

Design of an Advanced Smart Forum for Tesys e-Learning Platform

Paul Ștefan Popescu^{1(✉)}, Mihai Mocanu¹, Costel Ionașcu²,
and Marian Cristian Mihăescu¹

¹ Faculty of Automation, Computers and Electronics,
Department of Computers and Information Technology,
University of Craiova, Craiova, Romania
{stefan.popescu,mmocanu,mihaescu}@software.ucv.ro

² Faculty of Economics and Business Administration,
University of Craiova, Craiova, Romania
icostelm@yahoo.com

Abstract. This paper presents an application of Intelligent Data Analysis techniques in the area of online educational environments and more exactly, the discussion forums within them. The research area is also referred as Educational Data Mining and have many tools and techniques already developed. This work concentrates on the improvement that can be provided by the design and implementation of a forum that has “smart” capabilities and aims to be proactive to the user’s needs. The main issues addressed are the interaction design, student’s academic performance and the achievement of better models by completing the already gathered data with the logged data offered by the forum. We present here three methods that can solve the above mentioned issues: recommending subjects of interest, computing trends and offering smart alerts for users that are at risk for academic failure. Every method represents a tool that will be integrated in the forum and will take benefit from the extra logged data.

Keywords: e-Learning · Machine learning · Forum

1 Introduction

Online educational environments are a constantly growing field of education. This field is important because it creates the opportunity to learn even when there is no physical presence like in regular educational environments. The physical presence in the classroom is important in terms of interaction because students are able to talk easily, change opinions and help each others. From the teachers perspective they can easily response to questions, and create some sort of ranking in terms of knowledge. There will always be some learners that have better knowledge in some educational areas; these learners can help share their knowledge to others.

In online educational environments interaction of any kind is different [1] than in regular environments, the engagement in educational activities is also different [2] and this may lead to lower student’s results [3]. Going further from this lack of interaction

e-learning platforms integrates more and more smart capabilities; from better data logging that allows data analysts to accomplish more advanced data mining and machine learning tasks to advanced smart tools that offers many facilities (i.e. subjects of interest recommendations, predicting students failure or academic performance, better data logging).

In this paper we present a forum that encapsulates smart capabilities aimed to improve the student's interaction and to create a friendlier virtual environment included in e-learning platforms. The system's design aims to ensure enough flexibility to easily adjust and install in any e-Learning platform but first we will refer Tesys e-Learning platform that runs at the University of Craiova and offers us data. Tesys is a custom e-learning platform built at the University of Craiova which fulfils the professors and students needs in terms of online educational environments necessary for distance learning. The platform is under continuous development and based on the knowledge gained from we can make further development like this Smart Forum.

The forum is designed to be strongly integrated in the e-learning platform, being available in every discipline, and well isolated in terms of implementation. We don't design it as a plug-in that uses some sort of bridge to relate between forum and e-learning platform. This sort of integration ensures more specific and subject related topics of discussion.

2 Related Work

In the area of educational data mining there has been a lot of work. This paper lies on the border of machine learning and information retrieval offering a design of a system that encapsulates feature extraction techniques that are the prerequisites for employing data mining and machine learning algorithms.

Some other forums were used to employ data mining techniques and to gain more knowledge from discussions. The purpose of [4] is to present some data mining techniques that offer a strategy for data representation. For their approach the instructor's view of the output of a thread forum is somehow limited as the can review a transcript of the written dialogue produced by participants. Because of big amount of data that consists of forum contributions the paper seeks to intersect the information an instructor wish to extract from the forum with some useful information that the system may extract from the instructor's query. This will help the instructor to improve his ability to evaluate the progress of a threaded discussion.

There are also some other paper [5] that show that students who use discussion forums have a higher chance of finishing a course. This implies MOOCs (Massive Online Open Courses) and they use some machine learning techniques to extrapolate a small set of annotation to the whole forum. These annotations help in two main ways: summarize the state of the forum and the also allow researchers to deeper understand how the forum is implied in the learning process.

Some recent research [6] predicts the students performance based on on-line discussion forums which are considered as communities of people learning from each other. They aim that forums not only inform the students about their peers' doubts or problems

but it is a way to inform the professors about the learner's knowledge. In the paper they select some instances and attributes, run different supervised learning algorithms and then they measure the accuracy and comprehensibility of the prediction.

Some related work was done for the design and development of several machine learning algorithms [7]. For the supervised learning tasks we can use decision trees [8] or regression analysis [9] and we aim to evaluate the system using several machine learning metrics like area under the ROC curve [10].

3 Goals

Building and integrating *Smart Forum* aims to achieve three goals: create a virtual space for improving the users' interaction, generate more high quality logged data that can be analyzed and offering useful functionalities for improving the students' performance.

3.1 Interaction Improvement

Improving the users' interaction can be achieved by extending the means of interaction by number and quality. Forums are environments that allow message exchange, the particularity refers that the messages are most public or available for registered users and also the messages are grouped on subjects of discussion. For our approach, messages wrote on Smart Forum are available for students and instructors from a specific course. Adding a new mean of communication with specific features creates a environments which impact can be measured in terms of interaction. After the system is integrated and used for at least a semester, we evaluate the interaction and adjust some parameters like notification frequency or messages of interest alert frequency.

3.2 Data Logging

Data already logged in e-learning platforms refers users and learning resources. We threat Smart Forum as an important learning resource which can produce a important amount of data that refers both other learning resources and users. We can analyze what chapters of course were more discussed and what concepts from which chapters are referred in the most discussed topics. For user profiling we can add more data based on the actions that are perform on the forum. Logging more data can reveal several new patterns and also improve the actual models.

3.3 Improving the Student's Performance

This is a permanent topic of interest in the Educational Data Mining Area. There are so many solutions out there but there is also enough space for improvement. Collecting more relevant data and adding it to the existing models may lead the data analyst to better result. There are also many correlations that can be made between the forum activity and educational results, forum activity and the activity performed over other learning resources or the overall activity and final grades. One thing that may influence

the student’s performance is also the interaction between them and what they can learn from each other. By using the described platform one study can reveal if students that interact more and learn more from their colleagues will have better results or at least an ascending learning curve.

4 System Design

Smart forum creates a virtual space for interaction between all entities that perform activities in e-Learning platforms: students, professors and administrative staff. Every entity has his specific tasks that are related to their topics of interests; students discuss about the subjects related to the disciplines, professors may response to questions and may launch challenges, the administrative staff may respond to administrative questions of common interest.

Figure 1 presents how the disciplines are distributed over the Tesys e-Learning platform’s infrastructures. The e-learning platform holds 3 (but it can be more) years of study, every year of study have several modules (study programs) which are marked with M and every study program have a set of disciplines attached to it D. Because of the space problems we added in the figure disciplines only on some modules but there are usually a set of around five disciplines on every module. This figure is important to understand because smart forum is attached to every discipline being well distributed. The forum distribution over the disciplines ensures that the subjects discussed in the forum are strongly related to the course chapters that are used to create de discipline.

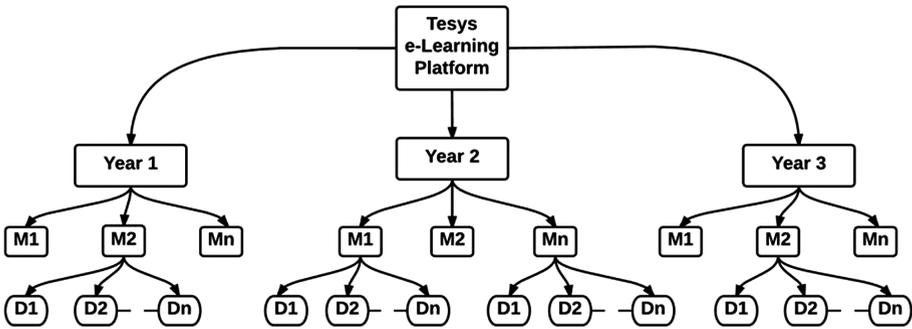


Fig. 1. Tesys disciplines distribution

Figure 2 presents the components that are encapsulated in a discipline. On the upper part of the picture we have the main actors that perform their activities on one discipline: students, professors, and secretaries. Every actor has his own interface and access to specific features of the e-Learning platform. Chapters, Homework, questions and references come at the same level with the forum and are related to the discipline covering

several learning areas. The chapters are represented by documents that form the course, homework are used to assign learners tasks that take more time while questions are used for tests and exams. The forum is used to discuss subjects related to a specific discipline; this makes the data collected from it to be more specific. For example, suppose a data analyst wants to analyze the subjects related to discipline “x”, he needs to find the specific subjects related to the discipline “x” by using some analysis techniques which will most likely not offer 100 % accuracy or the topics will have tags that were allocated by the topic initiator which will also may not offer 100 % accuracy all the time. Distributing the forum over the discipline offers one more topic focus attribute over the existing ones. On the forum there will always be moderators which are represented by professors or/and best students, these moderators will make sure there will not be topics out of the discipline area on a specific forum.

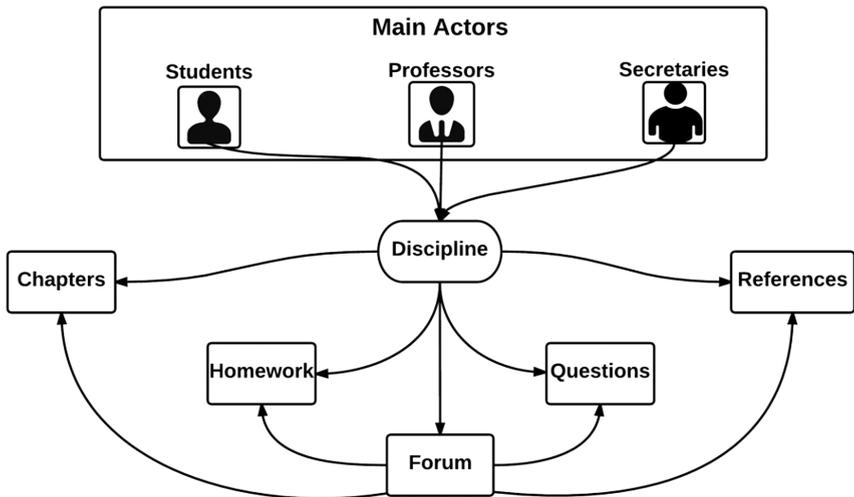


Fig. 2. Forum integration

Figure 3 presents the main architecture of the data analysis pipeline. There are two main components that contribute to data logging: Smart Forum and Tesys. Although the forum is integrated in the e-learning platform, it provides a specific amount of data that can be converted in a set of features that can be combined with the features extracted from Tesys in order to build better models. There are also some specific data analysis tasks that will be performed only on the data gathered from Tesys or Smart Forum and there will be some specific tasks that will be performed only on the combined model and these tasks will not be used for validation purposes.

Validation of the system in terms of data analysis can be performed by comparing the results obtained using the combined models with the models obtained using only the features from Tesys.

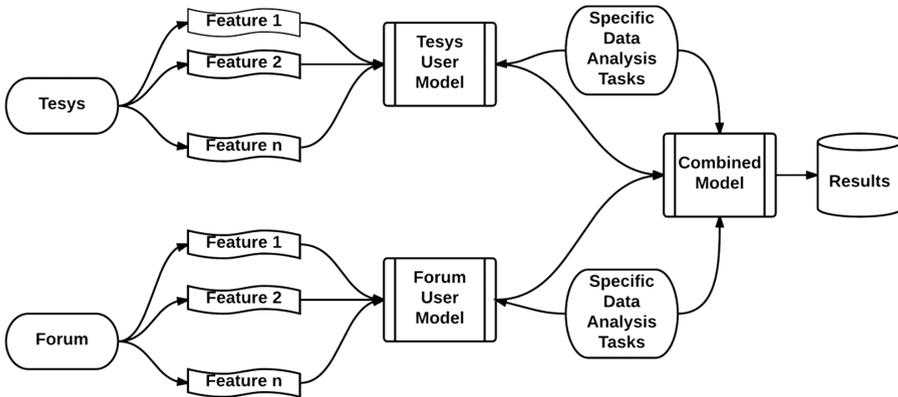


Fig. 3. Data analysis pipeline

5 Functionalities and Data Analysis Processes

In this section we present the main functionalities of the proposed system. These capabilities are specific to a forum that is integrated in an e-learning platform and are dependent on the data that is logged from the forum. There are two types of actions that are available in this forum: regular one and smart ones that includes the usage of several data mining and machine learning algorithms.

5.1 Proposed Features

Several features may be extracted using Smart Forum. These features ensure better user modelling for forum activity analysis and also may improve the overall student's profile.

- NO_POSTS – number of posts written on the forum
- NO_TOPICS – the number of topics started on the forum
- NO_QUESTIONS – the number of questions asked on forum
- NO_USEFUL_ANSWER – the number of answers written on the forum considered to be useful by other users
- NO_ACTIVE_DAYS_FORUM – the number of days when the user was active on the forum
- AVG_TIME_READ – the average time that passed between a new post/topic appeared and the moment of reading that post/topic
- NO_CONCEPTS_COURSE – the number of concepts extracted from a course that were referred by a students in his posts. This feature will give us a view of how focused on the course is a student when he posts.
- FRQ_POSTS – the frequency of posting messages on the forum
- FRQ_TOPICS – the frequency of posting new topics on the forum

Based on these features we can model the users and the activity but only after the experiments we can say which of the features will remain. Feature selection is a complex problem because we need to choose the features that are strongly related for the learner's activity but also we need to choose the ones that offer better results for the chosen algorithms.

5.2 Regular Functionalities

Posting New Topics. Posting messages and replies to a topic, in this capability we include the possibility to edit or delete a message for a specific amount of time. It is also important here to save the original message in the database for a period so the moderators can solve any complains.

Well Defined Profiles. User profiles include the number of messages posted on the forum, for the current discipline but also the number of the other posts. The profiles also refer the total number of answered questions and the number of questions that were correctly answered. Still regarding the user profiles we may have the number of questions from the forum that were answered by user and was marked as useful answer by the topic initiator or other users. This capability provides us with a vision over how trustworthy the student is when he answers at questions from the forum.

5.3 Smart Functionalities

The smart functionalities make the difference between a regular discussion forum and Smart Forum. These functionalities are enhanced by some data mining and machine learning algorithms. Extracting several features from this Smart Forum creates the possibility to implement the features that makes the forum to get the "smart" attribute. Below we present the main functionalities that can be implemented and a short description.

Subjects of Interest. Based on the educational results the forum may offer some subjects of interest for students. This capability will be accomplished by performing an analysis over the student's answered questions and text analysis over the posts. If we have a match between the concepts that can be extracted from the wrong answered questions and the concepts extracted from some specific posts or topics from the forum we can recommend those posts to the students. Going further for this capability we can also see what concepts are considered interesting by the student and we can recommend him/her some topics that weren't read by him/her.

Approach. This capability can be accomplished also by having a match between concepts extracted from the forum and the concepts extracted from the topics that were read or answered by the student and the questions that were correctly answered.

There are two directions to gain the subjects that may be interesting for a student: offering the ones that are addressed a lot or the ones that reveals lower results. In order to get the ones that are the most addressed we need to perform text analysis (i.e. concept extraction) on both the questions answered and on the messages posted on forum.

In Fig. 4 is presented an overview o the system. First we need to use clustering algorithms (i.e. SKM clustering algorithm [11]) for getting the learning that are most addressed by the students and the area from the forum that is most addressed then the concepts can be extracted using stemming algorithms from both collections. Having a bigger percent of concepts that match means a better matching and then we can recomend them to be read.

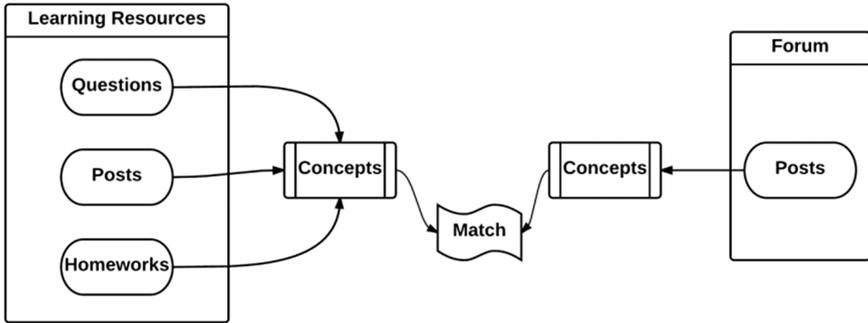


Fig. 4. Overview of the matching mechanism

In order to improve students’ performances we need to get the learning resources (questions and home works) on which the learner got lower results, then using stemming getting the common concepts from them. Based on posts analysis we can get the most related messages and offer them to the learner to be read.

Computing Trends. Computing trends for students will also be a smart capability available in Smart Forum. This capability is referring by drawing some charts that will reveal how the number of posts, number of questions answered that were marked as useful, number of questions asked or number of topics read are varying over the time. This can be useful for predicting but also for measuring the student’s engagement over the platform and, more specific, over the smart forum.

In Fig. 5 we present an example of trend that can be computed for a student over a full semester. On the OX axe we have the weeks from 1 to 12 and on OY axe we have the student’s activity. At a deeper analysis we can see that the student starts from a activity of 4.5 and have a small progress to the week 6, then there is a decline of two weeks and then he has an ascending trend until the last week of the semester. The green line with no dots on it is the trend computed based on the student’s activity. As we can see in this case is an ascending trend based on the student’s activity. This activity can be approximated using several attributes that defines it.

Figure 6 presents the trend discretized for three periods, on the OX axe we have the weeks and on OY on every of the third graphs we have the level of activity. In Fig. 6 the grey line with dots represents the activity, the green line withought dots ascending trends, the blue line represents a stable trend and the red line represents an descendent trend computed for the student.

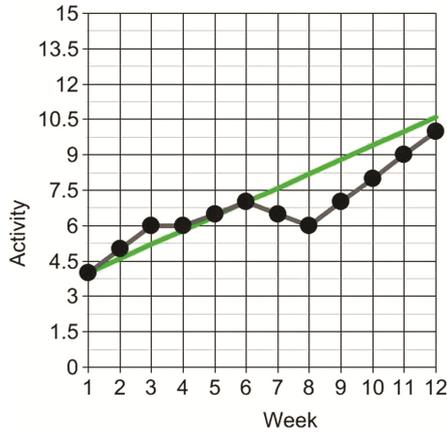


Fig. 5. Trend example

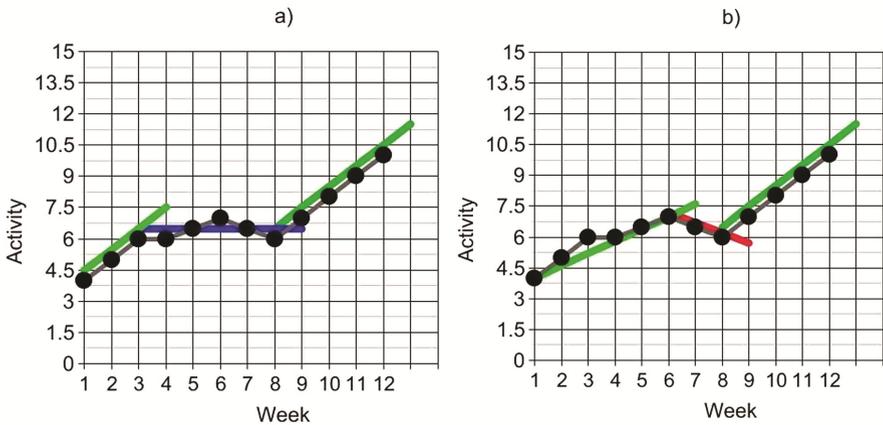


Fig. 6. Trend divided in periods

For the section (a) of the Fig. 6 the first period is represented by the first three weeks and have an ascending trend. We choose week number four to start the second period because there the trend is changing and we have stable trend until we reach week number 8 where is another trend switch. The last period consists of the last four weeks and is an ascending trend.

Another approach of computing trends is presented in (b) where we estimate the first 6 weeks at an ascending trend, then we have a descending trend between week six and 8 and then we have an ascending trend for the last weeks. This approach is easier to find if we parse the activity line during the period and brings up an alert for the user in week six as he might fail.

Approach. Computing trends based on students actions offers an overview of the student's evolution. This functionality can be accomplished using linear regression or

decision trees at specific checkpoints based on the extracted features. Usually we can produce two dimensional trends that take time and another attribute on both axes but that will be one chart for every attribute that can model a student and it is hard to estimate the trend. Our approach collects data from the forum and computes the features from Sect. 5.1; based on them we have two alternatives to address this issue.

Regression Approach: The regression approach aims to compute an estimated grade based on the features and add it on a chart having the time stamp and the grade on the axes.

$$Y_i = B_0 + B_1X_1 + \dots + B_nX_n + E \quad (1)$$

In Eq. (1) we present the computing of “ Y_i ” which is the grade computed for week “ i ”. The formula is standard for a linear regression and has three main components: “ B_i ” which are the coefficients assigned to a variable “ X_i ”, “ E ” which is the intercept and the variables “ X ” which represents one of the attributes that models a instance. For every week we need to compute an “ Y ” and place it on the plot; further analysis of the plot will lead us to the trend analysis.

Machine Learning Approach: This approach aims to estimate the grade via classification algorithms and more exactly decision trees. We will have 10 classes (one for every grade) and on the chart’s axes we will have the estimated grades and the time stamp.

The main difference between these two approaches is the grade which will have certain values for the classification approach and some continuous values for regression. Comparing the results from these two approaches will lead us to make the best choice.

Alerts. Offering alerts during the semester regarding the student’s failure is another capability that needs to be taken into consideration. First there are some tests that need to be performed in order to see if the student’s activity on the forum will predict its success/failure or even his final grade.

Approach. The approach for computing alerts when students are at risk there is a need for some pattern matching techniques. We plan to cover this approach with classification algorithms that can be employed using two or three classes. Based on the previous gathered data we can predict if a student is “at_risk” or “not_at_risk” but we also need to perform some tests having three classes like “at_risk”, “not_at_risk” and “avg_risk” where the last class can be used for students as a first warning level and then alert “at_risk” if the situation gets really bad.

6 Conclusions and Future Work

In this paper we presented the design and the integration approach for a forum that implements several smart functionalities that are based on underlying machine learning obtained models. Adding smart functionalities that address the goals described in the paper can improve the student’s experience of the learners that perform their activities in on-line educational environments. From the many possibilities of implementing the

prototype of smart forum we have chosen core algorithms that fit presented problems in terms of available input and needed task. Currently, the challenge is to choose proper machine learning algorithms and integrate with the data provided by Tesys e-Learning platform such that tasks within the smart forum may get practical and interpretable output.

The prototype version of the smart forum trains linear regression model in a predictive context for the output variable represented by the final grade. The segmented analysis provides finer grained trends for fixed or variable length timeframes and this approach is a more realistic one for the learning period of a student. Further analysis of trends in observed data is necessary as a mean to get an insight in the learning activity patterns of the students.

The second smart capability modelled as a classification problem and the linear regression model is used as a decision boundary between students “at_risk” or “not_at_risk”. Further validation of derived models is needed such that confidence values in obtained predictions and classification are obtained.

As future work, improving the logging mechanism from Tesys and smart form will provide better and comprehensive real world datasets and thus the ability to extract features for better describing the observations from the training dataset.

Another improvement regards implementing more functionality in smart forum for processing the activity traces with the task of creating better student models.

References

1. Chih-Hsiung, T., McIsaac, M.: The relationship of social presence and interaction in online classes. *Am. J. Distance Educ.* **16**(3), 131–150 (2002). doi:[10.1207/S15389286AJDE1603_2](https://doi.org/10.1207/S15389286AJDE1603_2)
2. Jennifer, C.: Richardson & tim newby, the role of students’ cognitive engagement in online learning. *Am. J. Distance Educ.* **20**(1), 23–37 (2006). doi:[10.1207/s15389286ajde2001_3](https://doi.org/10.1207/s15389286ajde2001_3)
3. Robinsona, C.C., Hullingera, H.: New benchmarks in higher education: student engagement in online learning. *J. Educ. Bus.* **84**(2), 101–109 (2008). doi:[10.3200/JOEB.84.2.101-109](https://doi.org/10.3200/JOEB.84.2.101-109)
4. Dringus, L.P., Ellis, T.: Using data mining as a strategy for assessing asynchronous discussion forums. *Comput. Educ.* **45**(1), 141–160 (2005). <http://dx.doi.org/10.1016/j.compedu.2004.05.003>. ISSN 0360-1315
5. Liu, W., Kidziński, L., Dillenbourg, P.: Semiautomatic annotation of MOOC forum posts. Chapter State-of-the-Art and Future Directions of Smart Learning Part of the series Lecture Notes in Educational Technology, pp 399–408. doi:[10.1007/978-981-287-868-7_48](https://doi.org/10.1007/978-981-287-868-7_48) Print ISBN 978-981-287-866-3 Online ISBN 978-981-287-868-7
6. Romero, C., López, M.-I., Luna, J.-M., Ventura, S.: Predicting students’ final performance from participation in on-line discussion forums. *Comput. Educ.* **68**, 458–472 (2013). <http://dx.doi.org/10.1016/j.compedu.2013.06.009>. ISSN 0360-1315
7. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: Extreme learning machine: theory and applications. *Neurocomputing* **70**(1–3), 489–501 (2006). <http://dx.doi.org/10.1016/j.neucom.2005.12.126>. ISSN 0925-2312
8. Purdila, V., Pentiu, S.-G.: Fast decision tree algorithm. *Adv. Electr. Comput. Eng.* **14**(1), 65–68 (2014). doi:[10.4316/AECE.2014.01010](https://doi.org/10.4316/AECE.2014.01010)

9. Chao-Ying, J.P., Leea, K.L., Ingersolla, G.M.: An introduction to logistic regression analysis and reporting. *J. Educ. Res.* **96**(1), 3–14 (2002)
10. Bradley, A.P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.* **30**(7), 1145–1159 (1997). [http://dx.doi.org/10.1016/S0031-3203\(96\)00142-2](http://dx.doi.org/10.1016/S0031-3203(96)00142-2). ISSN 0031-3203
11. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 881–892 (2002). doi:[10.1109/TPAMI.2002.1017616](https://doi.org/10.1109/TPAMI.2002.1017616)