

SENTIMENT ANALYSIS OF TEACHERS USING SOCIAL INFORMATION IN EDUCATIONAL PLATFORM ENVIRONMENTS

NIKOLAOS SPATIOTIS

*Department of Electrical and Computer Engineering, University of the Peloponnese,
Patras, Greece,
spatiotis@gmail.com*

ISIDOROS PERIKOS

*Department of Computer Engineering & Informatics, University of Patras
Patras, Greece
perikos@ceid.upatras.gr*

IOSIF MPORAS

*School of Engineering and Computer Science, University of Hertfordshire
United Kingdom
i.mporas@herts.ac.uk*

MICHAEL PARASKEVAS

*Department of Electrical and Computer Engineering, University of the Peloponnese
and
Computer Technology Institute and Press “Diophantus”
Patras, Greece
mparask@cti.gr*

Received (Day Month Year)

Revised (Day Month Year)

Accepted (Day Month Year)

Learners' opinions constitute an important source of information that can be useful to teachers and educational instructors in order to improve learning procedures and training activities. By analyzing learners' actions and extracting data related to their learning behavior, educators can specify proper learning approaches to stimulate learners' interest and contribute to the constructive monitoring of the learning progress during the course or to improve future courses. Learners-generated content and their feedback and comments can provide indicative information about the educational procedures they attended and the training activities that they participated in. Educational systems must possess mechanisms to analyze learners' comments and automatically specify their opinions and attitude towards the courses and the learning activities that are offered. This paper describes a Greek language sentiment analysis system that analyzes texts written in Greek language and generates feature vectors which together with classification algorithms give us the opportunity to classify Greek texts based on the personal opinion and the degree of satisfaction expressed. The sentiment analysis module has been integrated into the hybrid educational systems of the Greek school network that offers life-long learning courses. The module offers a wide range of possibilities to lecturers, policymakers and educational institutes that participate in the training procedure and offer life-long learning courses, to understand how their learners perceive learning activities and specify which aspects of the learning activities they liked and which did not. The experimental study shows quite interesting results regarding the performance of the sentiment analysis methodology and the specification of the users'

opinions and satisfaction. The feature analysis demonstrates interesting findings regarding the characteristics that provide indicative information for opinion analysis and embeddings combined with deep learning approaches yield satisfactory results.

Keywords: Text Mining; Machine Learning; Opinion Mining; Data Mining; Feature Extraction; Feature Selection.

1. Introduction

In recent years, there has been rapid technology development and global web proliferation, which has led to significant changes in how to extract and exploit data to improve the purpose that these data each time serve. Sentiment analysis is becoming an emerging topic mainly due to the vast amount of people's opinions and user-generated comments on the web [21]. The vast amount of textual big data of the user-generated content necessitates accurate methods to analyze and specify users' opinions and attitudes. In order to identify and extract opinion and subjective information in text, natural language processing methods, text analytics and machine learning approaches are widely used [22]. An important stage in sentiment analysis concerns the extraction and utilization of appropriate and meaningful features that will be indicative and assistive in the classification process. In general text mining is the process of examining and converting large amounts of unstructured text data into structured data for further analysis like visualization and model building. In this context, Sentiment Analysis, which is also known as opinion mining, refers to a broad field of Natural Language Processing (NLP), Computational Linguistics and Mining Text. The main methodological characteristic of the sentiment analysis concerns the Sentiment Classification of textual data. Sentiment Classification is characterized as the technique of identification and emotional grouping of proposals which relates to the Opinion Mining from a text and is divided into individual fields in emotional orientation and in emotional classification. Regarding the first, it recognizes and emotionally classifies a proposal on the polarity (positive, negative and neutral) and the degree of intensity (low, medium and high). The second category which concerns the emotional classification aims to examine and to categorize textual data and user-generated content concerning the objectivity or subjectivity that it conveys. The use of this taxonomy has been very popular for the last ten years. Pang & Lee [1] [6] have done significant work on sentiment analysis, sentiment classification, sentiment mining and opinion extraction. In one of them, they proposed a new machine-learning method to determine the sense of polarity. The document polarity classification, as it is called, is a major challenge in this area. Cambria et al. [3], [5] presented a concept-level sentiment analysis providing new approaches for opinion mining and sentiment analysis not only from a simple word-level analysis but also from unstructured textual data. Liu [2] focused on subjectivity sentiment classification, sentiment lexicon generation and opinion summarization.

In the field of education, where more and more educational institutions increasingly adopt web-based learning platforms such as learning management systems and other online applications, these changes are not only important but also imperative. The ability to extract, analyze, and categorize data across large databases makes these systems an important source of "hidden" knowledge using Educational Data Mining and Learning

Analytics techniques. The application of data mining in education is an emerging research field, known as Educational Data Mining (EDM) [11]. Educational Data Mining, like data mining techniques, provides modeling and analysis of information. This is a technology that integrates multidisciplinary techniques, information retrieval from the database using simple queries to use machine learning algorithms from the field of artificial intelligence. Among the most common techniques in educational data mining is the automatic grouping (clustering), the analysis of partnerships, and also the discovery of common patterns in sequences [11]. Educational Data Mining investigates the development of appropriate methods for investigating data from educational environments. These are data extracted from the textual content and the opinions of the students, the teachers and other people, they are directly related to the educational procedure and are analyzed with the aim to be classified according to the opinions that they convey. This application aims to provide a better understanding of pupils' behavior, to determine how pupils learn and to identify how the learning process could contribute to students' performance [12]. Each training area has its problems and needs. Through Educational Data Mining, it is possible to discover hidden knowledge, which is harmonized with the needs and resolves the problems of each educational environment. In particular, using Educational Data Mining tools, researchers are interested in understanding students' needs and methods to improve the students' experience, activities and performance. For example, with data mining, interesting activities can be proposed for students related to their interactions with the online learning tools, thus contributing to their high performance. The information extracted can be a source of information for parents about the progress of their child's learning [18][19].

Students' opinions constitute an important source of information that can be useful to teachers and educational instructors in order to improve learning procedures and training activities. Evaluation teachers can use data mining applications to determine how to organize and structure the curriculum, how to determine the best methods for providing lesson information and the tools they will use to attract their students for the best learning outcomes. By extracting students' data related to their learning behavior, teachers can find appropriate teaching methods to stimulate student interest and contribute to the constructive monitoring of the learning progress during the course or to improve future courses [19]. Students' generated content and their reviews and their comments provide indicative information about the lectures they attended and the training activities they participated in. In educational systems, opinion mining must possess mechanisms to collect learners' comments and automatically specify their opinions towards the courses and the learning activities that are offered to them.

In this article, we will utilize the power of text mining to do an in-depth analysis of learners' reviews on an e-learning platform. Learners' reviews are a great source of "Voice of the customer" and could offer tremendous insights into what customers like and dislike about the e-learning platform. Our research focuses on opinion analysis of the Greek language by creating a computer model of opinion processing and subjectivity that governs a text. For this research, we used data derived from questionnaires, collected by the Greek School Network. These data come from learners -who attended synchronous, life-long

courses via a distant learning service (teleconference) - expressing their views while answering open-ended questions in free text. This paper describes a Greek language sentiment analysis system that analyzes texts written in greek language and generates feature vectors which together with classification algorithms give us the opportunity to classify Greek texts based on the personal opinion and the degree of satisfaction expressed. We created an automatic sentiment analysis system that classifies opinions polarity into five classes, reporting quite satisfactory performance and accuracy percentages. More specifically, the sentiment analysis system classifies free-text user comments according to their polarity, into very positive, positive, neutral, negative and very negative opinions. The sentiment analysis module has been integrated into the hybrid educational systems of the Greek school network that offers life-long learning courses and the module aims to analyze learners' opinions and specify their attitude towards the learning activities and the courses that they attend. The module operates in real-time and gets textual reviews as input, analyzes the reviews and automatically determines the users' satisfaction and their attitude. The module offers a wide range of possibilities to lecturers, policymakers and educational institutes that participate in the training procedure and offer life-long learning courses, to understand how their learners perceive learning activities and specify which aspects of the learning activities they liked and which they did not. The experimental study shows quite interesting results regarding the performance of the sentiment analysis methodology and the specification of the users' opinions and satisfaction.

The remainder of the paper is organized as follows. Section 2 describes methods and approaches on sentiment analysis and opinion mining and presents related work. Section 3 presents the hybrid education platform formulated for distant learning, describes its characteristics and analyzes its functionality. Section 4 describes the sentiment analysis module that is integrated into the educational platform to analyze learners' textual comments and specify automatically and in real-time their opinions and their attitude towards the courses and the learning activities that they participate in. After that, Section 5 presents the implementation procedure of the sentiment analysis methodology. Specifically, it describes the methodology on the analysis of users' opinions, examines the impact of feature selection on the performance of machine learning methods, presents the results collected and discusses the main findings. Section 6 describes the results collected from the experimental evaluation and presents the main findings of the study while Section 7 presents the discussion. Finally, Section 8 concludes our work.

2. Background Topics and Related Work

2.1. *Description of sentiment analysis approaches*

The Sentiment and Opinion Analysis can be applied to various levels of analysis and classified according to the way we approach the text and the desired detail of the extracted information [7].

Document / Text - layer: At this level, we think the text expresses a single view of a product-service, which we are trying to define. The aim at this level is to determine the wider stance of the author, positive or negative, in a text that includes judgments and opinions (opinionated text). Not recommended for cases in which texts are describing different aspects of an object or comparing different objects.

Sentence - level: Assignment to sentence-level separates each document proposals whereas each sentence expresses a different emotion. The proposals are identified directly as positive or negative. Moreover, this approach is often associated with the classification of subjectivity (subjectivity classification).

Entity and Features Level: Assignment to Entity and feature level or aspect-based attempts a more detailed analysis compared to the previous two, and focuses on the same terms and not on an analysis of structural elements of the language (text, sentence, phrase). The opinion holders can give different opinions for different aspects of the same entity as this sentence [4]. For subjectivity elicitation, that is personal opinion - the view for multiple perspectives of a text and of polarity (the degree that is positive, negative or neutral positioning on the theme of a text), two main techniques in the literature that are followed: (a) Machine Learning and (b) Lexicon - based [23]. The first approach considers opinion mining as a text classification problem using machine learning algorithms for classification and syntactic and/or linguistic features. The second approach relies on opinion lexicons using the prior polarity of words or phrases.

2.1.1.1. *Sentiment Analysis Based on Lexicons*

One of the most common ways of research approaches in the sentiment analysis field is to create resource dictionaries composed of terms expressing an opinion, positive or negative (opinion words) which express emotion and to each is assigned a rating based on the designating percentage classification as positive, negative or neutral (sentiment words). We process these words as independent, ignoring the grammar and their syntactic role or even their position in the text. The expression attributed to this approach is "bag of words". Thus, the words are detected by a text analysis which are identical with each sentiment lexicon and the overall feeling of the text is determined by the sum of the scores of individual words and using thresholds (thresholds) in case of multi-level emotional classification. Such dictionaries are SentiWordNet, Linguistic Inquiry and Word Count, Multi-Perspective Question Answering Subjectivity Lexicon (MPQA), and General Inquirer. SentiWordNet is a lexical resource for opinion mining that assigns to each synset of WordNet three sentiment scores: positivity, negativity, and objectivity. It has a Web-based graphical user interface, and it is freely available for research purposes. The development of the resource is based on the quantitative analysis of the glosses associated to synsets and on the use of the resulting vector term representations for semi-supervised synset classification. Linguistic Inquiry and Word Count (LIWC) is a computer program for language analysis. It supports English, German, Spanish, Dutch, and Italian. This commercial word list lets you extract around 60 different word categories, including positive emotions, negative emotions, aggression, affective processes, anxiety, anger,

profanities, and so on. The MPQA (Multi-Perspective Question Answering) Subjectivity Lexicon is a list of subjectivity clues that is part of OpinionFinder which helps to determine text polarity.

Although the above method is simple and easy to understand and is classified as a popular choice because it is not necessary to make any training, it usually has not high success rates as it takes into account the interaction of words that sometimes changes the semantics of the opinion (negation, of intensity - Intensifiers, idioms, irony - sarcasm, the finite number of words in dictionaries, emoticons).

2.1.2. *Sentiment Analysis based on Machine Learning Methods*

By the term, Machine Learning is defined as a system capable of acquiring and integrating knowledge automatically. In opinion analysis and the classification of textual data, the techniques which are applied are mainly divided into the three following categories according to the degree of human intervention in the learning process [7]:

Supervised Machine Learning: In Supervised Machine Learning, the under training system takes as input the training data that are called the training instances of the data, which are already labeled with a feature vector in terms of their sentimental content, so that the classifier to be trained to recognize and to learn more representative differences between texts belonging to different categories.

Unsupervised Machine Learning: The system is powered only with inputs invited to discover possible hidden structures between them in order to classify the data into groups that show some similarity. The existence, the number and properties of groups are initially unknown to the system. No training set is required for the extraction of feature vectors. Instead, using pre-built dictionaries of emotion the various terms of the text are characterized and the overall polarity arises.

Semi-Supervised Machine Learning: The under training system is powered with little text data which are labeled as to their sentimental content and with enough data which are not labeled. Although the creation of a dictionary is a time-consuming process, this method is considered the most secure and is applied by many researchers because it allows the creation of dictionaries according to user requirements and enables the expansion of existing ones.

In our approach, we perform opinion analysis on text-level, where the whole user-generated comment is analyzed and classified in the proper category and we follow a supervised machine learning methodology where user-generated textual data were annotated by experts.

2.1.3. *Sentiment Analysis based on Deep Learning Methods*

Deep Learning utilizes algorithms with multiple layers/levels, usually in the form of Artificial Neural Networks. The learning method is focused on representation and abstraction on the features of the data. It was introduced around 1970, where concepts like “multi-layer perceptrons” and “neural networks with hidden layers” were first mentioned by researchers.

In the years following 2000, the deep learning industry began to rise in popularity alongside the rapid growth of the internet with the term deep (structured) learning gaining traction. A big variety of new algorithms and techniques emerged like CNNs and RNNs. Then after 2009 the first use of graphics processing units to train deep neural networks arose, as well as heavy use in image recognition.

Deep learning has emerged as a powerful machine learning technique that learns multiple layers of representations or features of the data. The applications are endless and some fields where deep learning achieves state-of-art performance are image and automatic speech recognition. Furthermore, it is very useful in natural language processing, sentiment analysis and emotion recognition. In recent years, the majority of competitions and challenges are won by Deep Learning models [50] [52] [53].

2.2. Related Work on Sentiment Analysis

The application of data mining in education is an emerging research field, known as Educational Data Mining (EDM) [25]. Educational Data Mining (EDM), like data mining techniques, is developing methods to analyze student's performance which can help in identifying the student fails or dropouts and allow the teacher to provide appropriate counseling.

Pandey and Pal [26] conducted data mining research on the performance of students based on the selection of 600 students from several colleges in India. Employing Bayes Classification on category, language and background qualifications, the impact of language on the presence of students in the classroom was examined. In another paper [27], authors applied the classification as a data mining technique to predict the performance of students considering gender, location of house, family income, medium of instruction along with previous semester grades, attendance. Surjeet Kumar Yadav, Brijesh Bharadwaj and Saurabh Pal [28] used tree decision classifiers, that are studied and appropriate experiments are conducted in order to find the best classifier in order to predict the probability of student's drop-out. He and Wu ([29] examines students' participation and learning behavior by using LVS (Live Video Streaming) of data mining and text mining techniques. They found that students use chat messages to communicate positive emotions, negative emotions, and expressions of social support. Also, there is no positive correlation between the number of chat messages and final grades.

Ahmad et al [30] conducted a study to predict student academic performance of the first-year bachelor students of computer science courses. The dataset lasted for 8 years collecting data about various aspects of students' records including demographic data, previous academic records and other family-related information. They used Decision Tree, Naïve Bayes and Rule-Based classifiers for classification and prediction of student performance. The experiments showed that the Rule-Based classifier was the best among the classifiers and its accuracy was found as 71.3%.

Ali [31] performed a study to predict the student's performance using data mining techniques. He collected the data by the students at the admission time. Data mining methods include classification and clustering applied to various students' data based on

behavioral, psychographic and demographic features. It helped him in describing the student's performance whether they are successful or unsuccessful depended on their GPA or percentage achieved during secondary school.

Yadav and Pal [32] conducted research using classification trees to build a prediction model for engineering student's performance using students' features about previous academic records, family background and demographics. The dataset was obtained from VBS Purvanchal University, Jaunpur (Uttar Pradesh) on the sampling method for the Institute of Engineering and Technology for the session 2010. Three decision tree algorithms (C4.5, ID3 and CART) were applied to predict student's performance in the final exam. In their study, the experiments showed that the C4.5 classifier was the best among other decision tree algorithms and its accuracy was 67.77%.

Bevinda et al [33] conducted a study on predicting drop out students' performance. They have used the attributes like Previous Semester Marks, Internal Grades to predict the student's final semester marks. They used various of classification algorithms like ID3, C4.5, CART, CHAID. Their result shows that CHAID algorithm predicted the performance of drop out students with the highest accuracy.

Amjad Abu Saa [34] research concluded that student's performance not only depends upon academics but also depends upon other personal, social and extra-curricular activities. He along with the Naïve Bayes algorithm used three decision tree algorithms for classification of data. Firstly he did a survey and collected students' data and then preprocessed and explored the data for data mining tasks. Secondly, the data mining algorithms were implemented on the data set to generate classification models for predicting student's performance.

Khobragade [35] proposed an approach where they have predicted the students' academic failure using a decision tree, Naive Bayes and using classifiers that are based on induction rule and decision tree. Data used for classifications involved social, academic and background information of the student. These data have been collected through surveys. A total of 11 features were used for prediction after applying the feature selection algorithm. Classifiers have then been evaluated based on accuracy. Naive Bayes provided the best accuracy of above 87%.

Osmanbegovic and Suljic in [36] proposed an approach to predict the grade in Business Information for first-year students based on demographic variables and scores earned at high school too. The score of the entrance exam, study material and average weekly hours devoted to studying have been found to have the maximum impact while several household member distances of residence and gender have been found to have the least impact. Naïve Bayes is found to be a better classifier than J48.

Kabakchieva [37] proposed an approach to predict the students' performance using students' personal, pre-university and university performance characteristics. The studies encompass the data of 10,330 students in the Bulgarian educational sector. The students are classified as: Excellent, Very Good, Good, Average or Bad. The decision tree classifier performed best having the highest overall accuracy, followed by the rule learner (JRip) and the k-NN classifier. However, all classifiers performed with an overall accuracy below

70%. The predictive accuracy for the Good and Very Good classes (which contained the most students) was between 60% and 75%. The experiments of predictions were evaluated using 10-fold cross-validation.

Bydzovska [38] developed models to predict students' university performance based on course characteristics and previous grades. Two different approaches were used. In the first approach, classification and regression were used to predict performance using academic-related data and data about student's social behavior. The findings were significant with a small number of students. In the second approach, collaborative filtering techniques were used to predict the student's performance based on the similarity of achievements. Classification algorithms, namely, support vector machine, decision tree, part, IBI, RF, Naive Bayes and rule-based classifier, were used, where support vector machine produced best predictions which were further improved by integrating social behavior data.

Xiao et al. [54] propose a neural network architecture that utilizes and efficiently encodes characters. In recent work, character-level processing shows benefits over classic word-level approaches. At first, the document is encoded into a character-based representation, then these features are used as input on a Recurrent Neural Network (RNN). The model named ConvRec utilizes both recurrent and convolutional layers, and achieves comparable performances with other approaches while using much less parameters. In particular, the convolutional layer comes before the LSTM layer and ReLU activation is used.

The proposed model is evaluated on eight large-scale classification datasets, two of which are sentiment analysis ones, the Yelp and Amazon review datasets. The comparison is performed against a model by Zhang et al. [55], which is most relevant to this paper. As previously mentioned, it performs around the same but with fewer parameters.

Poria et al. [56] propose a system for multimodal emotion recognition, where other types of data such as audio and video are also used alongside textual data. Sentiment analysis isn't limited to text, the content produced and uploaded on the internet nowadays is multimodal. The authors propose a method to extract features from visual and textual modalities using deep Convolutional Neural Networks (CNNs) to automatically extract sentiment information. Then, such features are fed into a multiple kernel learning classifier. RNN layers are also used, since standalone RNN/CNN performs worse than combining them both. Multiple kernel learning (MKL) is a feature selection method where features are organized into groups and each group has its kernel function.

Mesnil et al. [57] present an Ensemble of existing learning models for a sentiment classification task. They compare several machine and deep learning approaches on this problem and combine them to achieve a new state-of-the-art accuracy. The Ensemble method used is Weighted Voting with grid search for weight values and the base classifiers are Sentence Vector model by other researchers, Naïve Bayes SVM trigram and RNN-LM.

Tang et al. [58] attempt to incorporate user and product-level information into a deep neural network for a sentiment classification task. The main goal is to predict a user's sentiment about a product. By modeling user and product representations using a vector

space, we can capture important clues such as user preferences and trends regarding the products. The model is named User Product Neural Network (UPNN) and uses a convolution layer followed by pooling.

Rozental et al. [59] propose a system for WASSA 2018 [139] that uses an Ensemble of deep learning techniques. The ensemble system consists of language models together with BiLSTM Neural Networks as base classifiers and also utilizes an Amobee Sentiment Classifier (ASC). The LSTM-based networks contain a CNN attention mechanism.

3. The Educational Platform

In this section, we present the basic concept of the Hybrid Educational Platform (HEP), we present its characteristics and analyze its functionality. The Hybrid Educational Platform was designed and implemented by the Greek School Network (GSN) in order to establish a blended learning process. Currently, the Hybrid Educational Platform is administrated by the group of universities and research institutes that cooperate and maintain the Greek School Network [20]. Production service started in June 2013 at '<http://training.sch.gr>'. The system offers both synchronous and asynchronous learning procedures. The synchronous procedures are offered via the synchronous eLearning environment developed and integrated into the system for hosting Tele-education courses. These training sessions are based on Big BlueButton (BBB) open-source software, thus reducing implementation costs. The tele-education environment integrated transparently with front-end and back-end modules and the participants need only one-click for the transition into the synchronous education module. Also, seminars and thematic workshops, which offer practical training and hands-on experience to the participants, are organized and offered via the HEP system.

A typical use case considering the formal utilization of HEP is as follows: computer engineering and computer science educators can gain access to HEP using a valid account in Greek School Network, thus allowing them to register in forthcoming lectures and faceto-face seminars, edit profile data and extract personalized statistical data (i.e., total hours of successfully attended e-lectures), etc. Additionally, educators can have access to the available training resources during the e-lecture or in a future time in order to achieve in a more coherent way the desired learning objectives. Furthermore, the evaluation of the synchronous e-learning courses is performed through an electronic questionnaire survey (EQS) that is provided to the participants of each session. Finally, for each computer science educator a certificate of participation, that officially represents proof of training involvement, is provided upon successful completion of ten hours of e-training. As is depicted in Figure 1 and considering the initial design workflow, HEP comprises the modules discussed below.

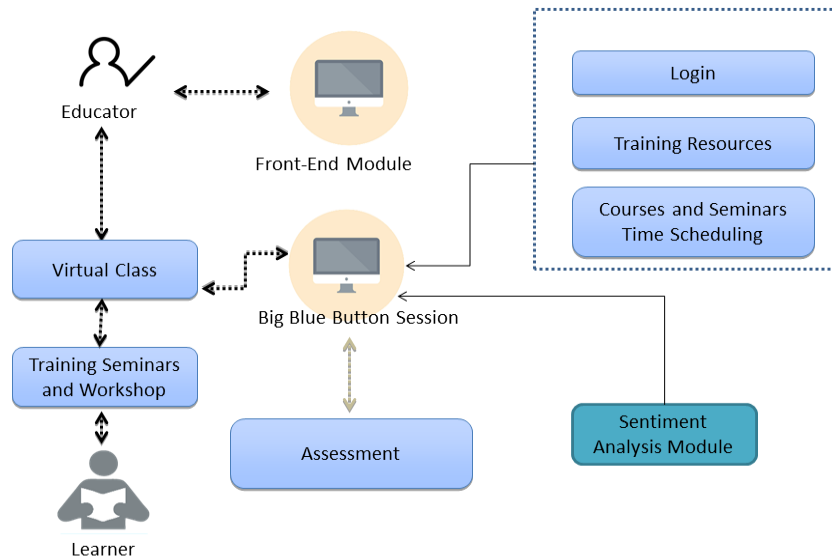


Fig. 1. The overview of the Hybrid Educational Platform

Frontend module: Upon successful login, time scheduling and basic administration operations can be efficiently performed. Also, some basic statistical data can be extracted and thoroughly presented to the user. Furthermore, a questionnaire survey is provided in order to evaluate, upon completion, the provided e-learning process, using an appropriate 5-grade Likert scale (see Table I) [24]. As is depicted in Table II, the evaluation tasks of the electronic questionnaire are related to (i) the functionality of HEP, (ii) the instructor's technical and communication skills and (iii) the objectives and the overall organization of the education topic. Additionally, a link to a virtual test-room is provided to verify software and hardware compatibility, in order to establish the functional synchronous e-learning environment described below.

Administration module: It provides the main administration interface of HEP, which is only accessible by the operators of the Greek School Network and the registered instructors with different administrative privileges. Basic operations such as monitoring the overall progress of the training course, scheduling or postponing a training task, tracking trainees who registered and participated in a training session, statistical data extraction, etc., are efficiently performed through this module.

Synchronous e-Learning Environment: The realization of each e-learning session is based on the BigBlueButton (<http://www.bigbluebutton.org>) open-source software, which is transparently integrated with the frontend module described previously. In principle, the administration of each real-time session is being performed by the instructor who can interact with the participants, upload slides to the presentation area and efficiently coordinate the related learning activities. The participants in each session that formulate the virtual class depicted in Fig. 1 are limited to 15.

Sentiment Analysis module: A vital part of the HEP platform is the sentiment analysis module which aims to analyze learners' opinions and specify their attitude towards the learning activities and the courses that they attend. The module operates in real-time and gets as input textual reviews (in the Greek language), analyzes them and automatically determines the users' satisfaction and their attitude. The module offers a wide range of possibilities to lecturers, policymakers and educational institutes that participate in the training procedure and offer life-long courses, to understand how their learners perceive learning activities and specify what aspects of the learning activities they liked and what did not. The Sentiment analysis module is presented in section 4 where characteristics and its functionality are described in detail.

Questionnaire surveys were appropriately formulated for the evaluation of the overall process, thus considering both face-to-face seminars and synchronous e-Learning sessions. The questionnaire survey related to the e-training activities comprises three distinct sections related to (i) the objectives and the overall organization of the specific educational topic, (ii) the instructors' technical and communication skills and (iii) the functionality of HEP. This questionnaire is complemented with optional open type questions, thus allowing a free response and detailed evaluation of the related e-training course and providing also critical hints for potential improvement. These open type questions provide participants the opportunity to express their opinions and experiences on several aspects of the training procedures. This repository of textual data is rich in opinions and their analysis and specifications of users' opinions and thoughts can provide valuable information. In this context, this work presents the approach used and integrated into the HEP system to automatically analyze textual data and specify their opinion.

Table 1. Comparison Scale of the Evaluation Process

Bad	Poor	Fair	Good	Excellent
1	2	3	4	5
Insufficient	Barely sufficient	Sufficient	Good	Excellent

The participants' evaluation tasks for the provided sections of the questionnaire and also the related percentage scores using the 5-grade scale mentioned above are depicted in Table 2. As can be observed, the high percentage scores recorded for the evaluation indexes 4 (Good) and 5 (Excellent) indicate the successful organization of each educational topic, the adequate technical background of the instructor and the robustness of HEP. Furthermore, the recorded responses to the electronic questionnaires provide all the necessary input for a systematic and coherent analysis, performed in the clustering module.

Table 2. Evaluation Tasks of the Electronic Questionnaire Survey

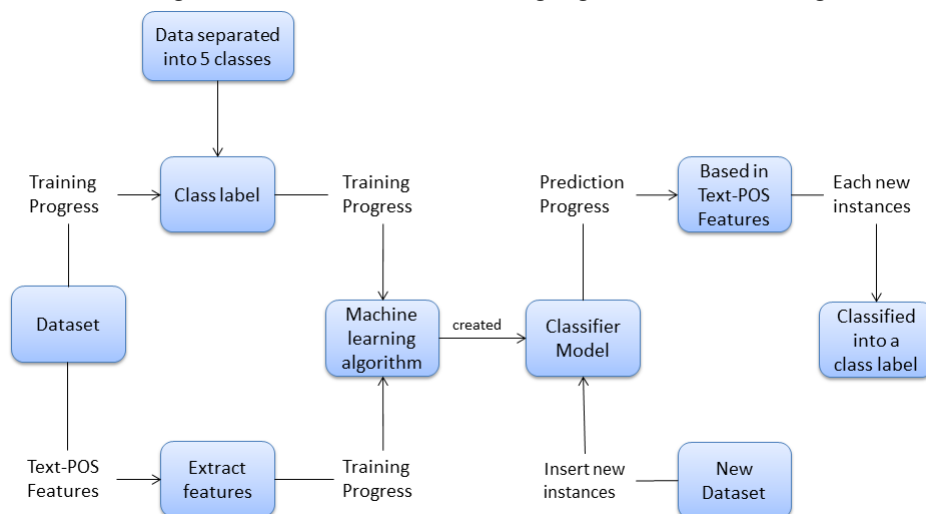
Sections	Tasks	% scores for the 5-level Likert scale				
		1	2	3	4	5
Objectives and the overall organization of the specific educational topic.	Were the objectives of the educational topics clearly presented?	0.1	0.3	4.7	45.1	49.7
	Did the thematic areas meet the objectives of the educational topic?	0.1	0.6	6.7	46.1	46.4
	Rare the provided reference training resources	0.1	0.8	7.4	45.4	46.3
	Was this e-learning session fruitful?	0.2	1.6	11.8	41.2	45.3
	Did the formulation and the presentation of the specific tasks help you to thoroughly understand the specific educational topic?	0.1	0.8	8.9	43.5	46.5
Instructors technical and communicational skills	Did the instructor motivate you in this session?	0.3	0.8	6.5	33.6	58.8
	Did the instructor analyze and present critical concepts and technical aspects simply and coherently, using appropriate examples?	0.1	0.7	5.5	33.2	60.6
	Was the instructor subject expert?	0.0	0.2	2.5	27.9	69.4
	Did the instructor assist you and provide useful hints for the realization of the activities?	0.6	0.9	6.2	36.3	56.0
Functionality of HEP	Was the functionality of HEP straightforward enough to perform basic administrator and time-scheduling tasks?	0.2	0.8	9.1	43.4	46.5
	Rate the sound quality	0.6	3.3	15.9	45.2	38.0
	Rate the video quality	0.5	2.4	12.6	41.8	42.6

4. Sentiment Analysis using Social Information

In this section, the sentiment analysis designed and integrated into the Hybrid Educational System is presented. The sentiment analysis module analyzes learners' textual and aims to provide teachers and educational instructors the opportunity to understand learners' opinions, their opinions and the degree of their satisfaction. More specifically, the sentiment analysis system classifies free-text user comments from according to their polarity, into very positive, positive, neutral, negative and very negative opinions. The sentiment analysis module has been integrated into the hybrid educational systems of the Greek school network that offers life-long courses and the module aims to analyze learners' opinions and specify their attitude towards the learning activities and the courses that they attend. The module operates in real-time and gets as input textual reviews (in the Greek language), analyzes them and automatically determines the users' satisfaction and their attitude. The module offers a wide range of possibilities to lecturers, policymakers and educational institutes that participate in the training procedure and offer life-long courses, to understand how their learners perceive learning activities and specify what aspects of the learning activities they liked and what did not. The overall workflow of the sentiment analysis module is illustrated in Figure 2.

Fig. 2. Workflow of the sentiment analysis procedure

In general, the workflow of the sentiment analysis module has two main stages. Firstly, the module is trained so as to be able to recognize the polarity of new user-generated data to it and to have satisfactory performance. To be achieved this, machine learning algorithms are trained with a large number of instances that are characterized by the text, POS and social features and the class [14], [23]. The classification algorithms are trained using a large number of data. In the second stage (prediction), the classifiers have the ability to classify new opinions represented as feature vectors in the appropriate class. For the classification stage, a number of machine learning algorithms were investigated. In the



following sections, the implementation procedure is thoroughly presented.

A characteristic of the sentiment analysis module concerns that it utilizes and analyzes social info [42][43][44]. The social info concerns mainly personal information and geographical data of the location of the participating users. More specifically, regarding the teachers' profile, we take into account the gender of the teachers, their school level, and their region and the development level of their prefecture. These are part of the social info.

5. Experimental Setup

In this section, we present the analysis performed and the knowledge extracted for the textual data of the users' opinion on the open text questions of the HEP with the aim to understand users' opinions and shed light on their behavior and their attitude towards HEP and the learning procedure.

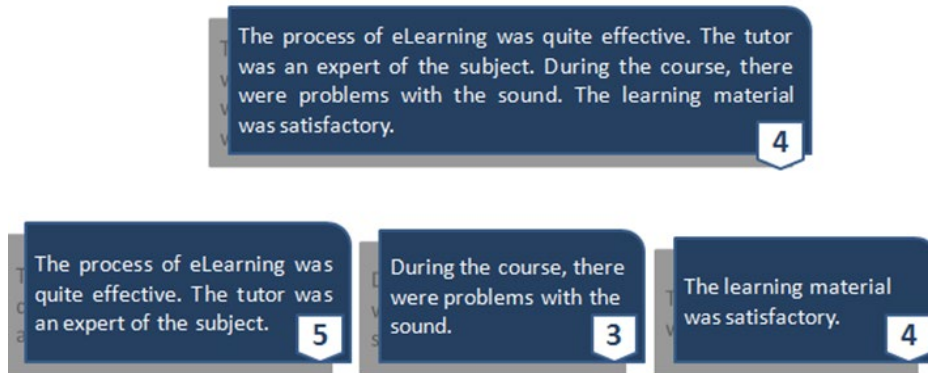
5.1. Description of the Textual Data

The data used in our work as mentioned above concern a set of opinions collected from over 2600 of teachers who participated in the e-learning courses through the Hybrid educational platform of the Greek School Network and expressed their opinions by answering open-ended questions in free text [18]. The data used in our work, are set of reviews of teachers that participated in lifelong learning courses which were conducted by the Greek School Network. More specifically, the reviews were made by over 2600 of teachers who participated in e-Learning courses on computer science [46][28][47] which were offered through teleconference services [19] and the teachers expressed their opinions and experiences by answering multiple choice and open-ended questions [8][45]. After the collection of the teachers' opinions and their textual data, their comments were analyzed by experts and they were manually annotated. During the analysis of the reviews, at the cleansing stage, misspelled words and typos were fixed in order for each sentence not to suffer from misspelling and orthographic errors.

More specifically, teachers' opinions were manually annotated by two experts in agreement on a five-level scale (1-5) where 1 was used to denote a very negative and 5 a very positive opinion respectively. Furthermore, we approached the users' reviews based on entity and feature level approaches, and when we determined that a review bears multiple meanings, then we divided it into parts manually according to the content carrying and we graded their emotional load from 1 to 5. Example reviews and the corresponding annotation are illustrated in Figure 3.

Fig. 3. Example of an opinion evaluation and classification in a class label

After that, a corpus of 11.156 annotated reviews was created and it was used to train machine learning algorithms. From these, 133 reviews were annotated with score 1, and belong to the very negative sentiment class, 584 were annotated with score 2 in negative

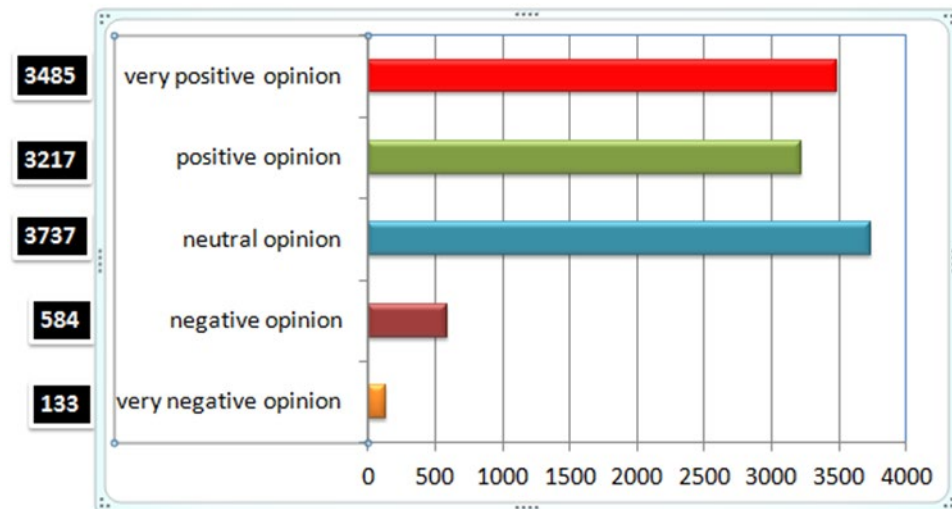


sentiment class, 3737 with score 3 in neutral sentiment class, 3217 with score 4 in positive sentiment class and 3485 were annotated with score 5 and belonged to the very positive sentiment class. In Figure 4, the distribution of the annotated reviews is illustrated.

Fig 4. Distribution of the annotated learners' opinions

5.2. Textual Analysis and Feature Extraction

In order for the textual user-generated data to be mined like any other type of data, the unstructured natural language document must be turned into a structured form. This stage



is called Information Extraction and it is the data generated by natural language processing systems. Initially, at the preprocessing stage stop-words, that are the words that do not bear meaningful information, are removed. The process of feature extraction constitutes a crucial procedure in text classification, which is directly influencing the accuracy of text classification [49]. Feature selection can be seen as a part of data pre-processing potentially followed or coupled with feature construction, but can also be coupled with the learning phase if embedded in the learning algorithm. An assumption of feature selection is that we have defined an original feature space that can be used to represent the data, and our goal is to reduce its dimensionality by selecting a subset of original features [48]. The original feature space of the data is then mapped onto a new feature space. In this work, three types of features are extracted which are the text-based features, the Part of Speech (POS) and the social features as illustrated in Figure 5.

The text-based features concern the number of (i) characters, (ii) words, (iii) capital letters, (iv) small letters, (v) special characters, (vi) average word length, (vii) sentences and (viii) digits which appear in a review. On the other hand, the POS-based features concern the number of (i) nouns, (ii) adjectives, (iii) verbs, (iv) proper nouns, (v) articles, (vi) pronouns, (vii) adverbs, (viii) prepositions and (ix) interjections.

The Part of Speech features are extracted with the use of the POS tagger of Xerox [17], a widely used tool for the morpho-syntactic analysis of text which also supports the Greek language. Furthermore, we extract additional features such as the number of (i) punctuations, (ii) verb tenses, (iii) foreign words.

The social features concern some personal characteristics of the participant teachers. So, in addition to the textual and the part of speech features, we take into account characteristics about the profile and the social data of the teachers. More specifically, regarding the teachers' profile, we take into account (i) the gender of the teachers, (ii) the school level, and (iii) their region and (iv) the development level of their prefecture. These are grouped in the category of the Social features of the data.

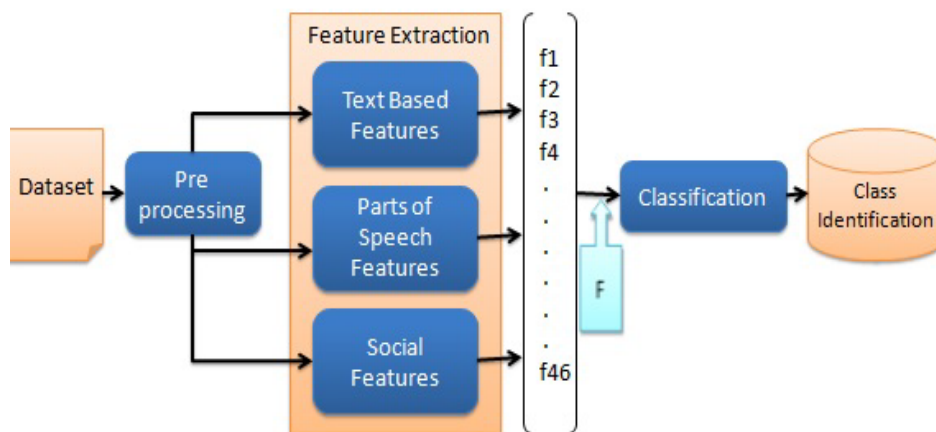


Fig. 5. The feature extraction procedure

In the context of the study, the Natural Language Toolkit (NLTK) was utilized with the aim to extract text-based features. The NLTK is a suite of program modules and data-sets for text analysis, covering symbolic and statistical Natural Language Processing (NLP). NLTK includes capabilities for tokenizing, parsing, and identifying named entities as well as many more features. The text-based features that were specified as the number of characters, words, capital letters, small letters, special characters, average word length, sentences and digits which appear in the uses reviews. So each review has been represented by a 46-feature vector that is comprised of text, POS and social features.

5.3. Machine Learning Algorithms Utilized

Supervised machine learning is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances [39] [40]. In particular, the main goal of supervised learning is to build a concise model of the distribution of the class label in terms of the predictor features in which the resulting classifier is then used to assign class labels to the testing instances where the values of the predictor features are known but the value of the class label is unknown. There are many categories of supervised learning algorithms and each category has its own characteristics but some issues are more or less generally applicable.

Specifically, in the context of our work, we used different kinds for classifying objects and the prediction was implemented over Decision Trees and k-nearest neighbors' classifiers. The experiments of the performance of the sentiment analysis module were conducted using the WEKA toolkit [41]. This provides implementations of state-of-the-art data mining and machine learning algorithms. It contains modules for data preprocessing, classification, clustering and association rule extraction. In the rest of this section, the main techniques used in data mining and their applications are briefly explained and discussed.

Decision trees C4.5 (J48) is an algorithm used to generate a decision tree developed by Ross Quinlan mentioned earlier. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The output given by J48 is the Decision tree. A Decision tree is the same as that of the tree structure having different nodes, such as root node, intermediate nodes and leaf node. Each node in the tree contains a decision and that decision leads to a result. The decision tree divides the input space of a data set into mutually exclusive areas, where each area has a label, a value or an action to describe or elaborate its data points. The splitting criterion is used in the decision tree to calculate which attribute is the best to split that portion tree of the training data that reaches a particular node.

Reduced Error Pruning Tree ("REPT") is a fast learning algorithm that uses decision trees. It builds a decision/regression tree based on information gain as the splitting criterion, and prunes it using the reduced error pruning method. It applies only to numerical data and results in the average value of the items of each sheet (ie it returns a numerical result again). The incomplete values are dealt with using a threshold, just as in the C4.5 algorithm.

IBK Classifier is a non-parametric method used for classification. In Weka, it's called IBK (instance bases learning with parameter k) and it's the lazy class folder. KNN is the k

parameter. IBK's KNN parameter specifies the number of nearest neighbors to use when classifying test instances, and the outcome is determined by the majority vote. Weka's IBK implementation has the "cross-validation option that can help by choosing the best value automatically Weka uses cross-validation to select the best value for KNN (which is the same as 'k'). An example of the instance-based learning algorithm is the k-nearest neighbor algorithm, kernel machines and RBF networks. These store (a subset of) their training set, when predicting a value/class for a new instance, they compute instances or similarities between this instance and the training instances to make a decision.

Support Vector Machines (SVMs) are supervised learning models with associated learning algorithms that are based on statistical learning theory. SVMs have been proven in many studies to be one of the most powerful learning algorithms for text categorization. SVM training algorithm mainly used to builds a model that assigns new textual data into a set of predefined categories. Given a set of N linearly separable points $S = \{x_i \in \mathbb{R}^n \mid i = 1, 2, \dots, N\}$, each point x_i belongs to one of the two classes, labeled as $y_i \in \{-1, +1\}$. A separating hyperplane divides S into 2 sides, each side containing points with the same class label only. Thus the goal of the SVM learning is to find the optimal separating hyperplane that has the maximal margin to both sides. Besides that, SVMs can be used as linear and non-linear classifiers. Since the feature space dimension is quite large in text classification, the classification problem is linearly separable and so, linear kernels are suitable to be used.

Recursive Neural Networks (RNN) are a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior. Derived from feedforward neural networks, RNNs can use their internal state to process variable-length sequences of inputs. This makes them applicable to tasks such as text classification and sentiment analysis.

Long-Short Term Memory (LSTM) constitute an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points but also entire sequences of data.

6. Experimental Results

In this section, we present the results collected from the experimental procedure. Initially, we present the feature extraction, ranking and selection procedure and how the selection of n-best features affects the performance of the classifiers. After that, we examine how the utilization of social information assists in the specification of users.

6.1. Feature Ranking

In total 46 features were extracted from the textual data. More specifically, 14 of them were text-based features, 29 of them were part-of-speech features and 3 were features concerning the users' social info. The aim of the study was to examine the performance of each classifier based on different feature sets like the 10-best features, the 20-better features and the 30-better features and test how the n-best features is able to affect the performance of

each classifier. As shown in Table 3, after using ReliefF, little deviation in the degree of fluctuation of characteristics is presented which demonstrates the lack of significant influence of the best features on the performance of classifiers. The ranking of the classification features indicates the importance of both the text and the POS features. Among the 20 first ranked characteristics, 10 belong to the text-based features and 9 of them to part-of-speech features and one relates to the social features. Specifically, the best rating score was achieved by the maximum length of words (0.15322696), after that was the average length of words with a relief score of 0.13064252 and then the total number of letters with a score of 0.12936403. An interesting factor concerns the average length of words which is rank as the second-best feature. We believe that is also to the fact that in Greek language many adjectives and adverbs that provide meaningful information are lengthy words. These may provide indicative clues in order to recognize neutral comments from comments that have polarity either positive or negative.

The comparison of the relief scores shows a typical deviation between the features and there is a substantial difference between the first twenty features. So, their selection will have a substantial impact on the performance of the machine learning classifiers.

Table 3 ReliefF ranking scores for the top-20 features

	ReliefF score	Feature description	Feature class
1	0.15322696	# maximum length of words	Text based
2	0.13064252	#average length of words	Text based
3	0.12936403	# total number of letters in a review	Text based
4	0.12469019	# total number of characters in a review	Text based
5	0.11616742	# total number of small letters in a review	Text based
6	0.09361249	# number of nouns in a review	Part of speech
7	0.07829686	# city/area of the teacher	Social
8	0.06859977	# total number of adjectives in a review	Part of speech
9	0.06184015	# total number of words in a review	Text-based
10	0.05628763	# total number of special characters in a review	Text-based
11	0.04493536	# total number of capital letters in a review	Text-based
12	0.04348413	# total number of possessive pronouns in a review	Text-based
13	0.04302483	# total number of pronouns in a review	Part of speech
14	0.03763799	# total number of articles in a review	Part of speech
15	0.02861032	# total number of adverbs in a review	Part of speech
16	0.02702445	# total number of verbs in a review	Part of speech
17	0.02683725	# total number of verbs in a specific tense	Part of speech
18	0.02663293	# total number of negative words in a review	Part of speech
19	0.02560867	# total number of sentences in a review	Text-based
20	0.02463112	# total number of the prepositions in a review	Part of speech

Based on the features and the annotated data, machine learning algorithms were trained and used to automatically analyze and determine the users' opinions.

6.2. Results

The performance of the algorithms was examined under different feature selection options and with/without the utilization of social info of the participants in the study.

Table 4: Sentiment Analysis without Social Information (baseline architecture)

Classifier	10-best	20-best	30-best	All features
IBk	49,5%	49%	51,91%	52,34%
Naïve Bayes	47,3%	47,14%	46,99%	47,57%
SVM (RBF)	51,76%	53,28%	55,31%	56,07%
SVM (poly)	48,2%	52,47%	54,06%	53,96%
MLP (1 hidden layer)	33,84%	35,48%	34,67%	35,48%
MLP (2 hidden layers)	34,65%	34,65%	33,96%	34,65%
MLP (3 hidden layers)	34,65%	33%	33,02%	34,67%
Bagging (REPTree)	47,04%	47,9%	47,7%	47,75%
Bagging (J48)	46,43%	48,18%	49,17%	49,42%
Boosting(AdaBoostM1)	42,15%	41,7%	42,15%	42,15%
J48	48,31%	47,83%	48,82%	49,15%

The results reveal quite interesting findings. Among the classifiers, the best performance is reported by the SVM (RBF) classifier which shows an accuracy of 56.07% when it was trained with all the features of the dataset. After that, the SVM (poly) classifier reported the second-best performance with an accuracy of 53.96% and after that the IBk classifier with an accuracy of 52.34%. Bagging J48 ensemble method follows with an accuracy of 49.42 and J48 with an accuracy of 49.15%. On the contrary, neural network procedures, MLPNN reports the lowest accuracy. Class imbalanced methods were also examined. Specifically, we examined oversampling and under-sampling methods without however getting better results on the dataset.

From a feature selection scope of view, the highest performance was reported when all the features were utilized. However, an interesting result concerns the fact that even when the 10-best features were utilized, the performance was quite good. The performance got better when 20-best features were utilized and even better when 30-best features were used. This stands for the performance of most of the classifiers. An interesting fact concerns the Naïve Bayes classifier where its performance was better when 10-best features were utilized than in case that the 30-best features utilized. The same stands also for the performance of AdaBoostM1 which was the same when 10-best features and 30-best features were utilized. This indicates that the selection of a subset of n-best features can provide indicative information about the data and assist the classifiers in modeling their classification tasks. This indicates that it has quite good accuracy and does not need expensive resources to operate.

The second phase of the experimental study concerns the utilization of social info of the participants, in addition to the text-based and POS features, and the examination of the performance of the sentiment analysis module of the system. Once again, the same classifiers were utilized and they were trained under the features sets including social info features. Their performance is depicted in Table 5 in the following.

Table 5: Sentiment Analysis with Social Information

Classifier	Baseline	With Social info
IBk	52,34%	53,46%
Naïve Bayes	47,57%	50,31%
SVM (RBF)	56,07%	59,64%
SVM (poly)	53,96%	57,71%
MLP (1 hidden layer)	35,48%	35,48%
MLP (2 hidden layers)	34,65%	33,84%
MLP (3 hidden layers)	34,67%	33,96%
Bagging (REPTree)	47,75%	47,9%
Bagging (J48)	49,42%	50,39%
Boosting	42,15%	42,15%
J48	49,15%	48,94%

The results indicate that the best performance is reported once again by SVM (RBF) which has an accuracy of 59,64%. This indicates that the utilization of social info improved the performance of the classifiers. Indeed, there was an increase of almost 3,5% in the accuracy of the classifiers and in the performance of the sentiment analysis module when social info was utilized. So, in this line, it could be very assistive to educational systems to integrate opinion mining procedures that utilize social info of the users in order to specify their opinions and attitude towards the courses and the learning activities that are offered to them more efficiently.

In addition, we examined the performance of deep learning methods and more specifically, we trained and used Recurrent Neural Networks (RNN) and Long Short Term Memory (LSTM) architectures[50][60][61]. In the table below we can see the performance of the models (Table 6).

Table 6: Performance of Deep Learning Methods

Classifier	Performance
LSTM	71.12%
RNN	66,21%

The implementation of the models was made in TensorFlow with Keras and their native hyperparameters and structure were used. The results indicate that the LSTM achieved superior performance between the models examined. For the representation of the textual data, word embeddings was adopted as a medium to handle data. A word embedding is a class of approaches for representing words and documents using a dense vector representation. The position of a word within the vector space is learned from text and is based on the words that surround the word when it is used. For learning word embeddings from text Word2Vec [62][63] was used.

7. Discussion

Sentiment analysis in the Greek language constitutes a very challenging area of research. Whereas the classification algorithms, machine learning or others like statistical, are important, the utilization of appropriate features and user social info are crucial and of extreme importance. The combination of social information of users in sentiment analysis procedures can substantially improve the performance and the accuracy of the sentiment recognition and it allows the extraction of certain information that is meaningful and assistive to the classification task.

The finding regarding the social group of features confirms the sociolinguistic studies supporting that the integration of user personal information proves to be important in the sentiment analysis measurements. As previously indicated, the social info that was utilized in the context of this study concern the gender of the teachers, their school level, their prefecture and the development level of their region. These features were taken into account and formulated the dataset of the experimental study. The feature ranking performed by ReliefF algorithm specified the prefecture to be the most meaningful feature of the social features group that provides the most information. The prefecture constitutes both a social and a geographical characteristic which is the base also for the determination of the development level of the region. In the next figure, the teachers' gender and prefecture development status distributions are illustrated.

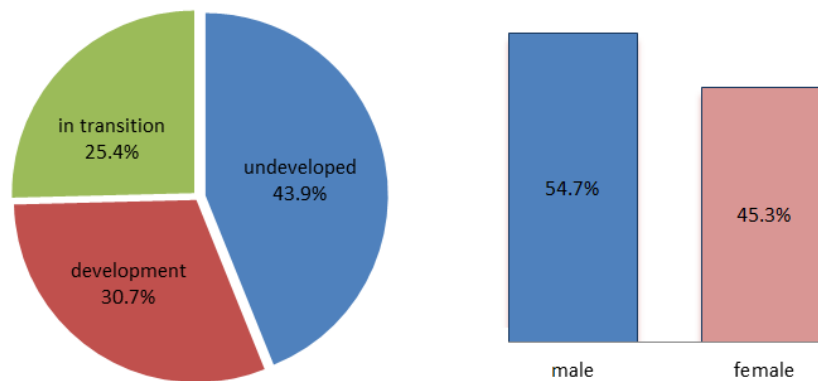


Fig. 6. The teachers' gender and prefecture development status distributions

Regarding the prefecture, according to ReliefF, this was ranked first in the social features group. After that, the development type of the region was ranked 2nd. The feature had three possible values indicating a region to be developed, undeveloped or in transition. The development according to the cohesion policy was specified in terms of income per capita based on the per capita income of the EU-27. A region is considered to be developed if it has a per capita income greater than 90% of the mean per capita value of EU-27. It is considered to be undeveloped if it has a per capita income less than the 75% of the mean per capita value of EU-27 and otherwise to be in transitions (per capita income greater than 75% and less than 90%). These findings indicate that the location in combination with the

development of the region constitutes aspects of the users that can be utilized in modeling their profiles and can assist in understanding their attitudes, opinions and behavior.

Finally, the gender feature was ranked next by the reliefF and is among the informative clues that prove that the knowledge of the gender improves the understanding of users' opinions and thus the performance of the classifiers. Concerning the social features, it appears that they are critical in understanding users and in understanding their opinions and their attitudes. Ultimately, the findings of the locational and social characteristics can provide important info about the teachers to service providers such as the Greek School Network in Greece as well as other school networks in other countries about the needs and the profiles of the educators and the teachers of the school community.

The second stage of the experimental studies provides interesting findings regarding the performance of classification algorithms. A wide range of classification methods was examined and specific classifiers were trained under various features sets and their performance was evaluated in terms of accurate predictions. The results indicated that the best classifier was achieved by kernels stream of methods and more specifically by the SVM (RBF) algorithm which reported accuracy of 59,64%. This indicates that the utilization of social info improved the performance of the classifiers. Indeed, there was an increase of almost 3,5% in the accuracy of the classifiers and in the performance of the sentiment analysis module when social info was utilized. Another interesting finding was that the social features improved the performance of almost all the classification methods and the greatest performance improvement was reported on the SVM (poly) algorithm where the improvement was approximately 3,75%. In contrast, the lowest performance was reported by neural network approaches. Specifically, a multiplier perceptron neural network was examined under different hidden layer architectures and it reported unsatisfactory performance. As a result, MLP neural networks seem not to be a suitable option in the sentiment analysis task and the recognition of users' opinions in the Greek language.

8. Conclusions

The vast amount of user-generated content necessitates accurate methods to analyze and specify opinions. Students' opinions constitute an important source of information that can be useful to teachers and educational instructors in order to improve learning procedures and training activities. By extracting students' data related to their learning behavior, teachers can find appropriate teaching methods to stimulate student interest and contribute to constructive monitoring of learning progress during the duration of the course or to improve future courses. Students generated content and their reviews and their comments provide indicative information about the lectures that they attended and the training activities that they participated in. Educational systems opinion mining must possess mechanisms to collect learners' comments and automatically specify their opinions and attitude towards the courses and the learning activities that are offered to them.

The paper describes a Greek sentiment analysis system based on the way of writing texts written in the Greek language in order to generate features vectors which together

with classification algorithms to give us the opportunity of the classification of Greek texts based on their personal opinion and the degree of satisfaction. We created an automatic sentiment analysis system that classifies opinions polarity into five classes reporting quite satisfactory performance and accuracy percentages. More specifically, the sentiment analysis system classifies free-text user comments from according to their polarity, into very positive, positive, neutral, negative and very negative opinions. The sentiment analysis module has been integrated into the hybrid educational systems of the Greek school network that offers life-long courses and the module aims to analyze learners' opinions and specify their attitude towards the learning activities and the courses that they attend. The module operates in real-time and gets as input textual reviews (in the Greek language), analyzes them and automatically determines the users' satisfaction and their attitude. The module offers a wide range of possibilities to lecturers, policymakers and educational institutes that participate in the training procedure and offer life-long courses, to understand how their learners perceive learning activities and specify what aspects of the learning activities they liked and what did not. The experimental study shows quite interesting results regarding the performance of the sentiment analysis methodology and the specification of users' opinions and satisfaction. Several text-based and part of speech based features are extracted from textual data and a generic approach to analyze text and determined opinion is presented. The sentiment analysis module has been integrated into the educational system to automate the analysis of the textual data and assist in the better and more complete understanding of their thoughts and opinions.

Experimental evaluation studies were performed and shed light on the effect that various algorithms have on the classification performance. The results revealed were quite interesting. Specifically, the evaluation results indicate that the approach has very good performance in specifying opinions in the Greek text. In addition, the social info of the users constitutes a crucial piece of information that can have a substantial impact on the performance of sentiment analysis procedures. The methodology has an acceptable accuracy and does not need expensive resources to operate. The feature analysis demonstrates interested findings regarding the characteristics that provide indicative information for opinion analysis and embeddings combined with deep learning approaches yield quite satisfactory results. The use of sentiment analysis module in educational systems is not only necessary but also imperative, because of the multitude of aspects users express opinions about, and the mixed sentiments that are present in reviews.

There are many directions for future work. A main direction concerns the utilization of aspect-based methods to specify participants' opinions towards specific aspects they address. Also, another direction concerns the examination of additional classification schemas, deep learning approaches and also the study of techniques for handling imbalanced data. Finally, the examination of the architectures of deep learning methods as well as their parameterization concerns another interesting direction that future work could focus on.

References

1. Pang, B., Lee, L., Opinion mining and sentiment analysis, *Found. Trends Inform. Retrieval* 2 (2008) 1–135.
2. Liu, B., *Sentiment Analysis and Opinion Mining*, Morgan and Claypool Publishers, 2012 (May)
3. Cambria, E., (2016). "Affective computing and sentiment analysis" (PDF). *IEEE Intelligent Systems* 31 (2): 102–107
4. Pang, Bo; Lee, Lillian (2008). "4.1.2 Subjectivity Detection and Opinion Identification". *Opinion Mining and Sentiment Analysis*. Now Publishers Inc.
5. Cambria, E. Schuller, B. Y.-Q. Xia, New avenues in opinion mining and sentiment analysis (extended abstract), in: *Proceedings of IJCAI, Buenos Aires 2015*.
6. Pang, Bo; Lee, Lillian (2005). "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales". *Proceedings of the Association for Computational Linguistics (ACL)*. pp. 115–124
7. Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.
8. Angelopoulos, P., Paraskevas, M., Perikos, I., & Zarouchas, T. (2014). A highly motivated blended learning model oriented to computer engineering educators. In *Computer Science & Education (ICCSE), 2014 9th International Conference on* (pp. 457-461). IEEE.
9. Athanasia Koumpouri, Iosif Mporas, Vasileios Megalooikonomou, "Evaluation of Four Approaches for "Sentiment Analysis on Movie Reviews": The Kaggle Competition", In *Proc. of the 16th International Conference on Engineering Applications of Neural Networks (INNS)*, Article No. 23, EANN 2015, Rhodes, Greece.
10. Athanasia Koumpouri, Iosif Mporas, Vasileios Megalooikonomou, "Opinion Recognition on Movie Reviews by Combining Classifiers", *Speech and Computer*, Volume 9319 of the series *Lecture Notes in Computer Science* pp. 309-316, *SPECOM 2015*.
11. Baker, R. S. (2010). Data mining for education. *International Encyclopedia of Education* (3rd edition), 7, 112–118.
12. Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19 (1–2), 205–220
13. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics EACL 1997*, pp. 174–181. Association for Computational Linguistics, Stroudsburg (1997)
14. Liu B (2012) *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers
15. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113
16. Pang B, Lee L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts In: *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, ACL '04*. Association for Computational Linguistics, Stroudsburg, PA, USA. 2004
17. [https://open.xerox.com/Services/fst-nlp-tools/Consume/Part%20of%20Speech%20Tagging%20\(Standard\)-178](https://open.xerox.com/Services/fst-nlp-tools/Consume/Part%20of%20Speech%20Tagging%20(Standard)-178)
18. Paraskevas, M., Zarouchas, T., Angelopoulos, P., & Perikos, I. (2015). Formulating an adaptable e-training framework to computer science educators. *International Journal of Continuing Engineering Education and Life Long Learning*, 25(2), 151-173
19. Perikos, I., Gkamas, V., Zarouchas, T., & Paraskevas, M. (2016). Educational capabilities of a novel teleconference service integrated into the Greek school community. In *Proceedings of the 20th Pan-Hellenic Conference on Informatics*. ACM.

20. Kalochristianakis, M., Paraskevas, M., Varvarigos, E., and Xypolitos, N., The Greek School Network, a paradigm of successful educational services maturing based on open source technology, *IEEE Transactions on Education*, vol. 50, issue 4, pp. 321-330.
21. Cambria, E., Schuller, B., Xia, Y., Havasi, C.: New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems* (2) pp. 15–21. (2013)
22. Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W.: Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pp. 427-434. IEEE. (2003)
23. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.
24. Likert, R. A technique for the measurement of attitudes. *Archives of psychology*, 1932.
25. Baker, R. S. (2010). Data mining for education. *International Encyclopedia of Education* (3rd edition), 7, 112–118.
26. Pandey, U. K., and Pal, S., “A Data mining view on class room teaching language”, (*IJCSI International Journal of Computer Science Issue*, Vol. 8, Issue 2, pp. 277-282, ISSN:1694-0814, 2011.
27. Gurmit Kaur, Williamjit Singh, “ Prediction of Student Performance Using WEKA Tool”, *An International Journal of Engineering Sciences*, 2016, Vol 17, ISSN:2330-0332(online)
28. Surjeet Kumar Yadav, Brijesh Bharadwaj, and Saurabh Pal, *Mining Education Data to Predict Student's Retention: A Comparative Study*, 2012
29. He, Wu. “Examining students' online interactions in alive video streaming environment using data mining and text mining” *Computers in Human Behaviour*, 29 (2013): 90-102
30. Ahmad, F., N.H. Ismail, and A. Abdulaziz, *The Prediction of Students' Academic Performance Using Classification Data Mining Techniques*. *Applied Mathematical Sciences*, 2015. 9(129): p. 12
31. Ali, M. M. 2013. "Role of Data Mining in Education Sector" in *International Journal of Computer Science and Mobile Computing*, 2(4), 374-383.
32. S. K. Yadev & Pal., S. 2012. Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification, *World of Computer Science and Information Technology (WCSIT)*, 2(2), 51-56.
33. Bevinde Alisha Pereira, Anusha Pai, "A Comparative Analysis of Decision Tree Algorithms for Predicting Student's Performance", *International Journal of Engineering Science and Computing*, Vol 7, 2017
34. Amjad Abu Saa, 2016. Educational Data Mining & Student's Performance Prediction. *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 5, 2016.
35. Khobragade, L.P. (2015), "Students' academic failure prediction using data mining", Vol. 3 No. 5, pp. 2321-7782.
36. Osmanbegovic E., and Suljic, M., "Data mining Approach for Prediction of Student Performance" *Economic Review - Journal of Economics & Business* Vol. 10, issue 1, 2012.
37. D.Kabakchieva, "Predicting Student Performance by using Data Mining methods for classification.", *Cybernetics and Information Technologies*, Volume 13, 2013
38. Bydžovská, H. (2016), "A comparative analysis of techniques for predicting academic performance", *Proceedings of the 9th International Conference on Educational Data Mining*, pp. 306-311.
39. Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3), 159-190.
40. Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160, 3-24.

41. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
42. Ghose, A., & Ipeirotis, P. G. (2008). Estimating the socio-economic impact of product reviews: Mining text and reviewer characteristics.
43. Stavrianou, A., Velcin, J., & Chauchat, J. H. (2009). A combination of opinion mining and social network techniques for discussion analysis. *Revue des Nouvelles Technologies de l'Information*, 25-44.
44. Simaki, V., Aravantinou, C., Mporas, I., Kondyli, M., & Megalooikonomou, V. (2017). Sociolinguistic features for author gender identification: From qualitative evidence to quantitative analysis. *Journal of Quantitative Linguistics*, 24(1), 65-84.
45. Spatiotis, N., Mporas, I., Paraskevas, M., & Perikos, I. (2016). Sentiment Analysis for the Greek Language. In Proceedings of the 20th Pan-Hellenic Conference on Informatics. ACM.
46. Zarouchas, T., Perikos, I., Paraskevas, M., & Pegiazis, T. (2015). A hybrid training framework oriented to computer engineering educators. In Proceedings of the 19th Panhellenic Conference on Informatics (pp. 33-37). ACM.
47. Spatiotis, N., Perikos, I., Mporas, I., & Paraskevas, M. (2018, July). Evaluation of an Educational Training Platform Using Text Mining. In *Proceedings of the 10th Hellenic Conference on Artificial Intelligence* (p. 42). ACM.
48. Spatiotis, N., Perikos, I., Mporas, I., & Paraskevas, M. (2019, July). Examining the Impact of Discretization Technique on Sentiment Analysis for the Greek Language. In *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)* (pp. 1-6). IEEE.
49. Spatiotis, N., Paraskevas, M., Perikos, I., & Mporas, I. (2017, September). Examining the Impact of Feature Selection on Sentiment Analysis for the Greek Language. In *International Conference on Speech and Computer* (pp. 353-361). Springer, Cham.
50. Goodfellow, I., Bengio, Y. and Courville, A. (2017). Deep learning. Cambridge, Mass: The MIT Press.
51. Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, 77, 236-246.
52. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.
53. Tang, D., Qin, B., & Liu, T. (2015). Deep learning for sentiment analysis: successful approaches and future challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(6), 292-303.
54. Xiao, Y., & Cho, K. (2016). Efficient character-level document classification by combining convolution and recurrent layers. arXiv preprint arXiv:1602.00367.
55. Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In *Advances in neural information processing systems* (pp. 649-657).
56. Poria, S., Chaturvedi, I., Cambria, E., & Hussain, A. (2016, December). Convolutional MKL based multimodal emotion recognition and sentiment analysis. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on* (pp. 439-448). IEEE.
57. Mesnil, G., Mikolov, T., Ranzato, M. A., & Bengio, Y. (2014). Ensemble of generative and discriminative techniques for sentiment analysis of movie reviews. arXiv preprint arXiv:1412.5335.
58. Tang, D., Qin, B., & Liu, T. (2015). Learning semantic representations of users and products for document level sentiment classification. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (Vol. 1, pp. 1014-1023).

59. Rozental, A., Fleischer, D., & Kelrich, Z. (2018). Amobee at IEST 2018: Transfer Learning from Language Models. arXiv preprint arXiv:1808.08782.
60. Arras, L., Montavon, G., Müller, K. R., & Samek, W. (2017). Explaining recurrent neural network predictions in sentiment analysis. *arXiv preprint arXiv:1706.07206*.
61. Li, D., & Qian, J. (2016, October). Text sentiment analysis based on long short-term memory. In *2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI)* (pp. 471-475). IEEE.
62. Church, K. W. (2017). Word2Vec. *Natural Language Engineering*, 23(1), 155-162.
63. Rong, X. (2014). word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*.