets have been created based on the attributes presented in Table 1 and on the class distribution.

- DATA$_1$: It contains the attributes concerning the students' performance during the 1st semester.
- DATA$_2$: It contains the attributes concerning the students' performance during the 1st and 2nd semesters.

Table 1 List of features used in our study

Attributes	Type	Values	Mean	SD
Oral grade of the 1st semester	Integer	[0,20]	15.24	3.69
Grade of the 1st test of the 1st semester	Real	[0,20]	14.19	4.59
Grade of the 2nd test of the 1st semester	Real	[0,20]	14.73	4.76
Grade of the 1st semester's final examination	Real	[0,20]	14.42	4.48
Final grade of the 1st semester	Integer	[0,20]	15.58	3.13
Oral grade of the 2nd semester	Integer	[0,20]	14.83	4.37
Grade of the 1st test of the 2nd semester	Real	[0,20]	14.44	4.64
Grade of the 2nd test of the 2nd semester	Real	[0,20]	14.23	4.73
Grade of the 2nd semester's final examination	Real	[0,20]	14.40	4.42
Final grade of the 2nd semester	Integer	[0,20]	15.87	2.98
Grade in the final examinations	Ordinal	"Fail", "Good", "Very good', "Excellent"		

3.2 2-Level classifier

Our primary goals in the present research are the accurate and early identification of the students who are at-risk and the accurate classification of the students who have successfully passed the course.

For this purpose, we consider a two-level classification scheme, aiming on achieving higher classification accuracy than any individual classifiers. Two-level classification schemes are heuristic pattern recognition tools that are supposed to yield better classification accuracy than single-level ones at the expense of a certain complication of the classification structure [3, 17, 41].

On the first level of our proposed classification scheme, we utilize a classifier to distinguish the students who are likely to "Pass" or "Fail" in the final examinations. More specifically, this classifier predicts if the student's performance is between 0 and 9 (Fail) or between 10 and 20 (Pass). In the rest of our work, we refer to this classifier as A-level classifier. Clearly, the primary goal of this classifier is to identify the students' who are at-risk. In case the verdict (or prediction) of the A-level classifier is "Pass" in the final examinations, we utilize a second-level classifier in order to conduct a more specialized decision and distinguish between "Good", "Very good" and "Excellent". This classifier is titled as the B-level classifier. An overview of our proposed 2-level classifier is depicted in Fig. 1.

Furthermore, it is worth mentioning that the corresponding training sets TS_A and TS_B of the A-level and B-level classifier are generated by the original training set as follows:

Let (x, y) be an instance contained in the training set, where x stands for the vector of attributes (as listed in Table 1) while y stands for the output variable. In case, "y = Fail" then it is immediately imported to the training set TS_A. In contrast, in case "$y \neq$ Fail" then the instances $(x, $ "Pass") and (x, y) are imported in the training sets TS_A and TS_B, respectively.

Essentially, TS_A contains all the instances in which the respective output variable y is "Fail" and the rest of instances the training set with the output variable y changed into "Pass", while TS_B contains all the instances of the training set in which the respective output variable y is not "Fail".

It is worth mentioning that the rationale behind the development of the presented 2-level classifier lies in the diversified pedagogical actions to be taken in order to support a student's performance so as to succeed (i.e., classified as "Pass" or "Fail") and those needed so as to obtain a better final mark (i.e., classification among "Good", "Very good" and "Excellent"). On this basis, any other classification scheme (e.g., "Excellent" vs ("Very good", "Good", "Very good")) may be technically preferable in terms of accuracy but pedagogically invalid.

3.3 Experimental results

In this section, we report a series of tests in order to evaluate the performance of our proposed 2-level classification scheme with that of the most popular and commonly used classification algorithms.

The multi-layer perceptron (MLP) [37] and the RBF algorithm [26] were representatives of the artificial neural networks which has been established as well-known learning algorithm for building and training a neural network [20]. From the support vector machines, we have selected the sequential minimal optimization (SMO) algorithm since it is one of the fastest training methods [30] while Naive Bayes (NB) algorithm was the representative of the Bayesian networks [11]. From the decision trees, C4.5 algorithm [31] and logistic model tree (LMT) [19]
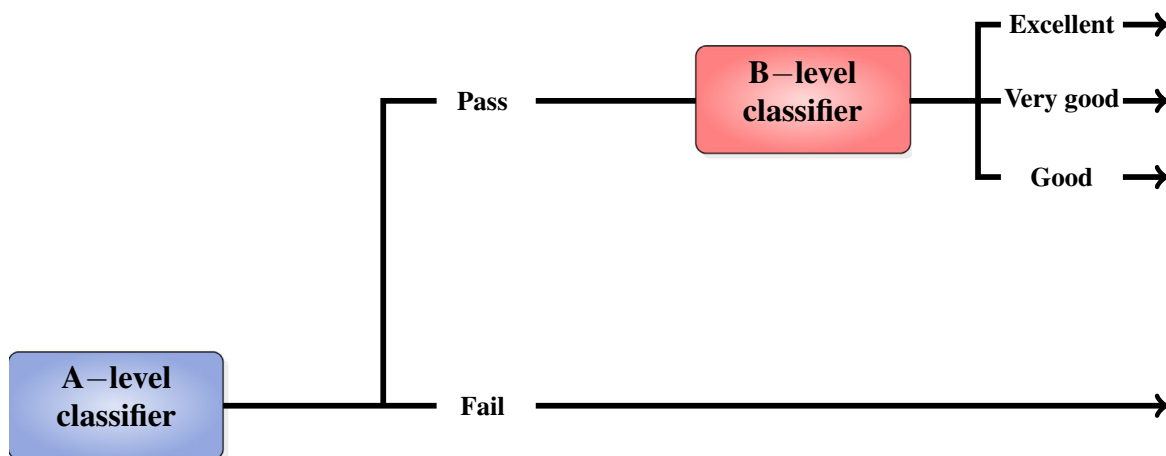
Fig. 1 An overview of the 2-level classifier

were chosen for our study and RIPPER (JRip) [8] and PART [13] algorithms were selected as typical rule-learning techniques since they probably consist the most usually used methods for producing classification rules. Finally, 3-NN and 10-NN algorithms were selected as instance-based learners [1] with Euclidean distance as distance metric. Studies have shown that the above classifiers constitute some of the most effective and widely used data mining algorithms [40] for classification problems. Moreover, in our numerical experiments voting stands for the voting scheme using JRip, 3-NN, MLP and SMO as base classifiers presented in [23].

The classification algorithms have been implemented in WEKA 2.9 Machine Learning Toolkit [15] and the classification accuracy was evaluated using the stratified ten-fold cross-validation, i.e., the data were separated into folds so that each fold had the same distribution of grades as the entire data set. Moreover, in order to minimize the effect of any expert bias, instead of attempting to tune any of the algorithms to the specific datasets, all algorithms were used with their default parameter settings included in the WEKA software.

We evaluate the performance of our proposed 2-level classification scheme in terms of accuracy which consists as one of the most frequently used measures for assessing the overall effectiveness of a classification algorithm and is defined as follows:

$$\text{Accuracy} = \frac{\text{Number of correctly classified students}}{\text{Total number of students}}$$

Table 2 summarizes the accuracy of each individual classifier and the accuracy of the 2-level classification scheme, relative to both datasets. Clearly, our proposed 2-level scheme considerably improved the performance of each individual classifier from 3.8 to 9.5%.

Moreover, it is worth to mention that most classifiers present lower classification accuracy in case where the students' grades in the 2nd semester are also taken into consideration. In contrast, LMT and NB improved their accuracy while SMO and C4.5 present almost identical performance.

In the sequel, motivated by the efficiency of our proposed 2-level classification scheme, we consider to perform a performance evaluation utilizing different classification algorithms at each level and explore its classification accuracy. Our aim is to find which of these classifiers is best suited for A-level and B-level for producing the highest performance. Therefore, we consider three performance metrics:

$$\begin{align} \text{Accuracy} \atop \text{(Fail)} &= \frac{\text{Number of students correctly predicted to fail}}{\text{Total number of failed students}} \\ \text{Accuracy} \atop \text{(Pass)} &= \frac{\text{Number of students correctly predicted to pass}}{\text{Total number of passed students}} \\ \text{Accuracy} \atop \text{(Grade)} &= \frac{\text{Number of passed students correctly predicted their grade}}{\text{Total number of passed students}} \end{align}$$

The first two metrics evaluate the performance of A-level classifier while the last metric evaluates the performance of B-level classifier. Furthermore, since the number of students who failed in the examinations is about 10%, it is crucial for a prediction model to correctly identify them. As a goal of this study is to identify the students at-risk, it is significant to achieve the highest possible predictive accuracy for the student who failed in the examinations. Therefore, we present an additional performance metric:

$$F_1 = \frac{2\text{nFiF}}{2\text{nFiF} + \text{nFiP} + \text{nPiF}},$$

where nFiF stands for the number of students who failed and correctly identified, nFiP stands for the number of students who failed and identified as passed and nPiF stands for the number of students who passed and identified as failed.

It is worth mentioning that the performance metric F_1 constitutes a harmonic mean of precision and recall. In

Table 2 Individual classifier and 2-level classifier Accuracy

Classifier	DATA₁		DATA₂	
	Individual (%)	2-Level (%)	Individual (%)	2-Level (%)
MLP	81.13	89.12	76.22	89.08
RBF	83.76	89.69	83.49	89.56
SMO	84.09	89.94	82.71	89.93
NB	75.77	81.24	71.72	81.46
C4.5	84.22	89.83	84.13	89.78
LMT	85.20	89.56	86.46	90.09
JRip	84.68	90.04	86.09	89.64
PART	84.38	89.42	84.86	88.32
3-NN	85.21	88.14	78.67	85.33
5-NN	84.39	88.05	80.44	84.98
10-NN	84.86	86.90	80.44	85.97
Voting	85.81	88.22	84.42	87.03

particular, this metric takes into account accuracy for the students who passed and failed the examinations and weights more the accuracy for students who failed, than the students who passed [39]. From an educator's perspective, it is better to misidentify a "good" student than a "failed" student. Misidentifying a "good" student as a potential fail may encourage him/her to work harder and improve his/her

performance. In contrast, misidentifying a "failed" student as a potential "pass" may prevent him/her from taking the proper actions to pass the course and consequently fail in the final examinations.

Tables 3 and 4 present the performance evaluation of A-level and B-level classifiers utilizing various classification algorithms, for both datasets, respectively. The

Table 3 Performance evaluation of A-level and B-level classifiers on DATA₁

Classifier	Accuracy (Fail) (%)	Accuracy (Pass) (%)	F_1 (%)	Accuracy (Grade) (%)
MLP	82.55	99.51	88.16	90.92
RBF	83.49	99.27	87.62	90.77
SMO	84.45	99.37	88.61	**91.14**
NB	81.60	99.41	87.15	86.82
C4.5	**84.48**	**99.56**	**89.50**	91.02
LMT	83.96	99.46	88.58	90.92
JRip	83.96	99.37	88.34	91.02
PART	83.49	99.46	88.50	90.43
3-NN	81.60	99.31	87.15	89.36
5-NN	82.55	99.37	87.50	89.21
10-NN	80.66	99.51	87.02	89.11

Table 4 Performance evaluation of A-level and B-level classifiers on DATA₂

Classifier	Accuracy (Fail) (%)	Accuracy (Pass) (%)	F_1 (%)	Accuracy (Grade) (%)
MLP	82.55	99.41	87.72	90.72
RBF	83.96	99.32	88.12	90.63
SMO	84.43	99.37	88.61	**91.11**
NB	82.55	99.32	87.28	87.42
C4.5	**84.91**	**99.61**	**90.00**	90.92
LMT	84.43	99.37	88.61	**91.11**
JRip	84.43	99.17	87.75	90.87
PART	82.08	99.41	87.44	89.40
3-NN	82.55	99.12	86.42	89.84
5-NN	80.66	99.17	85.50	87.35
10-NN	80.19	99.27	85.64	88.18

accuracy measure of the best performing algorithm is highlighted in bold for each dataset.

Clearly, C4.5 illustrates the best performance as A-level classifier, since it exhibits the highest accuracy of correctly classified passed and failed students, relative to both datasets. More specifically, C4.5 predicts that a student will fail in the final examinations with probability 84.48% and 84.91%, relative to $DATA_1$ and $DATA_2$, respectively. Moreover, C4.5 predicts that a student will successfully pass the final examinations with probability 99.56% and 99.61% in the same situations. As regards F_1 metric, C4.5 reports the best performance, exhibiting 85.9% and 90% for datasets $DATA_1$ and $DATA_2$, respectively.

Finally, SMO reports the best performance as B-level classifier, illustrating the highest percentage of correctly classified students who have successfully passed the course, followed by LMT.

Tables 5 and 6 summarize the performance of the proposed 2-level classifier utilizing various A-level and B-level classifiers for both datasets, respectively. The best performing technique for each dataset is illustrated in boldface. Furthermore, Figs. 2 and 3 present the average performance of A-level and B-level classifiers, respectively.

Firstly, we observe that the C4.5 reports the best classification performance as A-level classifier, followed by LMT and JRip, relative to both datasets. In particular, it

Table 5 2-level classifier classification accuracy (%) on $DATA_1$

	B-level classifier										
	MLP (%)	RBF (%)	SMO (%)	NB (%)	C4.5 (%)	LMT (%)	JRip (%)	PART (%)	3-NN (%)	5-NN (%)	10-NN (%)
A-level classifier											
MLP	89.12	89.60	89.73	85.93	89.69	89.60	89.73	89.47	88.19	88.23	88.23
RBF	89.96	89.69	89.78	84.93	89.91	89.73	89.73	89.38	88.14	83.76	88.19
SMO	89.42	89.65	89.94	86.24	90.00	89.91	90.04	89.47	88.50	88.32	88.54
NB	84.38	85.13	84.96	81.24	84.91	84.82	84.96	84.38	83.41	83.81	83.27
C4.5	89.42	89.65	**90.13**	86.24	89.83	90.04	90.04	89.56	88.50	88.27	88.54
LMT	88.23	89.60	90.00	85.14	90.00	89.56	89.96	89.42	88.31	88.27	88.32
JRip	89.42	89.69	90.04	86.24	90.00	89.73	90.04	89.47	88.50	88.36	88.54
PART	88.45	89.51	90.00	86.15	89.96	89.42	90.00	89.42	88.45	88.19	88.23
3-NN	89.07	89.38	89.69	85.84	89.69	89.56	89.25	89.12	88.14	88.05	86.62
5-NN	89.34	89.32	88.23	85.66	89.69	89.51	89.51	89.16	87.92	88.05	87.96
10-NN	89.07	89.34	89.69	85.84	89.69	89.56	89.65	89.20	86.41	88.01	86.90

The highest accuracy is highlighted in bold

Table 6 2-level classifier classification accuracy (%) on $DATA_2$

	B-level classifier										
	MLP (%)	RBF (%)	SMO (%)	NB (%)	C4.5 (%)	LMT (%)	JRip (%)	PART (%)	3-NN (%)	5-NN (%)	10-NN (%)
A-level classifier											
MLP	89.08	89.03	89.91	81.77	90.00	90.09	89.73	88.36	86.68	85.80	86.15
RBF	88.72	89.56	89.65	81.68	89.34	89.78	89.47	89.07	85.66	85.44	85.88
SMO	88.01	89.65	89.93	80.49	88.58	88.67	88.41	88.58	85.27	85.93	84.73
NB	89.07	84.89	89.56	81.46	89.65	89.73	89.38	82.83	86.33	85.44	85.80
C4.5	89.60	90.10	**90.43**	81.80	89.80	90.09	89.80	88.98	86.60	85.88	87.11
LMT	88.85	89.65	89.91	81.77	90.00	90.09	89.73	88.58	86.68	85.93	86.64
JRip	89.42	89.47	90.22	81.77	90.00	89.96	89.64	88.41	86.68	85.80	86.15
PART	88.58	89.38	89.65	81.55	89.73	89.82	89.47	88.32	86.42	85.49	85.84
3-NN	87.43	88.23	89.65	79.91	88.01	88.10	87.74	87.12	85.33	84.03	83.85
5-NN	87.39	87.65	87.08	79.69	87.96	87.96	87.70	87.30	83.89	84.98	84.12
10-NN	89.25	87.65	90.00	81.59	89.82	89.91	89.56	86.55	86.50	83.76	85.97

The highest accuracy is highlighted in bold

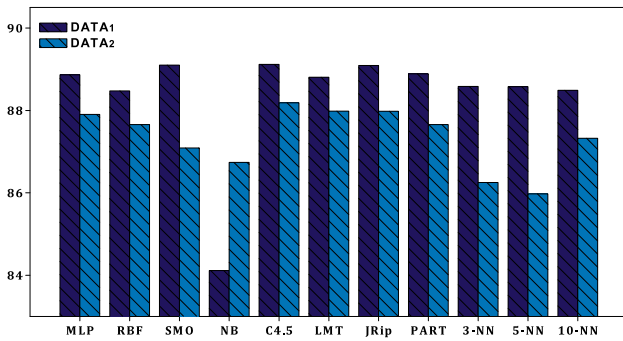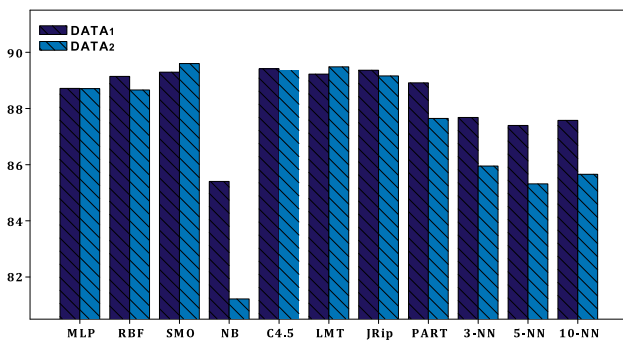**Fig. 2** Box plot for the average performance of A-level classifier for each dataset

Fig. 3 Box plot for the average performance of B-level classifier for each dataset

exhibits 86.24–90.13% and 81.80–90.34% classification performance, for $DATA_1$ and $DATA_2$, respectively. Moreover, the interpretation of Fig. 2 reveals that C4.5 reported the highest average classification accuracy for both datasets. As regards B-level classifier, SMO exhibited the best performance slightly outperforming C4.5 and LMT. More specifically, SMO reported 89.4% and 89.85% average classification performance, for $DATA_1$ and $DATA_2$, respectively while C4.5 exhibited 89.39% and 89.49% and LMT 89.19% and 89.64%, in the same situations.

Conclusively, it is worth noticing that based on the previous discussion, we conclude that the best classification performance of the 2-level classifier was presented in case C4.5 was selected as A-level classifier and SMO as a B-level one.

Finally, in order to illustrate the classification accuracy of our proposed algorithm with a more traditional approach, we compare the confusion matrix of 2-level classification scheme with that of best single classifier (LMT), relative to both datasets. Notice that confusion matrix gives an additional information about classes which are commonly mislabeled one as another. Tables 7 and 8 present the confusion matrices of the 2-level classification scheme and LMT, relative to both datasets.

Table 7 Confusion matrix of 2-level classification scheme

		Fail	Good	Very good	Excellent
$DATA_1$	Fail	179	21	9	3
	Good	7	618	45	20
	Very good	2	15	571	71
	Excellent	0	9	20	670

		Fail	Good	Very good	Excellent
$DATA_2$	Fail	180	22	10	0
	Good	7	620	48	15
	Very good	1	15	576	67
	Excellent	0	6	23	670

Table 8 Confusion matrix of LMT

		Fail	Good	Very good	Excellent
$DATA_1$	Fail	169	38	10	1
	Good	9	609	54	17
	Very good	3	48	524	84
	Excellent	1	5	65	629

		Fail	Good	Very good	Excellent
$DATA_2$	Fail	174	26	10	3
	Good	9	614	46	21
	Very good	3	39	536	81
	Excellent	0	23	41	635

Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class.

4 DSS-PSP: decision support software

For the purpose of this study, we present a user-friendly decision support software, which is called DSS-PSP[1] for predicting the performance of an individual student at the final examinations based on its grades on the 1st and/or 2nd semester. The software is based on the WEKA Machine Learning Toolkit and has been developed in JAVA, making it platform independent and easily executed even by non-experienced users. Notice that DSS-PSP is in fact a significantly updated version of the software presented in [23] with enhanced functionalities.

---

[1] The tool is available at http://www.math.upatras.gr/livieris/ DSSPSP.zip. Notice that Java Virtual Machine (JVM) 1.2 or newer is needed for the execution of the program.

Fig. 4 DSS-PSP: interface

Fig. 5 Selection of the classifier

Figure 4 illustrates a screenshot of our proposed decision support software DSS-PSP illustrating its main features:

- *Student personal data* This module is optionally used to import student's name, surname, father's name and remarks.
- *1st semester's grades* This module is used to import the student's grades of the first semester.
- *2nd semester's grades* This module is used to import the student's grades of the second semester.
- *Messages* This module is used to print the messages, warnings and outputs of the tool.

Subsequently, we demonstrate a use case in order to illustrate the functionalities of DSS-PSP. Firstly, the user/educator can use our data embedded in the software by clicking on the button "Import data" or he can load his/her data collected from his/her own past courses in XLSX (Microsoft Office Excel 2007 XML) file format. Notice that if the first row of the users' data refers to the names of the attributes then it is automatically ignored by the software.

Next, by clicking on the "Select classifier" button, DSS-PSP enables the user to choose between the old classifier based on a voting scheme [23] and the proposed 2-level classification algorithm (Fig. 5) which utilizes C4.5 as a A-level classifier and SMO as B-level classifier. Based on the previous discussion, we recall that the proposed 2-level classification algorithm is more accurate and it can be trained significantly faster than the voting scheme presented in [23].

Subsequently, the user can import the new student's grades of the 1st and/or 2nd semester in the corresponding fields. Then, the DSS-PSP is able to predict the student's performance at the final examinations by simply clicking on the button "Prediction". More specifically, the software presents the prediction and the confidence (probability of the prediction) of the A-level classifier for passing/falling of a given student in the final examinations. Moreover, in
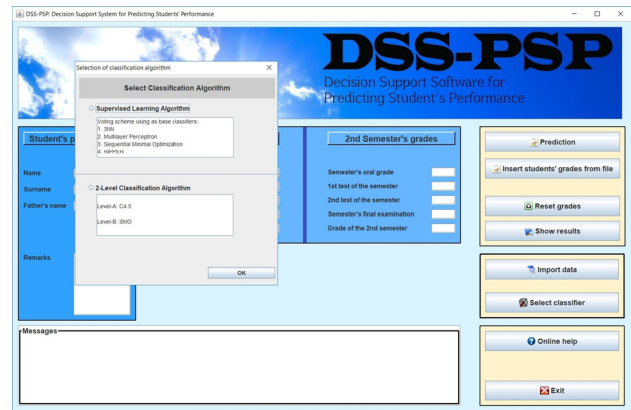
case that the A-level classifier verdict (or prediction) is "Pass", then the software also presents the prediction of the B-level classifier along with its confidence which distinguishes the student's final grade in three categories between "Good", "Very good" and "Excellent". Notice that the way in which the confidence predictions are measured is dependent on the type of utilized base learner (see [4, 38] and the references there in).

In the example presented in Fig. 6, the model predicts that the student is classified as "Excellent" based on the student's grades of both academic semesters. Furthermore, we have updated DSS-PSP adding a new functionality allowing the user to massively import the students' grades from an XLSX for evaluation. Firstly, by clicking on the button "Import students' grades from file" the tool asks if the user imports students' data concerning grades from the 1st and/or 2nd semester. After the grades have been imported, the software presents its predictions about each individual student as presented in Fig. 7.

Additionally, the DSS-PSP stores all its previous predictions and user actions and provides the ability to present them by simply clicking on the button "Show results" as it

Fig. 6 DSS-PSP prediction about the performance of a new student at the final examinations

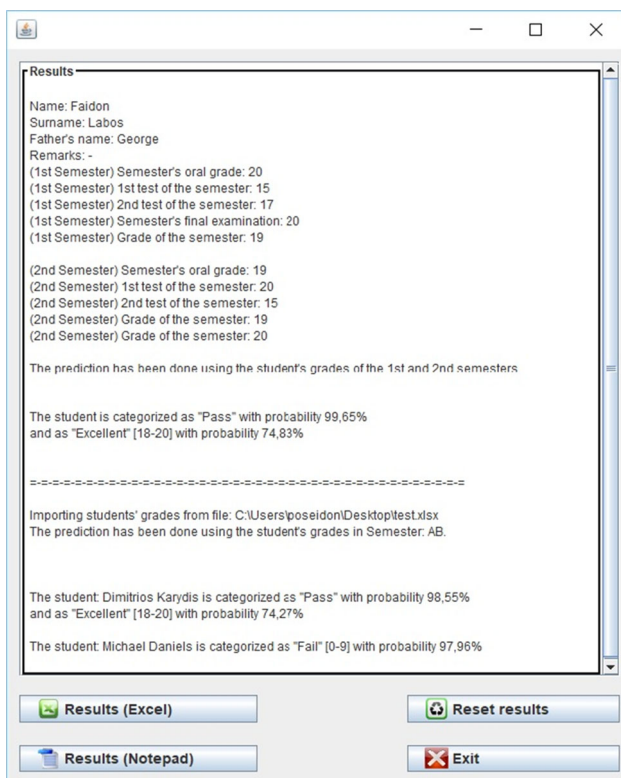**Fig. 7** DSS-PSP predictions about the performance of new students at the final examinations

Fig. 8 Stored predictions about students' performance at the final examinations

is illustrated in Fig. 8. Moreover, the tool provides online help, for novice users and the ability to see all previous predictions by clicking the button "Online help".

5 Conclusions

In this work, we provided an extended analysis of DSS-PSP, a user-friendly decision support system for predicting the students' academic performance at the final examinations which incorporates a 2-level machine learning classifier. Our numerical experiments revealed that the proposed scheme considerably improves the accuracy of each individual classification algorithm, producing better classification results. Moreover, the software is highly adaptable and expandable due to its modular design and implementation. Additional functionalities can be easily added according to the user needs.

The software was developed to provide an assistance to students' evaluation and mostly to the early identification of students' at-risk in order to take proper actions for improving their performance. Our objective and expectation is that this work could be used as a reference for decision making in the admission process and strengthen the service system in educational institutions by offering customized assistance according to students' predicted performance.

Since the experimental results are quite encouraging, a next step could be a systematic and extensive evaluation of the tool by several groups of external teachers in order to evaluate its usability. Furthermore, another direction for a future research would be to evaluate the performance of the presented 2-level classifier utilizing data from several lesson and apply our methodology for predicting the students performance at Panhellenic (national)-level examinations for admission to universities.

Compliance with ethical standards

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Aha D (1997) Lazy learning. Kluwer Academic Publishers, Dordrecht
2. Baker RS, Yacef K (2009) The state of educational data mining in 2009: a review future visions. J Educ Data Min 1(1):3–17
3. Barabash YL (1983) Collective statistical decisions in recognition. Radio i Sviaz, Moscow
4. Baumgartner D, Serpen G (2009) Large experiment and evaluation tool for WEKA classifiers. DMIN 16:340–346
5. Bresfelean VP, Ghisoiu N (2010) Higher education decision making and decision support systems. WSEAS Trans Adv Eng Educ 7:43–52
6. Chalaris M, Gritzalis S, Maragoudakis M, Sgouropoulou C, Tsolakidis A (2014) Improving quality of educational processes providing new knowledge using data mining techniques. Procedia Soc Behav Sci 147:390–397
7. Chau VTN, Phung NH (2012) A knowledge driven education decision support system. In 2012 IEEE RIVF international

conference on computing communication technologies, research, innovation, and vision for the future, pp 1–6

8. Cohen W (1995) Fast effective rule induction. In: International conference on machine learning, pp 115–123

9. Cortez P, Silva A.M.G (2008) Using data mining to predict secondary school student performance. In: Proceedings of 5th annual future business technology conference, pp 5–12

10. Deniz DZ, Ersan I (2002) An academic decision support system based on academic performance evaluation for student and program assessment. Int J Eng Educ 18(2):236–244

11. Domingos P, Pazzani M (1997) On the optimality of the simple Bayesian classifier under zero-one loss. Mach Learn 29:103–130

12. Dutt A, Ismail MA, Herawan T (2017) A systematic review on educational data mining. IEEE Access 29:103–130

13. Frank E, Witten IH (1998) Generating accurate rule sets without global optimization. In: 15th international conference on machine learning, pp 144–151

14. Grivokostopoulou F, Perikos I, Hatzilygeroudis I (2014) Utilizing semantic web technologies and data mining techniques to analyze students learning and predict final performance. In: International conference on teaching, assessment and learning (TALE), pp 488–494

15. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I (2009) The WEKA data mining software: an update. SIGKDD Explor Newsl 11:10–18

16. Kotsiantis S (2012) Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. Artif Intell Rev 37:331–344

17. Kuncheva LI (1993) "Change-glasses" approach in pattern recognition. Pattern Recogn Lett 14:619–623

18. Kužnar D, Gams M (2016) Metis: system for early detection and prevention of student failure. In: 6th international workshop on combinations of intelligent methods and applications (CIMA 2016), p 39

19. Landwehr N, Hall M, Frank E (2005) Logistic model trees. Mach Learn 59(1–2):161–205

20. Lerner B, Guterman H, Aladjem M, Dinstein I (1999) A comparative study of neural network based feature extraction paradigms. Pattern Recogn Lett 20(1):7–14

21. Livieris IE, Drakopoulou K, Kotsilieris Th, Tampakas V, Pintelas P (2017) DSS-PSP—a decision support software for evaluating students' performance. In: Engineering applications of neural networks (EANN), vol 744. Springer, pp 63–74

22. Livieris IE, Drakopoulou K, Pintelas P (September 2012) Predicting students' performance using artificial neural networks. In: Information and communication technologies in education

23. Livieris IE, Mikropoulos T, Pintelas P (2016) A decision support system for predicting students' performance. Themes Sci Technol Educ 9:43–57

24. Márquez-Vera C, Cano A, Romeroand C, Noaman AYM, Mousa-Fardoun H, Ventura S (2016) Early dropout prediction using data mining: a case study with high school students. Expert Syst 33:107–124

25. Mishra T, Kumar D, Gupta S (2014) Mining students' data for prediction performance. In: 4th international conference on advanced computing communication technologies, pp 255–262

26. Mitchell T (1997) Machine learning. McGraw Hill, New York

27. Nagy HM, Aly WM, Hegazy OF (2013) An educational data mining system for advising higher education students. World Acad Sci Eng Technol Int J Inf Eng 7(10):175–179

28. Noaman AY, Luna JM, Ragab AHM, Ventura S (2016) Recommending degree studies according to students' attitudes in high school by means of subgroup discovery. Int J Comput Intell Syst 9(6):1101–1117

29. Paz AM, Gerardo BD, Tanguilig III BT (2014) Academic decision support system for college completion model. In: International conference on advances in computer and electronics technology

30. Platt J (1999) Using sparseness and analytic QP to speed training of support vector machines. In: Kearns MS, Solla SA, Cohn DA (eds) Advances in neural information processing systems. MIT Press, Cambridge, pp 557–563

31. Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann, San Francisco

32. Ramaswami M, Bhaskaran R (2010) A CHAID based performance prediction model in educational data mining. Int J Comput Sci Issues 7(1):135–146

33. Romero C, Espejo PG, Zafra A, Romero JR, Ventura S (2013) Web usage mining for predicting final marks of students that use moodle courses. Comput Appl Eng Educ 21(1):135–146

34. Romero C, Ventura S (2007) Educational data mining: a survey from 1995 to 2005. Expert Syst Appl 33:135–146

35. Romero C, Ventura S (2010) Educational data mining: a review of the state of the art. IEEE Trans Syst Man Cyber Part C Appl Rev 40(6):601–618

36. Romero C, Ventura S, Pechenizkiy S, Baker M (2010) Handbook of educational data mining. Data mining and knowledge discovery series. Chapman & Hall, London

37. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart D, McClelland J (eds) Parallel distributed processing: explorations in the microstructure of cognition. MIT, Cambridge, pp 318–362

38. Triguero I, Sáez JA, Luengo J, García S, Herrera F (2014) On the characterization of noise filters for self-training semi-supervised in nearest neighbor classification. Neurocomputing 132:30–41

39. van Rijsbergen CJ (1979) Information retrieval, 2nd edn. Butterworths, London

40. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng AFM, Liu B, Yu PS, Zhou ZH, Steinbach M, Hand DJ, Steinberg D (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14(1):1–37

41. Xu L, Krzyzak A, Suen CY (1992) Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Trans Syst Man Cybern 22(3):418–435