Predicting Social Media Classrooms Efficiency using Educational Data Mining

Adel Sweisy
Information Systems Information Systems Department
Faculty of Computers and Information, Assiut University
Assuit, Egypt
adel.mohamed@fci.au.edu.eg

Ahmed Taloba
Information Systems Information Systems Department
Faculty of Computers and Information, Assiut University
Assuit, Egypt
ahmedtaloba@fci.au.edu.eg

Mohammed Fawzy
Information Systems Information Systems Department
Faculty of Computers and Information, Assiut University
Assuit, Egypt
mfseddik @fci.au.edu.eg

Kholoud Ali
Information Systems Information Systems Department
Faculty of Computers and Information, Assiut University
Assuit, Egypt
kholoudali.alomar@gmail.com

***Abstract*—In several fields, including higher education and scholarly practices, social media has a critical role to play. Over the years, exploring the facilities available for emerging technology for universities, educators, and students has increasingly become important for all of us employed in higher education. The Educational Data Mining (EDM) is considered as one of the new fields that is concerned with the expansion of the different techniques aiming to discover information from educational environments either over the traditional or distance learning. Over the last decade, increasing research interests in the use of data mining in education have been reported with a focus on various aspects of the educational process. Therefore, this review aims to construct a system to predict the efficacy of social media classrooms by applying input from students with Data Mining Techniques (DMT). The system consists of three main phases to test the developed prediction model, which are the pre-processing phase, the data mining phase, and the evaluation phase. Reasonable results are shown by the proposed model statistics.**

Keywords — Classroom; Social Media; Data Mining Techniques; Decision Tree; Prediction Model

I. INTRODUCTION

Lecturing is a term that most teachers seem to be either taking for granted or misusing. Lecturers typically go to class filled with course material to offer to students while being conscious of the limited time available to them each week, month, or semester [13]. Considering the advancement of social media technology such as Twitter and blogs, it can be a shared stimulus to allow both students and instructors to take an active and immediate part in and interact with each other on educational activities. It is possible for end-users to use mobile and web-based technologies to upload, co-create, discuss, and change user-generated content across these highly creative channels. People can now learn on online networks outside of the control of the institution, and depending on the nature of the connections made, the learning experience will vary. However, the big changes that were made in social and cultural environments do not appear to contribute to similar improvements in schools due to several factors, including restrictive school networking policies, hardware availability and the difficulty of successful technology integration [8]. Consequently, this study explores how social media supports the teaching and learning process, in addition to the different issues and concerns proposed in the previous studies and researches.

Web crawler is defined as a program used in computers to browse the World Wide Web in a methodical, automatic, or orderly manner. In addition, web crawlers are considered as text-based search engines that help users to access the web. Besides, web crawling is considered an effective method of gathering data from and keeping up with Internet that is rapidly growing. Users are capable to find their different resources using various hyperlinks. Search engines are used to retrieve useful knowledge from the internet [14].

Data mining is not just about methods or the used database applications. Data mining itself depends on the creation of an acceptable data model and structure that have the ability be utilized to process, define, and create the necessary decisions and knowledge [15]. In terms of the type and structure of the source data, social media and blogs are organized and formed in a layout that tolerates the process of data mining as a real model as possible [16]. Therefore, the goal of this study is to develop a prediction framework for the performance of the next classrooms by utilizing data mining techniques to the students' feedback to predict the desired outcomes. The rest of the paper is organized as follows; the related work is presented in Section II. This is followed by a portrayal of the classroom data and the parts of our framework in Section III. The experimental outcomes are reported in Section IV. Finally, a perspective to the future works and the paper's conclusion are proposed in Section V.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

II. Related Work

In 2012, The authors offered a qualitative experimental support for the principles of social learning offering techniques and examples of how it is possible to use social media to connect formal and informal learning [1]. One of the key concerns in this research study was the social media use to facilitate student learning. In this research, participants' experience suggested strategies for the integration of these approaches, not only in formal in-class environments, but also, more critically, for the promotion of student-centered and informal social and active learning. Research has confirmed that learning can be done if informal social networks are encouraged and guided by the agenda.

Sarah Edith et al. conducted a survey about Twitter's using as a medical education teaching tool [2]. A literature analysis was performed on a variety of databases, internet outlets and blogs investigating Twitter's usage in higher education. Twelve tips were created to use Twitter as a teaching tool and organized into: the mechanics of using Twitter, recommendations, and facts to integrate Twitter into many medical education situations. This is in addition to encouraging research into the use of Twitter in medical education. In addition, a literature research was performed using Twitter, education, medical education, health care, and microblogging combinations. In the conducted study, several databases were searched for: PubMed, which produced 20 papers, 12 of which were considered suitable. Besides, a search by Google Scholar produced 80 papers, 16 of which were considered suitable. Moreover, seven papers were published in a Google search and two papers associated to medical education and Twitter were published in an ERIC search. Reference sections have been assessed for all papers and more papers have been submitted. However, no new articles have been found in this methodology. Finally, a variety of blogs and web outlets have been discovered by references from other papers.

Tanja E Bosch [3] discussed the Facebook use for the students and engages lecturers with learners via modern social media. Based on virtual ethnography and qualitative interviews, this study has shown that Facebook using in teaching and learning, particularly for the growth of academic micro-communities, has potential beneficial advantages. However, this had some issues that are stay relevant, including ICT literacy and uneven access. A virtual ethnography and qualitative content analysis of 200 UCT student Facebook profiles were the main methodology for the research. Furthermore, in the study, semi-structured qualitative interviews were performed with a purposeful sample of 50 undergraduate learners and five presentations presently engaged via Facebook with their learners. The focus was on graduate students because they are usually the website's heaviest users. Additionally, many graduate students also work full-time employment, and as suggested by Stutzman may use Facebook more for networking than for social management. Facebook is considered as an example of web 2.0 tools, wikis, delicious, podcasts that stated as a prospective teaching and learning software. Moreover, it has been suggested that the young people new generation also referred to as Net Genres or Digital Natives, may be immune to traditional teaching and learning techniques. The quest for and retrieval of data on the Internet may have contributed to a shift in teaching styles, requiring more interactive environments and more realistic research-based learning methods.

The authors in [4] proposed a research study to investigate the teaching and learning approaches when using cell phones and smartphones devices in higher education. They provided a part of the results on the learning perceptions of learners with mobile devices and the social media roles they played. This qualitative study of researches concentrated on learners from three universities across the United States. The educators of the learners had integrated mobile computing devices like cell phones and smartphones into their classes. Data were gathered through interviews with the students. Two main topics were discovered from the interview data: (a) the advantages of student learning with mobile devices and (b) the frustration of learning with mobile devices. The use of mobile computing devices and social media has developed interaction possibilities, offered possibilities for cooperation. Besides, this have learners the ability to contribute in content development and communication using web 2.0 tools in addition to social media with continuous connectivity support. The key purpose of that study was to add importance and opportunities in greater education teaching settings for mobile computing systems and social media. The respondents acknowledged changes in their teaching regardless of the constraints identified, including fear of technology not functioning correctly, tiny mobile device keyboards making typing hard, and potential device distractions.

Douglas A.Powell et al. suggested a survey in academic practice in 2012 to use blogs and new media [5]. From their study, they found that researchers and extension staff at higher education organizations, and particularly publicly financed higher education institutions, should be encouraged to use new media including blogs to reinforce relationships with public stakeholders and enable interested people to communicate directly with research subject matter specialists. These connections can be proactively reinforced before crises or emerging threat events such as outbreaks of foodborne disease or natural disasters. While being more transparent and nimbler with outcomes does not substitute peer review rigors, blogs and other online communication forums constitute an extra mechanism for fast sharing of thoughts, methodologies, and findings from studies. Disclosure of the processes used for the procurement and transmission of data should be given and references should be quoted as suitable. Moreover, the accessibility of new media, such as blogs, helps to build a multi-way dialog and exchange of thoughts to complement traditional ways of communication used in studies, teaching, learning and expansion job in higher education organizations.

Eva Kassens produced a survey in 2012 to asking about the possibility of using Twitter to help learners to learn a certain topic [6]. As such, Twitter offers the advantages of more conventional instructional methods in learning environments. This exploratory research has shown the potential possibilities and drawbacks that Twitter causes to higher education in the e-learning society. In addition, this study filled the mostly uninvestigated territories of Twitter with some information gaps. Moreover, Twitter has a range of different advantages and drawbacks over conventional teamwork as an involved,

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

casual teaching method. The benefits are that Twitter can promote the pooling of information better than diaries and discussions between individuals, since Twitter encourages the exchange of ideas outside the classroom through an online forum that enables random access to such discussions to continue. Whereas, the drawbacks of Twitter are the restriction of critical thought and self-reflection due to the restricted personality of tweets. This study disputes the finding that Twitter learners were more quickly and easily accessible to emotions and to their own deficiencies. Instead, this research shows that diary students displayed a greater showing of self-reflection: more students defined their own flaws, while Twitter students recognized only other students' mistakes. There are several constraints to this research such as the limited size of sample restricts the performance generalizability. Having six learners in each group only permitted the teaching community to glimpse the benefits and disadvantages that Twitter could offer. Second, how this research measured the use, development and retention of information is obviously restricted. All techniques of evaluation involve apparent biases, as the implementation, development and retention of information is not equal to writing in diaries or on Twitter or to recalling facts in a quiz. However, the five mixed sources give significant lessons on the importance of using Twitter as a fresh method of learning. Third, the research only spanned more than a month. Because of these constraints, future research should sample a bigger study group, observe the development and implementation of information by learners over longer periods of time, and apply both methods of teaching to a multitude of subjects to provide further insight into Twitter's advantages and pitfalls.

In 2011, Rita Kop introduced a research paper that raised questions about the learner's level of presence, autonomy, and critical literacy needed for active learning connectivism [7]. In this study, a mixed-method method was used. In addition, a lot of surveys were performed, consisting of a combination of quantitative and qualitative issues, while observations, discourse analysis and secondary data analysis were also performed in the form of teaching analytics to capture and analyze information. There was also a group of lurkers, since it was hard to gain awareness of their interactions from events in the learning environment because they were invisible to the audience. Data were gathered using the #PLENK2010 tag on the Moodle course forums and wiki, the participant blogs and Twitter messages as well as on any other internet events. A restricted quantitative analysis of blog posts, Twitter and Moodle participation was possible because of the volume of information produced by the respondents and facilitators and the time constraints for producing this document, and the qualitative analysis of information for this article was restricted to the environment of Moodle and a selection of participant blogs. The study showed that there are some other conditions that clearly promoted the involvement and commitment of people to teaching in a connectivism teaching environment, including the "social presence" of facilitators and participants, which improved the formation and sense of belonging of the community that built trust and stimulated active participation.

In 2014, Jin Mao [8] proposed a survey depending on the design of a sequential blended method. The research examined the social media affordances of high school learners, their attitudes and views about these new techniques, and associated barriers and problems. The results of affordance showed that learners rely on social media for recreation and social relations in their daily life. The educational use of teachers for learning and teaching in the classroom is intermittent, while the use of students for learning purposes on their own appears comprehensive, but often incidental and informal. The proposed quantitative findings indicated that in education, learners generally demonstrate favorable attitudes and views about social media using. Exploratory factor assessment disclosed three elements that accounted for a total of 65.4% of the variance: (a) social media usage advantages, (b) social media usage disadvantages, and (c) present social media usage in schooling. The social media's understanding for teaching; close-minded versus open-minded, inherent social media uses; and the altered teaching ideas represented the main issues emerged from the interviews. Besides, the research results indicate that complex attempts are needed to design, scaffold, and interact with learners during the process to allow social media to be used as efficient teaching tools and to adjust the previous affordances of learners with these tools.

In 2012, there was G. The article suggested by Veletsianos [9] is to comprehend the naturalistic methods of academics in social networks particularly on Twitter. In order to arrive at dominant topics describing online social network practice, tweets from 45 academics were evaluated qualitatively. The results of this research showed that Twitter academics (1) shared their professional practice-related data, tools, and media; (2) shared classroom and student data; (3) requested assistance from and offered feedback to others; (4) engaged in social comment; (5) engaged in digital identity and impression management; (6) attempted to network and communicate with others; and (7) highlighted their participation in social comment. These results help the field understand the evolving practice of internet networks involving academic involvement. The information corpus consisted of each recognized participant's recent 100 tweets, yielding a total of 4,500 tweets. All information was gathered on a single day, although these tweets were published by respondents over a 9-month period with different posting frequencies between respondents. The data was gathered using a mixture of techniques: as outlined above, the first four users were manually found. The Twitter Application Programming Interface (API) was subsequently used to collect lists of user supporters and counts of supporters, finding all their supporters with 2,000 or more supporters. The involvement found on Twitter provides possibilities for such a vision, but it is still too early to transform social spaces online into closely knit groups of scholars. The snapshot presented in this document is promising, but further inquiry is required for the emerging nature of internet social spaces and internet scholarship.

Milos Jovanovic et al. [10] in 2012 applied classification models to predict the achievement of learners and cluster models to group learners in an e-learning setting based on their cognitive styles. Classification models outlined in this article should assist educators, students, and company individuals to

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

engage with learners at an early stage who are probable to be outstanding on a chosen subject. Clustering learners based on cognitive styles and general results should allow the teaching materials to be better adapted to their learning styles. The method is tested using well-established algorithms for information mining and assessed through multiple assessment measures. Model construction method included pre-processing of information, optimization of parameters and choice of attributes, which improved general efficiency. In addition, they suggested a Moodle module enabling the automatic extraction of information required for the assessment of instructional information mining and the deployment of models established in this research. However, there was one issue that was the research had to solve is the absence of information to analyze more thoroughly. This problem is targeted at many distance learning institutions.50 Usually, there are not many students registered in distance learning systems, which means fewer prospective participants in information mining studies. As a distance learning scheme was implemented at our University only a few years ago, further evaluation and verification of the hypothesis tested in this article will be possible in the future.

Reynol Junco in [11] produced a survey to investigate the connection between Facebook use frequency, Facebook involvement, and student engagement. In three aspects, student engagement was assessed: a scale of 19 items based on the National Student Engagement Survey, the time needed in preparing for school, and time needed for co-curricular activities. The outcomes indicated that Facebook use was considerably predictive of the score of the commitment scale and positive predictive of the time invested in co-curricular activities. In addition, some Facebook operations predicted the dependent variables positively, while others predicted negatively. All students were surveyed at a medium,4-year, government institution, mainly residential in the Northeast. Students were approached through their on-campus email accounts during the Autumn 2010 semester and sent a connection to a study hosting a commercial survey-hosting website on SurveyMonkey.com. There were sent two extra reminders, each week apart. In addition, there was an offer for the participants to join a drawing to have the chance to win one of Amazon.com 90 gift cards of the $10. An approximation of 2368 studies were conducted with a 44 percent response rate. They found that it can be positive, negative predictive, or positive and negative predictive depending on the outcome variable, including hours spent on Facebook and time spent engaged in such Facebook operations. For instance, time spent on Facebook is a positive prediction of co-curricular time spent while playing matches on Facebook is a negative prediction.

In 2016 [12], Stefania Manca and Maria Ranieri suggested a study that would report the outcomes of a survey to the Italian academic employees in order to identify the social media using in the domain of university teaching methods. The reaction rate was 10.5%, which is 6139. The main questions for the participants were to define their use frequency, motivations, teaching methods and barriers associated with using multiple tools. These tools included generic social media sites like Facebook, websites of the academic and professional services such as LinkedIn, writing and commenting tools as blogs, wikis, and archiving and retrieving content for presentations and group work like SlideShare. Besides, this study included an online investigation of professors employed in field of the higher education in Italy. Data analyzes evaluated which socio-demographic factors mainly influenced the frequency of use and the relationships between motivations, methods of use, barriers to use and scientific discipline. The study's findings indicated that social media using is still limited and restricted. Besides, scholars are found not willing to integrate the social media platforms through their activities. This was due to several reasons, such as cultural opposition, pedagogical problems, or institutional limitations. However, there are variations between scholars in how they use or perceive social media that highly depended on the teaching science discipline. Overall, the findings highlight ambivalent attitudes towards social media's benefits and difficulties in the higher education's context, with barriers prevailing over benefits.

Winner Dominic proposed a research in 2017 that integrated Twitter and blogs into two undergraduate courses provided at Mzuzu University's Department of Library and Information Science, a government university in Malawi [13]. Two main ways were used to collect data. The first was that blog posts and Twitter posts were evaluated by learners, and the second was a questionnaire that had been sent to 64 students to assess their experience of using blogs and Twitter in a classroom setting. The research findings showed that Twitter and blogs are catalysts for the heavily hyped learner-centered learning strategy if properly applied, because learners exchanged and discussed the different materials of course using these strategies, posted their course reflections, and interacted 24/7 with each other and their lecturer. The challenges encountered include Internet information bundles costs, inaccessible Wi-Fi, bad bandwidth, and inadequate computers. Based on the research results, the author indicates some suggestions that can encourage the social media using at MZUNI or universities with comparable financial and technological backgrounds if applied. First, MZUNI and other greater learning institutions should allow the Internet to be available to all learners freely or highly subsidized. Wi-Fi should be designed to cover the entire campus and be easily installed on the different devices used by learners. Second, it should not be taken as a matter of fact that that learners can find ways of accessing the Internet when an instructor introduces Internet-based systems into a course. Instead, MZUNI lecturers focused on the integration of social media or associated techniques into their classes should create previous agreements with the MZUNI Library and the MZUNI ICT Directorate to ensure that learners have no difficulty accessing computers

III. THE RESEARCH'S ARGUMENT

The study in [5] is considered the most suitable framework due to using different classification models in predicting achievement of learners and cluster models to group learners in an e-learning setting based on their cognitive styles. they applied clustering technique to cluster learners. In the evaluation phase they tested their cluster's results using multiple assessment measures. They suggested a Moodle module enabling the automatic extraction of information

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

required for the assessment of instructional information mining and the deployment of the proposed models. It will be great if they apply further evaluation and verification of the hypothesis tested in this article as they mentioned in the future work section.

IV. Material and Methodes

A. Classrooms Data

Data from 200 students from different universities in Egypt were collected by crawling the classrooms and an online survey. Many extensive experimental studies had been done to judge and assess the model's performance. Feature selection algorithm has been applied on dataset as shown in Figure. 1.

B. Prediction Model

In this research, a framework has been built to predict the class rooms efficiency by applying Data Mining Technique (DMT) on the extracted data from the actual classroom using web crawling and survey answers, then extract the result of DMT to be a knowledge base for the application to perform the prediction process. Figure. 2 shows the proposed framework which consists of three main phases. The first is the data preprocessing phase in which data are prepared in order to apply the DMT technique. The second is the DM phase in which the DMT is applied. The third is evaluation phase that is used to evaluate the built model's performance and accuracy by the use of a certain data mining evaluation technique.

In Data Preprocessing phase, a sequence of steps is to be applied to clean, split, and select the most appropriate topographies for the model from the classroom data for applying DMT and resample the data sets into training and test sets. The data cleaning phase is to clean the data and remove the records that contains empty values. The feature selection phase is to select a subset of topographies relevant to the target DMT from all the topographies of the data set. In the filtering approach, the feature selection algorithm did not depend on the DMT which applied to the selected topographies. The feature selection is applied on the classroom data and the selected attributes were sharing multimedia. sharing course content, answering urgent questions, course discussion, receiving notes, helps in learning and getting help without the lecturer. The class label representing the result is considered as the feedback

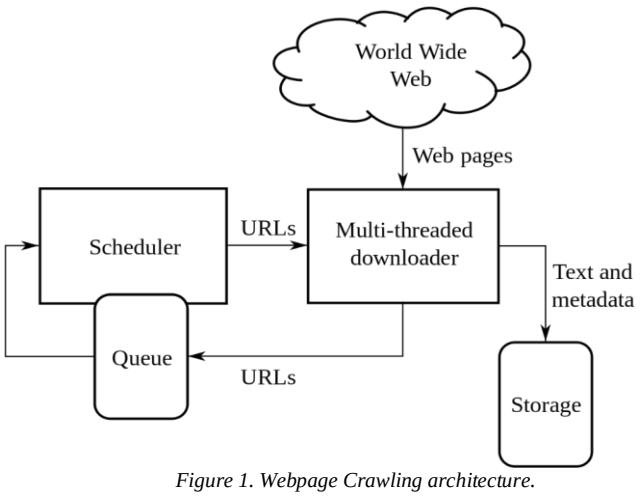

Figure 1. Webpage Crawling architecture.

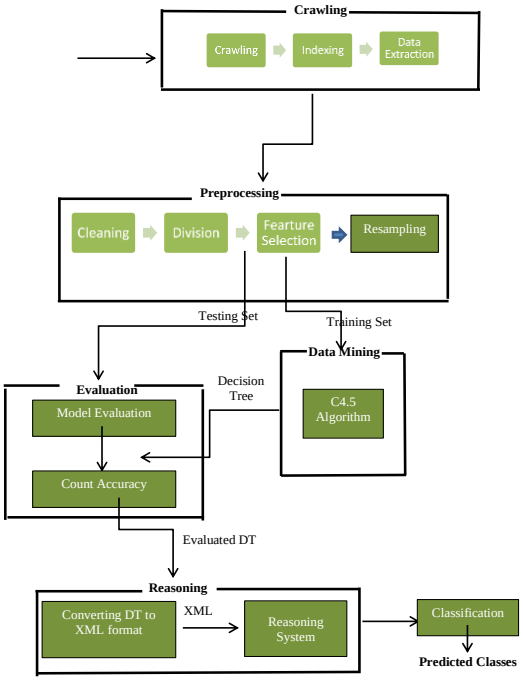

Figure 2. The Framework of predicting the classroom efficiency

of each student. For resampling, the data were divided into 90% as a training set and 10% as a testing set using unsupervised resample filter in Weka

In the data mining phase, Weka implementation of C4.5 (Weka J48) decision-tree learning algorithm was applied the training data sets. C4.5; is based on the ID3 algorithm and tries to find simple (or small) decision trees (DT's). The premises on which the algorithm is based will be provided in the following sections. some premises that guide the algorithm are as follows; If all cases are of the same type of class, the tree is considered as a leaf and that leaf is returned labelled with that type of class. For each attribute, the potential information is calculated based on the probabilities of each case to have a specific value for the attribute. Besides, the information gain is calculated using a test on the attribute to determine the probabilities of each case having a specific attribute's value to belong to a particular class. Depending on the current selection criterion, best attribute to branch on is to be find.

1. *Counting gain*

   The "Entropy" is used in this process as a measurement for the disorder of the data. The Entropy of $\vec{y}$ is calculated using equation (1).

$$Entropy\ (\vec{y}) = -\sum_{j=1}^{n} \frac{|y_i|}{|\vec{y}|} \log \frac{|y_i|}{|\vec{y}|} \qquad (1)$$

   Iterating over all possible values of $\vec{y}$, The conditional Entropy is calculated using equation (2)

$$Entropy\ (j|\vec{y}) = \frac{|y_i|}{|\vec{y}|} \log \frac{|y_i|}{|\vec{y}|} \qquad (2)$$

   and finally, we define Gain by equation (3)

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$Gain\ (\vec{y}, j) = Entropy\ (\vec{y}) - Entropy\ (j|\vec{y}) \quad (3)$$

2. *Pruning*

   The pruning is an important step to the result because of the outliers. All data sets include a little subset of instances which are not well-defined and vary from the other ones in its neighborhood. After the whole creation processes of the tree, which classify all the training set instances, it is pruned. This is to minimize classification errors which can be occurred because of specialization in the training set; we do this to make the tree more concise [15].

3. *Results*

   To show concrete examples of the C4.5 algorithm, WEKA software tool has been used on training sets. Fig.3 shows the resulting classes are about the students' recommendation about classrooms.

In the evaluation phase, the universality of the model was validated using the test data sets by the hold-out validation method. The holdout approach is known to be the simplest form of cross validation. The data set was split into two main sets, which are the training set and the testing set. The function approximator only suits a function using the training set. Then, it is used, in the test set, to estimate output values for the data that have never seen such output values before. Then the produced errors are accumulated to give the mean absolute error for the test set that will be used in the model evaluation. This approach is typically superior to the residual method and does not require much time to calculate. The estimate is pessimistic since only part of the initial data is used to derive the model as shown in Figure 4. Random subsampling is a

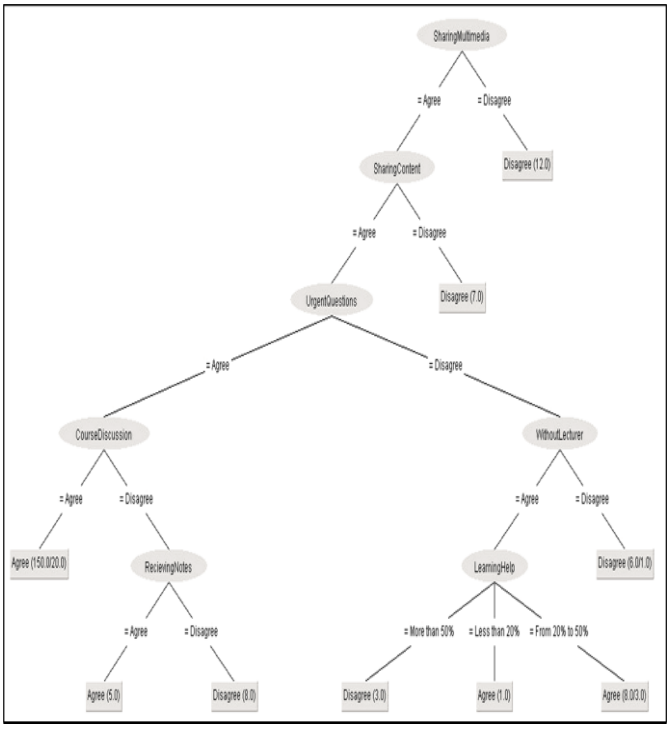

Figure 3. DT of Classroom data Created by the C4.5 algorithm.

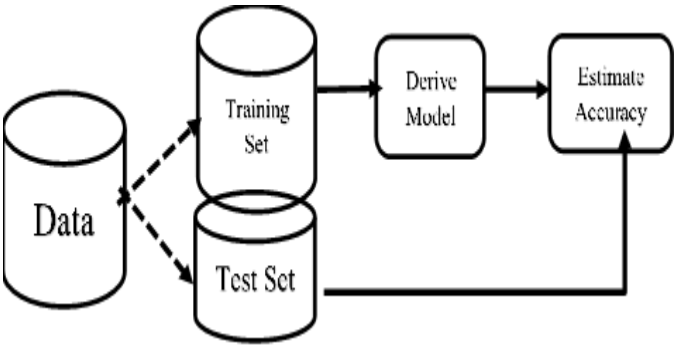

Figure 4. Estimating accuracy with the holdout method.

variant of the holdout methods where the holdout method is repeated with k several times. The cumulative estimation of accuracy is taken as the average of the accuracy of each iteration.

In the reasoning phase an application has been developed with C# programming language to perform the prediction operation. The knowledge base of this application is the model of decision tree algorithm. It is applied on WEKA which delivers rules in newick format as shown in Figure 5 which has been converted into XML as shown in Figure 6. This is the rules format to be used as the knowledge of our DSS application.

It allows the user to input the classroom attributes' values, and the evaluation of the classroom will be shown on the screen by matching the user input with the rules which inserted into the application in the previous step.

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

SharingMultimedia = Agree
|   SharingContent = Agree
|   |   UrgentQuestions = Agree
|   |   |   CourseDiscussion = Agree: Agree (150.0/20.0)
|   |   |   CourseDiscussion = Disagree
|   |   |   |   RecievingNotes = Agree: Agree (5.0)
|   |   |   |   RecievingNotes = Disagree: Disagree (8.0)
|   |   UrgentQuestions = Disagree
|   |   |   WithoutLecturer = Agree
|   |   |   |   LearningHelp = More than 50%: Disagree (3.0)
|   |   |   |   LearningHelp = Less than 20%: Agree (1.0)
|   |   |   |   LearningHelp = From 20% to 50%: Agree (8.0/3.0)
|   |   |   WithoutLecturer = Disagree: Disagree (6.0/1.0)
|   SharingContent = Disagree: Disagree (7.0)
SharingMultimedia = Disagree: Disagree (12.0)

Number of Leaves  :     9

Size of the tree :      16
```

Figure 5. DT rules in newick format, Created by the C4.5 algorithm.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
<?xml version="1.0" encoding="utf-8"?>
<Root>
  <Inference Name="r6">
    <Cluster Name="r6">
      <Rule Name="r0" Disorder="CAgree">
        <Tuple Cpt="classroom" Prop="SharingMuimedia" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="SharingContent" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="UrgentQuestions" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="CourseDiscussion" Val="=Agree:" />
      </Rule>
      <Rule Name="r1" Disorder="CAgree">
        <Tuple Cpt="classroom" Prop="SharingMuimedia" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="SharingContent" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="UrgentQuestions" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="CourseDiscussion" Val="=Disagree" />
        <Tuple Cpt="classroom" Prop="RecievingNotes" Val="=Agree:" />
      </Rule>
      <Rule Name="r2" Disorder="CDisagree">
        <Tuple Cpt="classroom" Prop="SharingMuimedia" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="SharingContent" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="UrgentQuestions" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="CourseDiscussion" Val="=Disagree" />
        <Tuple Cpt="classroom" Prop="RecievingNotes" Val="=Disagree" />
      </Rule>
      <Rule Name="r3" Disorder="CDisagree">
        <Tuple Cpt="classroom" Prop="SharingMuimedia" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="SharingContent" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="UrgentQuestions" Val="=Disagree" />
        <Tuple Cpt="classroom" Prop="WithoutLecturer" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="LearningHelp" Val="=More than 50%" />
      </Rule>
      <Rule Name="r4" Disorder="CAgree">
        <Tuple Cpt="classroom" Prop="SharingMuimedia" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="SharingContent" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="UrgentQuestions" Val="=Disagree" />
        <Tuple Cpt="classroom" Prop="WithoutLecturer" Val="=Agree" />
        <Tuple Cpt="classroom" Prop="LearningHelp" Val="=Less than 20%" />
      </Rule>
```

Figure 6. Rules in XML format.

V. EXPERIMRNTAL RESULTS

This section demonstrates the evaluation to the performance of the proposed framework and results using the applied DM techniques. At first, serial of problems like there are not enough datasets via internet have been countered. Then a crawler has been created to fetch the data from social media classrooms and blogs, we have created an online survey for the student's feedback. We discovered that using real data of our local community, raises credibility of data records. Many extensive experimental studies had been conducted in order to assess the performance of the proposed model. Feature selection algorithm has been applied on the dataset. Sharing multimedia. sharing course content, answering urgent questions, course discussion, receiving notes, helps in learning and getting help without the lecturer features were selected. A subset of 10% of the data had been selected to test the model and the 90% used to build the classifier.

After applying the model, a large scale of statistical information was obtained. These performance measures had been used to evaluate the model as shown in Table 1. This table shows the performance for Diseases Diagnosis. The DT had 93.3% sensitivity, 50% specificity and 82.5% accuracy.

Table1: The evaluation metrics of the classifier

Size of the tree	Number of leaves	TP	TN	Sensitivity	Specificity	AUC %	Accuracy %
19	9	33	7	93.3%	50%	73.2 %	82.5%

VI. CONCLUSION AND FUTUR WORK

In this research, a framework has been built to predict the efficiency of the classrooms by applying Data Mining Techniques (DMT) on the extracted data from the actual classroom using web crawling. Data from 200 students from different universities in Egypt which collected by crawling the classroom blog page has been analyzed, it contains the most effective factors of the education process, the feedback of the students.

Feature selection algorithm has been applied on dataset. Dataset has been divided into two main categories, which are the training dataset and testing dataset. Training set has been used to construct our classifier. Decision Tree J48 Algorithm has been applied on our training dataset for the classifier constructing. Testing dataset has been used in the classifier's evaluation. A statistical metrics has been analyzed. After that extraction of the result of DTs was performed by converting the WEKA newick format into XML format using our developed tool. This XML format rules were used to construct a knowledge base for our reasoning application to perform the prediction operation. It has been shown in the experimental results that the classifier gives acceptable results.
The DT which resulted from WEKA was in newick tree format which converted into XML format by our developed tool. The rules in XML format will be stored into the reasoning application as a knowledge base. In the reasoning phase an application has been developed with C# programming language to perform the prediction operation. It can be considered as the predicting system to predict the following classrooms performance and efficiency.

Since it is believed that the value of social media is not limited to social communication, but it plays a vital role in the process of education [17]. In addition, it is considered that apart from looking for improving the social media classrooms, future research should look for collecting more dataset by using different blogs and social media pages. More surveys are supposed to be done to collect students' feedback and suggestions. These data sets will be used to train other classifiers and to try more experiments. Also, other techniques should be investigated and studied to be applied and combined more of them if needed to create a hybrid predicting framework to reach as high accuracy as possible.

REFERENCES

[1] Chen, B., & Bryer, T. (2012). Investigating instructional strategies for using social media in formal and informal learning. The International Review of Research in Open and Distributed Learning, 13(1), 87-104.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[2] Forgie, S., Duff, J.P., & Ross, S. (2013). Twelve tips for using Twitter as a learning tool in medical education. Medical teacher, 35 1, 8-14.

[3] Tanja E Bosch (2009) Using online social networking for teaching and learning: Facebook use at the University of Cape Town, Communicatio: South African Journal for Communication Theory and Research, 35:2, 185-200.

[4] Gekas, Joanne & Grant, Michael. (2013). Mobile Computing Devices in Higher Education: Student Perspectives on Learning with Cellphones, Smartphones & Social Media. The Internet and Higher Education. 19. 18–26.

[5] Powell, Douglas & Jacob, Casey & Chapman, Benjamin. (2011). Using Blogs and New Media in Academic Practice: Potential Roles in Research, Teaching, Learning, and Extension. Innovative Higher Education. 37.

[6] Kassens-Noor, Eva. (2012). Twitter as a teaching practice to enhance active and informal learning in higher education: The case of sustainable tweets. Active Learning in Higher Education. 13. 9-21.

[7] Kop, R. (2011). The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. The International Review of Research in Open and Distributed Learning, 12(3), 19-38.

[8] Chawinga, Winner. (2017). Taking social media to a university classroom: teaching and learning using Twitter and blogs. International Journal of Educational Technology in Higher Education. 14.

[9] Veletsianos, G. (2012). Higher education scholars' participation and practices on Twitter. Journal of Computer Assisted Learning, 28(4), 336-349.

[10] Jovanovic, M., Vukicevic, M., Milovanovic, M., & Minovic, M. (2012). Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study. International Journal of Computational Intelligence Systems, 5(3), 597-610.

[11] Junco, R. (2012). The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. Computers & Education, 58(1), 162-171.

[12] Manca, S., & Ranieri, M. (2016). Facebook and the others. Potentials and obstacles of social media for teaching in higher education. Computers & Education, 95, 216-230.

[13] Chawinga, W. D. (2017). Taking social media to a university classroom: teaching and learning using Twitter and blogs. International Journal of Educational Technology in Higher Education, 14(1), 3.

[14] Amudha, S et al., Web crawler for mining web data. Int. Res. J. Eng. Technol. (IRJET) 4(2) (2017).

[15] J. Han and M. Kamber, Data Mining Concepts and Techniques 3rd edition, 2011.

[16] Mohammed A.Farahat, et al, "Response Prediction for Chronic HCV Genotype 4 Patients to DAAs," International Journal of Advanced Computer Science and Applications, Vol. 7, No. 12, pp. 173-178, December 2016.

[17] Tess, P. A. (2013). The role of social media in higher education classes (real and virtual)–A literature review. Computers in Human Behaviour, 29(5), A60–A68.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500