Enhancing High School Students' Performance based on Semi-Supervised Methods

Georgios Kostopoulos
Educational Software Development Laboratory (ESDLab)
Department of Mathematics, University of Patras
Patras, Greece
kostg@sch.gr

Ioannis E. Livieris
Department of Computer Engineering Informatics (Disk Lab)
Technological Educational Institute of Western Greece
Patras, Greece
livieris@teiwest.gr

Sotiris Kotsiantis
Educational Software Development Laboratory (ESDLab)
Department of Mathematics, University of Patras
Patras, Greece
sotos@math.upatras.gr

Vassilis Tampakas
Department of Computer Engineering Informatics (Disk Lab)
Technological Educational Institute of Western Greece
Patras, Greece
vtampakas@teimes.gr

***Abstract*—High school educators evaluate students' performance on a daily basis using several assessment methods. Identifying weak and low performance students as soon as possible during the academic year is of utmost importance for teachers and educational institutions. Well planned assignments and activities, additional learning material and supplementary lessons may motivate students and enhance their performance. Over recent years, educational data mining has led to the development of several efficient methods for the prediction of students' performance. Semi-supervised learning constitutes the appropriate tool to exploit data originated from educational institutions, since there is often a lack of labeled data, while unlabeled data is vast. In our study, several well-known semi-supervised techniques are used for the prognosis of high school students' performance in the final examinations of the "Mathematics" module. The experiments results demonstrate the efficiency of semi-supervised learning methods, and especially Self-training, Co-training and Tri-training algorithms, compared to familiar supervised methods.**

Keywords*—*Semi-supervised learning; Self-training; Tri-training; Co-training; Naïve Bayes; C4.5 Decision tree; SMO; kNN; prediction; student performance; high school;

I. Introduction

Over recent years, the need for exploitation and analysis of data originated from educational institutions has given rise to a substantial growth of data mining applications in education. Educational Data Mining (EDM) is the key tool for understanding students' learning behavior and predicting their academic performance [22]. The latter is regarded as the most interesting and well-studied aspect of EDM as confirmed by the development of various machine learning methods.

The purpose of this paper is twofold. Initially, we examine the effectiveness of semi-supervised learning (SSL) methods for the prognosis of high school students' grade in the final examinations of the "Mathematics" module at the end of the academic year. The students' grade has been classified into four classes and is based on several time-variant quantitative attributes of students, such as written assignments, oral performance, short tests and exams that have been performed during the two academic semesters. In addition, we investigate the possibility to identify low performance students in a good time during the academic year quite accurately. Identifying strengths and weaknesses of such students is of utmost importance for teachers and educational institutions. Students that are likely to fail in the final examinations need extra help and learning support. Well planned assignments and activities, additional learning material and supplementary lessons adapted to the different needs and knowledge levels of students may motivate them and enhance their performance. To the best of our knowledge there are a limited number of studies dealing with the implementation of semi-supervised learning (SSL) methods in the educational field, and particularly classification methods. In [11] is shown the effectiveness of semi-supervised classification methods for predicting students' performance in distance higher education.

The rest of this paper is organized as follows: In Section II we present recent studies of data mining applications in education, most of which concern the supervised learning, and especially the classification task. In Section III we briefly refer to SSL and provide a short report of the algorithms that are used in the experiments. A description of the data set is given in section IV together with a detailed analysis of data attributes. In section V we analyze the experiments carried out in this study using familiar SSL algorithms and present their results while making a comparison to well-known supervised methods. Finally, in Section VI we conclude writing down some thoughts for future work.

II. A RECENT REVIEW OF DATA MINING APPLICATIONS IN EDUCATION

Several studies deal with the implementation of machine learning techniques to evaluate students' performance attending courses in educational institutions. A large proportion of these studies examines the efficiency of supervised methods, especially classification (usually pass or fail), while SSL methodologies have been rarely applied to the educational field. Surveys of EDM applications have been presented in [1, 21, 22]. A number of rewarding studies have being carried out in recent years and some of them are presented below:

Cortez and Silva [3] parsed data originated from two secondary schools to predict students' performance (pass or fail) in "Mathematics" and "Portuguese language" modules in the final examinations at the end of academic year. Four familiar data mining methodologies, particularly Decision Trees, Random Forest (RF), Neural Networks (NNs) and Support Vector Machines (SVM), were tested in several demographic, social and school attributes showing a high predictive accuracy, especially in the case where the past school period grades were known.

Kotsiantis et al. [12] proposed an online ensemble of supervised algorithms to predict the performance on the final examination test of students attending distance courses in higher education. Naïve Bayes (NB) classifier, WINNOW (a linear online algorithm) and k-NN classifier constituted an online ensemble operating in incremental mode, while using the majority voting methodology for the output prediction (pass or fail). The proposed ensemble of classifiers outperformed well-known algorithms, such as the RBF, BP, C4.5, k-NN and SMO algorithm, and could be used as a predictive tool from tutors during the academic year to underpin and boost low performers.

Osmanbegovic and Suljic [18] tested the efficiency of three classification techniques (C4.5 Decision tree, NB and Multilayer Perceptrons) to predict students' performance during the summer semester at the Faculty of Economics in Tuzla, of the 2010 academic year. The NB method prevailed over the other two methods, with a prediction accuracy measure at 76.65%.

Kabakchieva [10] studied the impact of demographic and performance attributes for the prediction of students' access at the University of National and World Economy in Bulgaria. For the prediction of the five class output attribute, several experiments were conducted using popular Weka classifiers (J48 decision tree, NB and BayesNet classifiers, k-NN (IBk) algorithm, OneR and JRip rule learners). The results were not remarkable, showing that the best performer was the J48 classifier (66% accuracy), while the less accurate were OneR (54-55%) and NB classifiers (below 60%).

Mashiloane and Mchunu [15] studied the performance of three well-known classification algorithms (J48 decision tree, NB and Decision Table) for predicting first year students' failure in the School of Computer Science at the University of Witwatersrand. Student data from recent years were used for the training phase identifying J48 classifier as the best performer. In the testing phase, 92% of the instances were predicted correctly, indicating that decision trees can be a powerful tool in predicting first year students' performance precisely from the middle of the academic year.

In more recent works, Kostopoulos et al. [11] applied SSL methods for predicting students' performance in distance higher education. Several experiments were con-ducted using a variety of SSL algorithms from KEEL [27]. The experimental results showed the effectiveness of SSL methods, especially the Tri-training algorithm, in contrast to familiar supervised methods such as the C4.5 decision tree.

Livieris et al. [14] presented a user-friendly decision support software for predicting the students' performance, together with a case study concerning the final examinations in the course of "Mathematics''. Based on their preliminary results the authors concluded that the application of data mining can gain significant insights student progress and performance.

Sweeney et al. [26] examined the efficiency of RF, Factorization Machines (FM) and Personalized Linear Multiple Regression methods to predict students' success and retention rates in higher education. Moreover, a hybrid recommender system technique combining RF and FM is developed to predict students' grades based on the performance of previous terms.

Spoon et al. [25] presented a method named Individualized Treatment Effects (ITE) for evaluating students' performance and identifying students at risk in a statistics course. ITE is based on RF, ensembles of classification and regression trees, which split students into similar performance groups. Specifically, students who enrolled to an introductory statistics course could voluntarily enroll in a supplemental instruction section. This supplemental section is the core of the method and is used to identify students' performance as well as the factors influencing students' success based on data available at the beginning of the semester.

It is evident that several approaches, methodologies and algorithms have been developed over recent years exploring and exploiting the educational data by using classification, regression and visualization techniques to understand the academic "behavior" of students and predict their performance.

III. SEMI-SUPERVISED LEARNING

SSL is a mixture of supervised and unsupervised learning aiming to obtain better results from each one of these methods by using a small amount of labeled examples together with a large amount of unlabeled ones. Depending on the nature of the output variable, SSL is subdivided into two main categories: semi-supervised classification for discrete output variable and semi-supervised regression for real-valued.

Various SSL algorithms have been implemented in recent years with remarkable results in many scientific fields, such as Self-training [29], Co-training [2], Democratic Co-learning [32] and Tri-training [31], De-Tri-training [5] and RASCO [28]. These methods are trying to take as much advantage of the unlabeled data as possible, since the utilization of unlabeled data is essential for their efficiency [23].

Self-training or self-teaching is considered to be a simple and widely used SSL method. According to Ng and Cardie

(2003) self-training is a “single-view weakly supervised algorithm” [16]. Initially, a small amount of labeled data constitutes the training set. A classifier is trained which is subsequently used in classifying the unlabeled data. The training set is gradually augmented using the most confident predictions and the procedure is repeated until all unlabeled data are finally labeled. Self-Training is a bootstrapping method since it is based on its own predictions to teach itself, so wrong predictions of the classifier on the initial steps often lead to misclassifications of the labeled data [6].

Co-training is a semi-supervised method proposed by Blum and Mitchell (1998) and is based on the following three assumptions [2]. Each example of the data set can be partitioned in two distinct views that are not perfectly correlated (multi-view assumption), which are conditionally independent given the class label (independence assumption). Moreover, each view can effectively be used for classification (compatibility assumption). In this framework, two classifiers are trained separately in each view using a small set of labeled examples, and the most confident predictions of each algorithm on unlabeled data is used to augment the training set of the other. The efficiency of the Co-training algorithm depends mainly on the fulfillment of the above assumptions [17] as well as the proper choice of classifiers. A significant amount of research deals with the implementation of the Co-Training algorithm for SSC. Although the assumptions about the existence of sufficient and redundant views can hardly be met in practice, several extensions of the Co-Training algorithm have been developed such as Tri-training [31], De-Tri-training [5], Democratic Co-training [32], Co-Forest [13] and CoBC [7].

The existence of two independent views on a data set can hardly be met. In most cases, such views are not presented. Democratic Co-learning and Tri-training tackle this problem, since they do not require two sufficient and redundant views such as the original Co-training algorithm.

Democratic Co-learning [32] is a single view extension of the Co-training algorithm exploiting a small amount of labeled data together with a large amount of unlabeled data. Three different supervised learning algorithms train a set of classifiers separately on the same set of labeled data. More specifically, every learner predicts a label for an unlabeled example, which is labeled and added to the labeled subset if the majority of learners agree on the label. The augmented labeled data set is used to retrain the learners and the procedure is repeated until all unlabeled data are finally labeled.

Tri-training algorithm is also based on the co-training paradigm [31]. In contrast to Democratic Co-learning algorithm, Tri-training does not require different supervised algorithms, leading to greater applicability and implementation of the algorithm in many real world data sets. It uses three classifiers that are initially trained on labeled examples. If two of the classifiers agree on labeling an unlabeled example, then this example is used to train the third one.

Differential Evolution Tri-training (De-Tri-training) algorithm is a semi-supervised clustering method which is built on the Tri-training approach to enlarge the scale of the initial seeds set [5]. In addition, a k-NN rule based data editing technique is applied to decrease the impact of misclassified instances during the initial stages of the learning process and improve the efficiency of the algorithm.

Random subspace method for Co-training (RASCO) is an extension of the Co-training algorithm to the multiple view setting [28]. RASCO chooses multiple random subspaces of the feature space and trains a supervised classifier in each subspace such as a decision tree classifier (J4.8 was originally used). These classifiers complement one another and are used for Co-training enlarging the data set with the most confident predictions. Critical points for the efficiency of the method are the dimensionality and the number of subspaces, as well as the construction and cooperation of the classifiers.

IV. DATA DESCRIPTION

The data set used in our study has been provided by the Microsoft showcase high school “Avgoulea-Linardatou” in Athens. For a time period of five years (2007-2012), data of 340 students of ages 14-15 years have been collected concerning the “Mathematics” module. During the academic year, teachers are required to use a variety of assessment methods including written assignments, oral examination, short tests and exams. Moreover, students are obliged to attend the final examinations of the module at the end of the academic year. The final exam is marked out of 20, and is of prime importance to the overall final grade of the specific module.

TABLE I. ATTRIBUTES DESCRIPTION

Attribute	Type	Values	Description
ORAL_A	integer	[1, 20]	1st semester’s oral grade
TEST_A1	real	[1, 20]	1st semester’s test1 grade
TEST_A2	real	[1, 20]	1st semester’s test2 grade
EXAM_A	real	[1, 20]	1st semester’s exam grade
GRADE_A	integer	[1, 20]	1st semester’s overall grade
ORAL_B	integer	[1, 20]	2nd semester’s oral grade
TEST_B1	real	[1, 20]	2nd semester’s test1 grade
TEST_B2	real	[1, 20]	2nd semester’s test2 grade
EXAM_B	real	[1, 20]	2nd semester’s exam grade
GRADE_B	integer	[1, 20]	2nd semester’s overall grade
EXAMS	ordinal	0-9, 10-14, 15-17, 18-20	Grade in final examinations

Each instance in the data set is characterized by the values of 10 time-variant at-tributes (Table I). The assessment of students during the academic year consists of two 15-minute pre-warned tests, oral examination, several written assignments and a 1-hour exam in each semester (semester A, semester B). The 15-minute tests (TEST_A1, TEST_A2, TEST_B1, TEST_B2) include short answer problems and multiple choice questions. The 1-hour exams (EXAM_A, EXAM_B) cover a wide range of the curricula and include several theoretical and multiple choice questions, as well as a variety of problems requiring arithmetic skills, solving techniques and critical analysis, explaining mathematical situations and understanding of the basic mathematical terms and concepts. Several written assignments and frequent oral questions assess students’ understanding of important concepts and topics in mathematics daily in each semester (ORAL_A, ORAL_B). Finally, the overall semester performance of each student, which addresses the personal engagement of the student in the lesson and his progress, corresponds to attributes GRADE_A (semester A) and GRADE_B (semester B). The output attribute "EXAMS" corresponds to the students' grade in the final examinations (2-hour exam) according to the following four-level classification: 0-9 (poor), 10-14 (good), 15-17 (very good), 18-20 (excellent).

V. EXPERIMENTAL SETUP AND RESULTS

We divided the data set into 10 equally sized folds using the 10-fold cross validation procedure provided by KEEL, each of which was divided into two parts, the training set and the test set. The training set constitutes 90% of the data and was used to train the model, while the rest 10% constitute the test set and was used for the evaluation of the model. For each one of the training sets we used a label ratio of 20%, that is to say 20% of data instances are labeled and the rest 80% are unlabeled.

Our experiments were conducted in two distinct phases of two sequential steps each time. In each phase, the first step consists of the five attributes (ORAL_A, TEST_A2, TEST_A2, EXAM_A, GRADE_A) referred to the assessment of a student during the first semester, while in the second step all attributes of both semesters are used. It should be mentioned that all the attributes are being added gradually during the academic year.

A. The 1st Phase of Experiments

In the 1st phase of experiments we evaluate the performance of various SSL algorithms included in KEEL, and in particular Self-training, Co-training, Tri-training, De-Tri-training and Democratic Co-learning. Several supervised classifiers are used in each algorithm, such as the NB [9], the C4.5 Decision tree [20], the k-NN [4] and the Sequential Minimal Optimization (SMO) [19]. The SSL procedure that is used in our experiments is depicted below (Fig. 1):

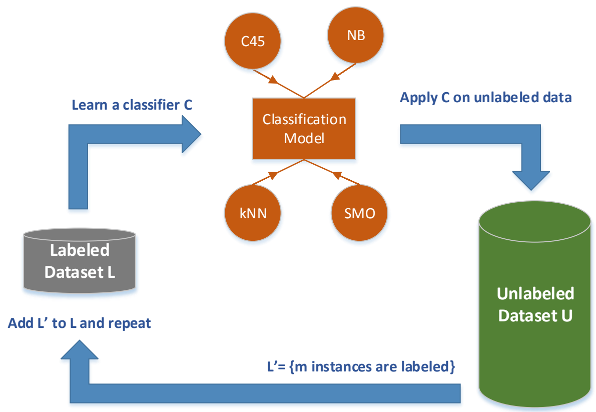

Fig. 1. SSL Procedure

Initially, we measure the accuracy of these algorithms, which corresponds to the percentage of the correctly classified instances. The accuracy performance of the SSL algorithms is presented in Table II. Self-training (NB), Tri-training (NB), Co-training (NB), Co-training (C4.5) and Democratic algorithms appear to be superior in the 1st step of the experiments based on the attributes regarding the first semester's assessment, with an accuracy measure between 63.53% and 67.35%.

TABLE II. THE ACCURACY (%) OF THE SSL ALGORITHMS

Algorithm	1st step (semester A)	2nd step (end of semester B)
Self-Training (C4.5)	58.82	63.53
Self-Training (kNN)	61.76	62.65
Self-Training (NB)	67.35	72.94
Self-Training (SMO)	58.24	64.41
De-Tri-Training (C4.5)	58.53	60.29
De-Tri-Training (kNN)	60.59	63.24
De-Tri-Training (NB)	62.65	67.94
De-Tri-Training (SMO)	61.18	67.35
Tri-Training (C4.5)	60.88	65.88
Tri-Training (kNN)	58.24	57.06
Tri-Training (NB)	65.59	71.76
Tri-Training (SMO)	59.71	64.71
Co-Training (C4.5)	65.88	67.94
Co-Training (NB)	64.41	72.06
RASCO (C4.5)	49.41	54.12
RASCO (kNN)	52.06	52.06
RASCO (NB)	58.53	67.06
RASCO (SMO)	46.47	48.82
Democratic	63.53	68.24

In the second step, Self-training accuracy measure is 72.94%, while Tri-training (NB) and Co-training (NB) exceed 71%. Moreover, there is an increase of accuracy measure for all SSL algorithms by adding the attributes of the second semester (2nd step). We evaluate the performance using the Friedman Aligned Ranks nonparametric test [8]. According to the test results (Table III) the algorithms are ranking from the best performer to the lower one.

TABLE III. FRIEDMAN ALLIGNED RANKS TEST

Algorithm	Rank
Self-Training (NB)	2.50
Tri-Training (NB)	4.50
Co-Training (NB)	4.50
Co-Training (C4.5)	7.25
Democratic	9.50
De-Tri-Training (NB)	11.75
De-Tri-Training (SMO)	14.50
Tri-Training (C4.5)	16.50
RASCO (NB)	20.25
Tri-Training (SMO)	20.50
Self-Training (kNN)	21.00
De-Tri-Training (kNN)	22.00
Self-Training (C4.5)	24.00
Self-Training (SMO)	25.25
De-Tri-Training (C4.5)	29.25
Tri-Training (kNN)	30.75
RASCO (kNN)	34.50
RASCO (C4.5)	34.50
RASCO (SMO)	37.50

B. The 2n Phase of Experiments

In the 2nd phase of experiments we make a comparison between the SSL algorithms that outweigh in the 1st step (Self-training, Tri-training, Co-training, De-Tri-training and Democratic) and a familiar supervised algorithm, in particular the Naïve Bayes (NB) classifier. NB is considered to be a very effective and simple classification algorithm, a representative form of the Bayesian network. Its effectiveness is based on the conditional independence assumption, according to which, all attributes are independent given the value of the class attribute [30].

The results (Table IV) show that the accuracy measure of the NB classifier ranges from 65.30% in the first step to 71.47% in the second step. Moreover, the SSL algorithms are comparatively better than the respective supervised algorithm in both steps, verified also from the Friedman Aligned Ranks test (Table V). Self-training (NB), Tri-training (NB) and Co-training (NB) take precedence over the Naïve Bayes method. The most efficient algorithm is Self-training (NB) scoring an accuracy measure of 67.35% in the 1st step and 72.94% in the 2nd, while the Naïve Bayes scores 65.30% and 71.47% respectively. Moreover, Self-training (NB), Tri-training (NB) and Co-training (C4.5) score between 65.59% and 67.35% at the end of the first semester showing that an accurate prognosis of weak and low performance students may be done in sufficient time.

TABLE IV. ACCURACY (%) COMPARISON (NB BASE CLASSIFIER)

Algorithm	1st step (semester A)	2nd step (end of semester B)
Naïve Bayes	65.30	71.47
Tri-Training (NB)	65.59	71.76
Co-Training (NB)	64.41	72.06
Self-Training (NB)	67.35	72.94
De-Tri-Training (NB)	62.65	67.94

TABLE V. FRIEDMAN ALLIGNED RANKS TEST

Algorithm	Rank
Self-Training (NB)	1.5
Tri-Training (NB)	4.5
Co-Training (NB)	5.5
Naïve Bayes	6.5
De-Tri-Training (NB)	9.5

VI. CONCLUSIONS

The purpose of this paper is to examine the effectiveness of semi-supervised methods for the performance prediction of high school students in the final examinations in the "Mathematics" module. More specifically, attributes related to written assignments, oral examinations, short tests and exams during the academic year are marked according to specific assessment criteria and are used to evaluate the final grade in exams using SSL methods with a considerable accuracy, as reflected from the experiment results. Self-training, Tri-training, Co-training prevail over efficient supervised methods, such as the NB classifier. SSL seems to be the appropriate tool for predicting students' performance in educational institutions, since it requires having labels for a limited data set, while at the same time it is difficult for educators to obtain a relatively large amount of labeled data.

One of the main queries of our study is how early can we predict students' performance in the final examinations of the academic year. As illustrated in Table II, teachers may recognize possible weak and low performance students before the end of the first half of the academic year based on the continuous assessment of students during the first semester. Self-training (NB) scores 67.35% accuracy at the end of the first semester showing that a confident prognosis of the final performance of students can be done. Fairly similar accuracy percentages to the previous SSL classifier achieve the Tri-training (NB) and Co-training (C4.5) classifiers (65.59% and 65.88% respectively).

This study was based on an off-line learning, since the learning methods were applied after the data was collected. There is need for an automatic on-line learning environment, by using a student prediction engine as part of a school management support system. That will allow the collection of additional variables (e.g. grades from previous school years). Another interesting topic is the implementation of semi-supervised regression (SSR) and active learning methods [24] in the educational field. EDM is principally engaged with classification problems and mostly supervised methods such as classification and regression for predicting students' performance in higher education and distance learning. Educational data mining using semi-supervised techniques is a hot topic in machine learning in recent years. So, it be-comes evident that the implementation and application of SSR as well as active learning methods in EDM are of particular importance.

REFERENCES

[1] Baker, R. S., and Yacef, K., The state of educational data mining in 2009: A review and future visions. JEDM-Journal of Educational Data Mining, vol. 1(1), pp. 3-17, 2009

[2] Blum, A., and Mitchell, T., Combining labeled and unlabeled data with co-training. In: 11th annual conference on Computational learning theory, pp. 92-100. ACM, 1998

[3] Cortez, P., and Silva, A. M. G., Using data mining to predict secondary school student per-formance, 2008

[4] Cover, T., and Hart, P., Nearest neighbor pattern classification. IEEE transactions on infor-mation theory, vol. 13(1), pp. 21-27, 1967

[5] Deng, C., and Zu Guo, M., Tri-training and data editing based semi-supervised clustering algorithm. In: Mexican International Conference on Artificial Intelligence, pp. 641-651. Springer Berlin Heidelberg, 2006

[6] Goldberg, A. B., New directions in semi-supervised learning. Doctoral dissertation, Uni-versity of Wisconsin-Madison, 2010

[7] Hady, M. F. A., and Schwenker, F., Combining committee-based semi-supervised learning and active learning. Journal of Computer Science and Technology, vol. 25(4), pp. 681-698, 2010

[8] Hodges, J. L., and Lehmann, E. L., Rank methods for combination of independent experi-ments in analysis of variance. The Annals of Mathematical Statistics, vol. 33(2), pp. 482-497, 1962

[9] John, G. H., and Langley, P., Estimating continuous distributions in Bayesian classifiers. In Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, pp. 338-345. Morgan Kaufmann Publishers Inc., 1995

[10] Kabakchieva, D., Predicting Student Performance by Using Data Mining Methods for Classification. vol. 13(1), pp. 61-72. Cybernetics and Information Technologies, 2013

[11] Kostopoulos, G., Kotsiantis, S., and Pintelas, P., Predicting Student Performance in Distance Higher Education Using Semi-supervised Techniques. In: Model and Data Engineering, pp. 259-270. Springer International Publishing, 2015

[12] Kotsiantis, S., Patriarcheas, K., and Xenos, M., A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. Knowledge-Based Systems, vol. 23(6), pp. 529-535, 2010

[13] Li, M., and Zhou, Z. H., Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 37(6), pp. 1088-1098, 2007

[14] Livieris I.E., Mikropoulos T.A., and Pintelas P., A decision support system for predicting students' performance. Themes in Science and Technology Education, vol. 9(1), pp. 43-57, 2016

[15] Mashiloane, L., and Mchunu, M., Mining for Marks, A Comparison of Classification Algo-rithms when Predicting Academic Performance to Identify "Students at Risk". In Mining Intelligence and Knowledge Exploration, pp. 541-552. Springer International Publishing, 2013

[16] Ng, V., and Cardie, C., Weakly supervised natural language learning without redundant views. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, vol. 1, pp. 94-101. Association for Computational Linguistics, 2003

[17] Nigam, K., and Ghani, R., Analyzing the effectiveness and applicability of co-training. In Proceedings of the ninth international conference on Information and knowledge management, pp. 86-93. ACM, 2000

[18] Osmanbegović, E., and Suljić, M., Data mining approach for predicting student performance. Economic Review, vol. 10(1), 2012

[19] Platt, J., Sequential minimal optimization: A fast algorithm for training support vector machines, 1998

[20] Quinlan, J. R., C4.5: Programs for Machine Learning. Elsevier, 1993

[21] Romero, C., and Ventura, S., Educational data mining: A survey from 1995 to 2005. Expert systems with applications, vol. 33(1), pp. 135-146, 2007

[22] Romero, C., and Ventura, S., Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40(6), pp. 601-618, 2010

[23] Seok, K., Semi-supervised regression based on support vector machine. Journal of the Ko-rean Data and Information Science Society, vol. 25(2), pp. 447-454, 2014

[24] Settles, B., Active learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 6(1), pp. 1-114, 2012

[25] Spoon, K., Beemer, J., Whitmer, J., Fan, J., Frazee, J., Stronach, J., Bohonak, A., and Levine, R., Random Forests for Evaluating Pedagogy and Informing Personalized Learning. JEDM-Journal of Educational Data Mining, vol. 8(2), pp. 20-50, 2016

[26] Sweeney, M., Rangwala, H., Lester, J., and Johri, A. Next-Term Student Performance Prediction: A Recommender Systems Approach. arXiv preprint arXiv:1604.01840, 2016

[27] Triguero, I., García, S., and Herrera, F., Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. Knowledge and Information Systems, vol. 42(2), pp. 245-284, 2015

[28] Wang, J., Luo, S. W., and Zeng, X. H., A Random Subspace Method for Co-training. In: IEEE International Joint Conference on Neural Networks, pp. 195-200. IEEE, 2008

[29] Yarowsky, D., Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In: Proceedings of the 33rd annual meeting on Association for Computational Linguistics, pp. 189-196. Association for Computational Linguistics, 1995

[30] Zhang, H., The optimality of naive Bayes. AA, vol. 1(2), pp. 3, 2004

[31] Zhou, Z. H., and Li, M., Tri-training: Exploiting Unlabeled Data Using Three Classifiers. In: IEEE Transactions on Knowledge and Data Engineering. vol. 17(11), pp. 1529-1541. IEEE, 2005

[32] Zhou, Y., and Goldman, S., Democratic Co-learning. In: ICTAI 2004, pp. 594-602. IEEE, 2004