# Application of Learning Analytics techniques on blended learning environments for university students

Sheila Lucero Sánchez López[1], Rebeca P. Díaz Redondo [2] and Ana Fernández Vilas[3]

School of Telecommunications Engineering. I&C Lab. AtlantTIC Research Center.
University of Vigo. 36310 Vigo. Spain.
[1]sheila.lucero@det.uvigo.es
[2]rebeca@det.uvigo.es
[3]avilas@det.uvigo.es

**Abstract.** The educational process is constantly changing. On the one hand, traditional educational methods have been modified and, on the other hand, the model of educational transmission has also changed. According to different authors, technological resources, specifically the eLearning platforms, and social interaction are responsible for these changes. Based on these approaches, this article applies Learning Analytics techniques with the aim of analyzing social interaction in blended-learning environments. For this, an exploratory analysis will be carried out in the messages published in the forums with the objective of qualitatively analyzing the students' interaction with the educational platform.

**Keywords:** Learning Analytics, e-learning, Forums, Learning Acquisition, Educational Data Mining, social interaction.



## 1 Introduction

Technology is innovating almost all areas of our life and education, is not the exception. The educational process has undergone several changes as a result of the implementation of technological resources inside and outside the classroom. In fact, numerous institutions add a fundamental role to technology in education. Among them, the European Parliament emphasizes that digital learning has the potential to help the European Union to respond to the challenges of the knowledge society, improve the quality of learning, solve special needs and allow a more effective learning and training [1]. For its part, the Department of Education of the United States of America argues that computers are "the new basic" of education and the Internet is the "blackboard" of the future [2]. The United Nations Organization for Education, Science and Culture (UNESCO) emphasizes the potential of Information and Communication Technologies (ICTs) to disseminate and improve teaching and learning in a wide variety of contexts [3].

One of the contributions with the greatest impact that Information and Communication Technologies (ICTs) have made to the education sector is the implementation of e-Learning platforms. These platforms are defined by [4], as web applications that integrate a set of tools for the online teaching-learning process, with the aim of allowing the creation and management of teaching and learning spaces on the Internet, where teachers and the students can interact during their training process. The boom of these platforms has been so great that they are currently used in different educational levels and in different parts of the world. In fact, according to the combination of technological resources with the degree of presence that the student has while learning, we can find three widely accepted teaching modalities: traditional modality, e-learning and blended-learning.

The traditional modality is when the student receives the knowledge in its entirety inside the classroom, in the same space-time as the teacher without the presence of technological resources provided by the (ICTs). The e-learning modality is also called online education modality. In this, the teaching is taught entirely remotely over the Internet, without the need for students to interact with the platform at the same time or in the same geographical location as the teacher. This allows the student to advance at his own pace, making his learning process more flexible and favoring his autonomy [5]. The modality blended-learning or mixed education, arises when the lessons in the classroom complement each other with the educational platform. Fusing this way, two pedagogical approaches that combine the effectiveness and opportunities of socialization of the class with the technological improvements of online learning [6].

These platforms have the capacity to store an innumerable amount of data from the interaction of users (students and teachers) with them and through them. Despite the success and acceptance that these platforms are having, these do not have per se any tool to facilitate the interpretation or analysis of this data. However, these data have aroused the interest of many researchers, thus emerging two specialized fields of study: Learning Analytics (LA) and Educational Data Mining (EDM).

According to the First International Conference on Learning and Knowledge Analysis (LAK 2011) [7], LA "is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.”. On the other hand, EDM is defined as "the development, research and application of computerized methods to detect patterns in large collections of educational data that would otherwise be difficult or impossible to analyze due to the huge volume of existing data" [8].

Both areas share different challenges. However, this work has been developed under the proposal of LA, based on the approach proposed by [9] "Learning Analytics refers to the interpretation of a wide range of data generated and collected on behalf of the student to evaluate their academic progress, predict their future performance, and locate potential problems. The data is collected from explicit student actions, such as performing evaluable exercises or tests, and from unspoken actions, including social interactions, extracurricular activities, publications in a discussion forum and other activities not directly evaluated as part of the educational progress of the student. The goal of Learning Analytics is to support teachers and schools in the process of adapting their learning opportunities to the level of need and ability of their students in real time (or with a fairly tight margin)".

It is also necessary to emphasize that not only the teaching modality has changed, also the model of transmission of knowledge has been transformed. According to

[10], two types can be distinguished: on the one hand there is the model where the teacher plays the central role as wise on stage, called "sage on the stage" and, on the other hand, there is the model where the teacher and the student jointly create the learning environment, called "guide on the side". In this case the role of the teacher is to be a side guide.

Several authors support the idea that interpersonal interaction provide the advantages of the second model. This type of interaction is generated when students react to the content and share concerns, they teach each other learning in a tangible way when they express with words (through publications on the platform) their own understanding and assumptions, which allows them to appropriate new skills and ideas, at all times being focused and deepened by the lateral guide, without it hindering the development and learning experience of the students [11]. Likewise, in the literature we can find numerous studies that prove that a greater participation in terms of quality and quantity can increase learning. Otherwise, by controlling the design elements of technological resources and the execution of the course, participation and learning can be increased [12], [13], [14].

Starting from the premises that e-Learning platforms and interpersonal interaction (or also called, social interaction) are of great importance in the new changes that are arising in the educational process. In this paper we will study the application of LA techniques in blended-learning environments focused on university studies. To this end, a methodology will be presented that allows qualitatively analyzing the interpersonal interaction of students in the e-Learning platform. Our approach tries to take advantage of the information exchange in the online forums to discover new knowledge about the students' way of learning or behave. In this paper, our work done in [15] is broadened by analyzing social interaction from a qualitative perspective, since in the work cited, social interaction is only approached from a quantitative perspective.

To do this, the data extracted from the official platform of the University of Vigo belonging to a programming course along three different academic years of Telecommunications Engineering will be used.

This document is structured as follows. The following section (section 2) provides a description of the data set and the methodology. Subsequently, in section 3 we will analyze the students qualitatively, analyzing their messages and publications in the forums of the e-Learning platform. Finally, in section 4 the results will be analyzed.

## 2 Description of Dataset and Methodology

To perform our experiments, we use data from a course related to the programming skills of the third year course of a Bachelor Degree on Telecommunication Engineering. This is a blended course of fourteen weeks between September and January. The dataset was gathered from the official e-Learning platform, Moodle-based, of the university where the subject was taught.

The assessment mechanism of this course is based on three mandatory assignments distributed along the course (from the fourth to the last week) as **Fig. 1** shows. The forum is accessible to all students and used to debate about different aspects related to the course (content or administrative issues), answer questions, solve doubts, etc. However, it should be noticed that it does not represent any mandatory activity. The required assignments are divided in two types:

- Laboratory: To determine if the student has acquired all the knowledge and skills corresponding to the laboratory practices (3 practices).
- Applied: To determine if the student knows how to apply the knowledge of the course to solve problems (2 exams).

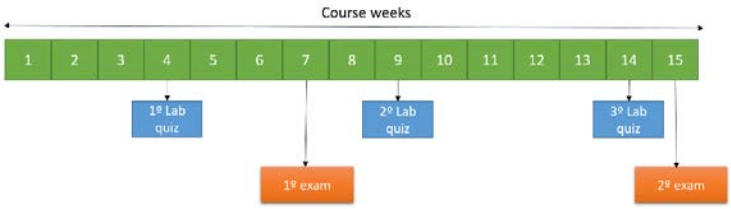


**Fig. 1.** Temporal representation of the course

The Moodle platform stores in its database not only all the information related with the courses (course contents, personal data of students and professors, students' grades, etc.), but also all the information about the students' interaction with the platform. In fact, Moodle distinguishes between different types of interactions, which are classified in ten different modules (Assignment, Blog, Choice, Course, Forum, Notes, Resource, Upload, User, and Quiz) as **Table 1** shows.

**Table 1.** Detail of the information contained in each module

| Module | Information |
|---|---|
| Assignment | Files, notes, deliveries of work requested by the teacher. |
| Blog | Advertisements |
| Choice | Selection of information such as dates, places, excursions attendance lists, etc. |
| Course | Assignment of teachers and students by subject. |
| Forum | Everything related to forums (questions, news, discussions) that create teachers and students. |
| Notes | Notes – additional information |
| Resource | Educational resources, notes, slides, presentations. |
| Quiz | Assessments, quizzes, tests, etc... |
| Upload | Updates/changes in resources |
| User | All personal user information |

For our analysis, we gathered data (73,849 interactions) from three academic years. We analyze data of 435 students organized from 2014 to 2017, as shown in **Table 2**.

**Table 2.** Initial Data

| Academic Year | Students | Interactions |
|---|---|---|
| *2014/2015* | *132* | 25.333 |
| *2015/2016* | *168* | 28.410 |
| *2016/2017* | *166* | 20.106 |
| *Total* | *466* | 73.849 |

As mentioned above, we will analyze two types of interaction: interaction with content and social interaction. Initially, we have divided the events into two groups: (i) actions related with some contents or class notes and (ii) actions related to interpersonal activities. The objective is to find the group of activities that have a higher relation with one of both interaction types.

We consider the following classification: the modules *Assignment, Course, Notes, Resource, Upload, and Quiz* are related to content. *Blog, Choice and Forum* are related to interpersonal interaction. *Use* is outside of both classifications, because it does not provide information related to this.

Modules *Blog*, *Choice* and *Forum* are considered as interpersonal participation because the students can show their own ideas in module Blog. On module *Choice* they can choose and propose surveys and discussions, and finally, in module Forum, students can participate in a more active way.

Our methodology is divided into three stages. The qualitative analysis begins with data preprocessing, continues with the classification of the messages in three groups and ends with an exploratory analysis of the content of the messages.

## 3 Qualitative analysis

As mentioned earlier, this analysis is divided into three stages: the first one corresponds to the data preprocessing; the second one is a classification of the messages in three categories (content, code and other); and finally the exploratory analysis of the content of these categories.

### 3.1 Data Preprocessing

It is necessary to prepare and transform the gathered information to classify the messages. Initially, a corpus of specific content has been created for the experiment, extracting the main words (topic words) from 12 pdf files: teaching material (4 pdf files), educational resources (3 pdf files), notes (3 pdf files), slides (2 power point files), three practices (3 files) and references in the presentations of the class (2 pdf files). All information is available to any student enrolled in this subject and with access to the official e-Learning platform of the university. From these documents, we have obtained a total of 15,704 words. This set is latter reduced to a corpus of 587 words, after removing stopwords, carrying out a lemmatization and extracting the topic words. This corpus will be called "*Content Corpus*".

As it was previously commented, the content of the subject is related to computing, especially to two programming technologies: Java and HTML. For this reason, the use of programming codes is very frequent. Therefore, we use a second corpus that will be called "*Code Corpus*", created by RANKS NL that contains the top words of all programming languages.

RANKS NL [16] is a keyword analyzer tool for URLs, websites, texts and documents to improve search engine optimization and other purposes. It has available a collection of stopwords' lists in more than 40 languages and the list of reserved words of Perl, Mysql, Javascript, C, C++ and HTML. In the same way, the stopwords raised by RANKS NL will be removed of all the messages.

As a summary, the two corpus that we will use to classify the messages are:

- *Content Corpus*: created by the extraction of the main words (topic words) of the teaching material, educational resources, notes, slides, practices and references available in the e-Learning platform. It is composed of 587 words.
- *Code Corpus*: this corpus will serve to classify messages that contain programming codes and it is based on the corpus armed by RANKS NL. It is composed of 2,500 words.

### **3.2** Classification.

The next step will be classifying the exchanged messages in the forums. Naive Bayes classifier and the two corpus (Code and Content) will be used for this task. By Bayes theorem, the probability can be defined as:

$$p(C|w_1 = y, w_2 = n, w_i = \cdots) = \frac{p(C)p(w_1 = y, w_2 = n, w_i = \cdots)}{p(w_1 = y, w_2 = n, w_i = \cdots)} \quad (1)$$

Where $p(C)$ is the probability of belonging to the specific corpus (Code or Content); $w_i$ is the identifier of word; $y$ represents if it belongs to the corpus; and $n$ if it does not belong to it.

Our interest is the relative probabilities of the messages being a code message or content message. In other words, the exact value of the probability is not important because the classification will be assigned according to the highest percentage of belonging to any of the corpus. Therefore, we can factor out any terms that are constant, namely the denominator of the above equation is a constant because it depends on the total number of messages (from both types – content and code -). For this reason, the numerator of equation (1) can then be written as:

$$p(C)p(|w_1 = y, w_2 = n, w_i = \cdots) = p(C)p(w_1 = y|C)p(w_2 = n|C)p(w_i = \cdots) \quad (2)$$

Each message will follow the same process. First, divide each message word by word. Second, stopwords are removed and lemmatization is executed. Third, the messages are classified using the Naïve Bayes classifier if the message has 33% membership in the code or content corpus. This percentage is recommended by RANKS NL, creator of the code corpus. This percentage is recommended when using this classifier for detecting spam in emails. Finally, we obtain three classifications.

1. *Code messages*: messages that 33% of its content belongs to the code corpus.
2. *Content messages*: respecting the same percentage, these are messages that 33% of its content belongs to the message corpus.
3. *Other messages*: the rest of messages that do not belong to any of the two previous classifications.

The procedure considers that the same message can contain words that belong to the two corpus. As shown in **Fig. 2**, first, it calculates the probability of each word of belonging to the code corpus and get the value of the probability. Then, it calculates the belonging to the content corpus, word by word, until exceeding the percentage of belonging to the code corpus or finishing by analyzing all the words of the message. Finally, the classification with the highest percentage is assigned, as long as it exceeds 33%.

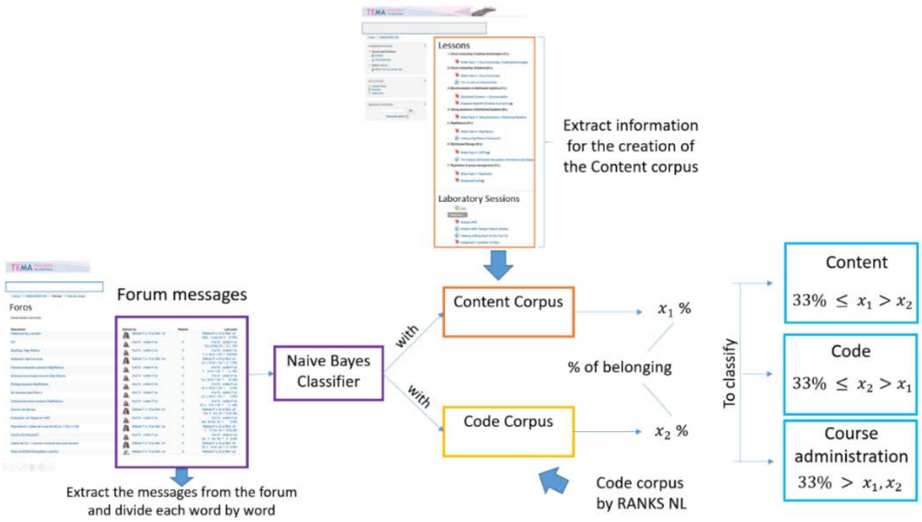


**Fig. 2. Data preprocessing & classification**

To check the classification, an expert in the programming area reviewed each message to classify them manually in the three identified groups, obtaining that only 7.2% of messages correspond to another category different from the one assigned by the Naïve Bayes classifier. With this information, we have calculated other interesting measures of accuracy like precision, recall and F score, summarized in **Table 3**.

**Table 3.** Test's accuracy

| | Precision | Recall | F score |
|---|---|---|---|
| Code | 88.5% | 88.5% | 3.54 |
| Content | 93.7% | 92.8% | 3.77 |
| Other | 92.7% | 91.9% | 3.68 |

As previously mentioned, we have decided to use the threshold of 33% to decide if a message would belong to the *Content* category, keeping the recommendation by RANKS. This value supported good result. However, we decided to perform some tests changing this threshold. After doing an exhaustive work, we detected that our results were optimized using a higher threshold of 52%: our error decreased to 5.7% and the total recall increased from 92.8% to 94.3%. This encouraged us to check what happened if the threshold of the *Code* category was also altered. After the same analysis, we optimized our results increasing this threshold to 35%.

Finally, **Table 4** summarizes the distribution of messages per academic year: content messages are clearly the most frequently exchanges and code messages the less frequent. Since the percentages are quite similar in the three academic years, we have decided to focus the analysis in a single data set formed by the information of the three academic years (2014-15, 2015-16, and 2016-17).

**Table 4.** Distribution by classification

| | **2014-15** | | **2015-16** | | **2016-17** | | **Total** | |
|---|---|---|---|---|---|---|---|---|
| Code | 18 | 13% | 18 | 8% | 25 | 11% | 61 | 13% |
| Content | 68 | 49% | 102 | 48% | 113 | 49% | 283 | 49% |
| Other | 52 | 38% | 92 | 43% | 92 | 40% | 236 | 38% |

### 3.3 Analysis of messages

It is important to emphasize that we will analyze the content of the three classifications by performing an exploratory analysis. We will search the most frequently used words in the previously classified messages. This will allow us to know which are the top words and if there is a relationship between the classifications. Moreover, the next step is to plot networks of these co-occurring words, so these relationships are clearly displayed, as **Fig. 3** shows.

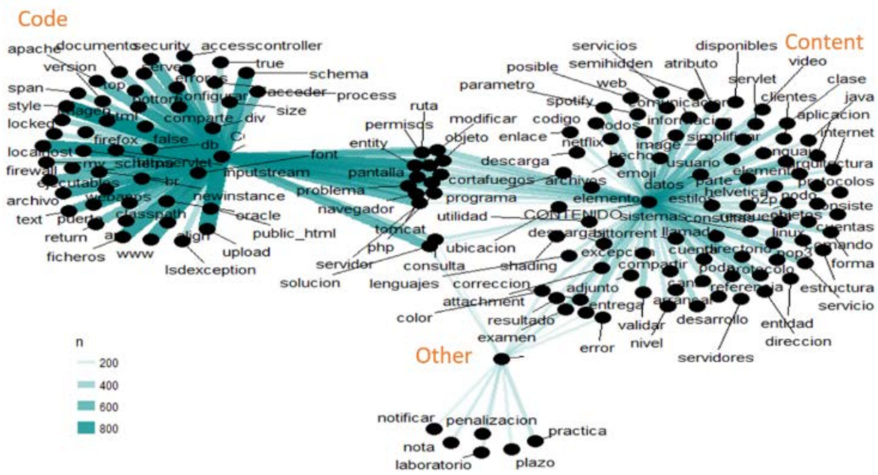


**Fig. 3.** Co-occurring words

Considering that *n* represents the co-ocurrence of the words in **Fig. 3**, it was found that several words are indistinctly used in the *Code* category and in the *Content* category, such as *entity*, *permissions*, *firewalls*, *browser*, or *route*. Besides, there are words that appear in the *Other* category and in the *Content* category, such as *exam*, *results*, *deliver*, *attachment* or *correction*. Finally, there are words that appear in the three corpus: *php*, *tomcat, query, server*, etc. We can see that the words of the *Other* category are referring to the course administration, therefore this classification will be named as such.

Additionally, **Fig. 4** shows the distribution of the messages of each category along the academic term, showing the temporal evolution of the exchanges messages.

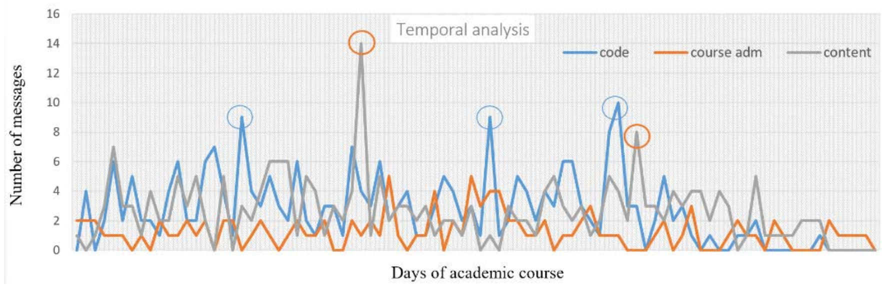


**Fig. 4.** Temporal analysis

As **Fig. 4** shows, we have 3 peaks (blue circles) of code messages, the first corresponds to the delivery of the first practice, the second to the revision of the second practice and the third to the delivery of the third practice. We also have 2 peaks (orange circles) in content messages corresponding to the session prior to the exams.

Regarding the messages of the course administration, there is no pattern depending on the academic organization of the course.

To finish our exploratory analysis, it is important to know who initiates the posts: a teacher or a student. It was obtained that 71% of code conversations, 63% of content conversations and 55% of information conversations are started by students in each group.

Knowing that the student starts mainly the posts, the next point would be to know which messages respond most frequently, those sent by the teacher or by the other students. For this reason, we have calculated the percentage of the student's response to conversations initiated by one of their classmates, knowing that 89% of the messages sent by another student is answered. Only 11% is initially answered by the teacher.

Additionally, we have analyzed the themes and topics of the exchanged messages with a program called DepPattern [17]. It is a linguistic package providing a grammar compiler, PoS taggers, and dependency based parsers for several languages including Spanish and Galician. This is a very important feature, because the messages in the forum are written in two languages (Spanish and Galician). **Fig. 5** shows an example of the results obtained by the software. The list of infinitive verbs, punctuation marks and nouns of the messages were obtained by DepPattern.

```
1	En	en	PRP	_	_	_
2	que	que	PRO	_	_	_
3	consiste	consistir	VERB	0		ROOT	ROOT:0
4	la	el	DT	5	SpecL	SpecL:5
5	estrategia	estrategia	NOUN	3		DobjR	DobjR:3
6	Tit-for-Tat	Tit-for-Tat	NOUN	5		AdjnR	AdjnR:5
7	(	(	Fpa	8	PunctL	PunctL:8
8	TFT	TFT	NOUN	5	AdjnR	AdjnR:5
9	)	)	Fpt	8	PunctR	PunctR:8
10	de	de	PRP	3	CircR	CircR:3
11	Bittorrent	Bittorrent	NOUN	10		Term	Term:10
12	y	y	CONJ	_	_	_
13	de	de	PRP	_	_	_
14	que	que	CONJ	_	_	_
15	ambito	ambito	NOUN	19	SubjL	SubjL:19
16	de	de	PRP	15	CprepR	CprepR:15
17	la	el	DT	18	SpecL	SpecL:18
18	Teoria@de@Juegos	Teoria@de@Juegos	NOUN	16
19	proviene	provenir	VERB	0		ROOT	ROOT:0
20	?	?	SENT	_	_	_
```

**Fig. 5.** Results by DepPattern

Therefore, when interpreting the results we have obtained the following topics:

1. Questions mainly about delivery schedules and tests' dates and a reminder of instructions.
2. Recommendations of alternative content, specific questions and questions about the relevance of the exercises.
3. Ask about exam dates and deliveries, make assumptions and the questions are more general.

4. Examples' requests, references to class slides, web links, feedback and answers to questions.
5. Ask giving answer options, ask several questions in the same message, give examples and alternatives, attach extra resources, the messages are longer.

## 4 Discussion and conclusions

As a brief summary, two corpus were used in the analysis. The first (content corpus) was created specifically with the academic content of the course, and the second (code corpus) was taken from the one created by RANKS NL. Applying the Naïve Bayes classifier and these two corpus, we have obtained three classes or categories: *code*, *content* and *course administration*. The first two are composed of messages with a high percentage of words related to code and course content, respectively. Those messages which are not classified in these two categories go directly to the third one, whose name was decided after checking that all the messages included reference to course administration (questions and/or information about the exams, revision dates, etc.). We chose this classifier because its structure is fixed and does not depend on the data, it follows a generative or discriminative criterion. Like the other Bayesian classifiers, the obtaining of the parameters is based on the maximum likelihood or a posteriori maximum estimations [18]. In addition, this classifier has shown good results in the classification of texts [19].

The analysis of the messages can give feedback from the students to the teachers, remarking those topics that are considered more interesting or those in which doubts usually arise. Having a direct feedback from the student is important to be able to take more concrete actions and improve the academic course, for example, reviewing certain concepts, solving concerns, repeating dates or instructions and, consequently, supporting the student in his acquisition of knowledge from a less formal environment (forums) than the classroom. Forums can encourage shy or absent students to interact with other students and, in the same way, they can encourage the more participatory students continue to reinforce their interaction. The proposed methodology can bring improvements inside and outside the classroom. It marks an important guide in the educational process, by facilitating the content analysis of the messages in the forum, identifying the main topics of discussion, the topics that more generate doubts, the answers and the recommendations that are given between students. This allows to analyze valuable data of student behavior, with which learning models and learning analysis could be applied to improve the quality of education and the participation of students.

As a future line, on the one hand, it would be interesting to integrate the classification of students to analyze the content of the messages for each profile proposed by the classification and, on the other hand, to use these messages to try to profile the student who sent them. This methodology could be an initial step to integrate a content recommender system into the eLearning platform.

## Acknowledgments

This work is funded by the European Regional Development Fund (ERDF) and the Galician Regional Government under the agreement of funding the Atlantic Research Center for Information and Communication Technologies (AtlantTIC); the Spanish Ministry of Economy and Competitiveness under the National Science Program (TEC2014-54335-C4-3-R, TEC2017-84197-C4-2-R).

## References

1. Parlamento Europeo, "Decisión nº 2318/2003/CE del Parlamento Europeo y del Consejo de 5 de diciembre de 2003 por la que se adopta un programa plurianual (2004-2006) para la integración efectiva de las tecnologías de la información y la comunicación (TIC) en eLearning," *Diario Oficial de la Unión Europea,* Vols. L-345, no. 2318, pp. 9-16, 2003.
2. U.S. Department of Education, "Getting America's students ready for the 21st Century: Meeting the technology literacy challenge.," *National Education Technology Plan,* p. 3, 1996.
3. Guttman, C., "Education in and for the Information Society," United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, 2003.
4. Fernández-Pampillón Cesteros, A. , "Las plataformas e-learning para la enseñanza y el aprendizaje universitario en Internet.," in *Las plataformas de aprendizaje. Del mito a la realidad.*, Madrid, Biblioteca Nueva, 2009, pp. pp. 45-73.
5. Cabero, J., "Bases pedagógicas del e-learning," *Revista de Universidad y Sociedad del Conocimiento,* vol. 3, no. 1, 2006.
6. Norberg, A., Dziuban C. D., and Moskal, P. D., "A time-based blended learning model," *On the Horizon,* vol. 19, no. 3, pp. 2017-216, 2011.
7. LAK, "Call for Papers of the 1st International Conference on Learning Analytics & Knowledge (LAK 2011)" in *LAK*, Banff, Alberta, 2011.
8. Romero, C. and Ventura, S., "Data mining in education," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, no. 3, pp. 12-27, Enero, 2013.
9. Johnson, L. , Conery, L. and Krueger, K., The NMC Horizon Project: 2011 K-12 Edition, Austin, Texas: The New Media Consortium, 2011, pp. 26 - 29.
10. Bento, R., Brownstein, B. and C. &.Schuster, Z., "Fostering Online Student Participation," *Journal of College Teaching and Learning,* vol. 2, no. 7, pp. 31-37, 2005.
11. Collison, G., Tinker, R., Elbaum, B. and Haavind, S., Facilitating Online Learning: Effective Strategies for Moderators, Madison: Atwood Publishing, 2000.
12. Harasim, L. M., Hiltz, S. R., Teles, L. and Turoff, M., Learning Networks: A Field Guide to Teaching and Learning On-Line, Cambridge: The MIT Press, 1995.
13. Kemery, E., "Developing On-line Collaboration," in *Web-Based Learning and Teaching Technologies: Opportunities and Challenges*, Baltimore, Idea Group Publishing, 2000, pp. 227-245.
14. Adrianto, D., Yesmaya, V. and Chand, A., "Increasing Learning Frequency through Education Based Game," *Journal of Computer Science,* vol. 11, no. 3, pp. 567-572, 2015.
15. Sánchez López, S. L., Díaz Redondo, R. P. y Fernández Vilas, A. «Predicting students' grade based on students behavior,» *International Journal of Engineering Education (IJEE).*, vol. 34, nº 3, p. 940–952, 2018..

16. Doyle, D., "Ranks.nl," Ranks, [Online]. Available: https://www.ranks.nl/about. [Accessed 11 2017].
17. Gamallo Otero, P. and Gonzalez, I., "DepPattern: a Multilingual Dependency Parser," in *The 10th International Conference on the Computational Processing of Portuguese*, Coimbra, Portugal, 2012.
18. Santafé, G., Lozano, A., Larrañaga, P., «Aprendizaje discriminativo de clasificadores Bayesianos,» *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial,* nº 29, pp. 39-47, 2006.
19. Palazuelos, C., García-Saiz D. y Zorrila, M. «Social Network Analysis and Data Mining: An Application to the E-learning Context,» *Proceedings of the 5th International Conference on Computational Collective Intelligence,* pp. 651-660, 2013.