Use of Social Networks Sites (SNSs) as A Collaborative Learning Technique: Survey Analysis and Mining Approach

Nevine M. Labib[1], Ahmed E. Sabry[1], Rasha H. A. Mostafa[2], Edward W. Morcos[1]

[1]*Computer and Information Systems Department, Sadat Academy for Management Sciences, Cairo, Egypt.*

[2]*Business Administration Department, Faculty of Commerce, Ain-Shams University, Cairo, Egypt.*

Abstract— *This study adopts a multi-disciplinary approach, relating social psychology and information sciences. It aims at measuring the significance of social networks usage in collaborative learning using different information science techniques. After extensive review for relevant literature it has been noticed that the implementation context namely Middle East and North Africa (MENA) region is starving for such stream of researches.*

A number of studies underscored various aspects of the relationship between Social Network Sites (SNS) and collaborative learning such as perception, satisfaction, collaboration, engagement, integration, innovation, performance, interaction, problem solving, motivation, knowledge sharing and discovery, information sharing, and communication.

A survey targeting about 300 students as a sample of relevant stakeholders (users) was conducted over a period of one-year. Three data mining models are implemented using the transformation methods, clustering techniques, and decision tree classification methods. They are all included as part of the triangulation of methods for providing the research analysis higher credibility, reliability and validity.

The originality of this research stems from the following: first, applying novel methodological techniques in social networks domain. Second, improved validity and reliability of the results through triangulation of methods applied.

Keywords: *Social Network Sites (SNS), Data Mining (DM), Decision Tree (DT), Triangulation of techniques, K-means, Clustering, Association Rules, MENA (Middle East and North Africa), Collaborative Learning (CL)*

I. INTRODUCTION

Nowadays, Social Network Sites (SNS) are being used by students, not only for social interactions but also for learning activities since they increase student engagement. Hence, SNS can lead to the creation of virtual communities of learners, which eventually increase the overall learning [1].

Learning activities may include sharing information, doing assignments, discussing issues and other activities that fall under the umbrella of collaborative learning. Nevertheless, there are still more activities to be explored in order to benefit from SNS in the domain of education. This study explores the dimensions of SNS use for collaborative learning, among undergraduate and graduate University students, by means of data mining techniques.

1.1 Problem Definition and Objectives

This study discusses the different dimensions of Social Networks (SNs) in Collaborative Learning among University students using different data mining techniques.

As for the objectives, they consist of the following:

- Analyzing SNS dimensions in the domain of Collaborative Learning among University students in Egypt (based on data collected via structured questionnaire).
- Conducting a comparative study between three different data mining techniques in this domain.
- Comparing the results with statistical outcomes previously revealed.

1.2 Originality and Value

The originality of the study stems from applying different data mining techniques for social networks' use in education in Egypt [2]. Moreover, results' validity and reliability is improved through triangulation of the methods applied.

Results drawn out of this study may help educators to foster student learning by incorporating social media into taught modules. In addition, they will be able to deal with the negative effects of social media on different types of learners.

1.3 Structure of the paper

It starts by reviewing similar researches that make use of data mining techniques related to the use of Social Networks Sites (SNS's) in collaborative learning among University students. Second, it provides a detailed description of the survey used in the study. Then it discusses the data-mining framework along with the data collection details and relevant results. Finally, conclusions, recommendations, as well as future work are drawn.

II. LITERATURE REVIEW

In order to analyze and discover the role of SNS in education and the different interactions between students, several techniques may be used, whether linear or non-linear. In this paper, we explore the use of data mining techniques to analyze the role of SNS in collaborative learning among undergraduate and graduate students in the Egyptian Universities as part of MENA region.

A survey paper [3] studied several data mining techniques, such as graph theoretic, clustering, recommender system, semantic web, and opinion analysis and classification were used in order to analyze different aspects of SNS. This study showed that data mining techniques are very useful when it comes to retrieving information from a huge amount of data. The selected technique should be based on the kind of data to be analyzed [3].

Another study discusses the usefulness of social media for collaborative learning in higher education by using a social media platform, Graasp. It is implemented in a project-based course and evaluated from different perspectives, such as collaboration and knowledge management. It was found that students were satisfied with using Graaspas it was able to enhance knowledge management and collaboration [4].

A study used Spectral clustering as a data mining method to discover students' behavioral patterns performed in an e-learning system. In order to do so a software was developed. It allowed the tutor to define the data dimensions and input values to obtain appropriate graphs with behavioral patterns that meet his/her needs. Then, the discovered behavioral patterns were compared with students' study performance and evaluation with relation to their possible usage in collaborative learning [5].

A study tackled some SNA techniques, namely community mining, in order to discover relevant structures in social networks. Using new ideas in a toolbox, named "Meerkat-ED", which automatically discovers relevant network structures, visualizes overall snapshots of interactions between the participants aiming to facilitate fair evaluation of students' participation in online courses [6].

III. DESCRIPTION OF CONDUCTED SURVEY

A random sample of three Egyptian public universities students was drawn. The usable sample consisted of 300 students divided evenly between undergraduate and postgraduate. It is a common practice to rely on students' sample, specifically that they are considered heavy users of SNSs [7]. The characteristics of the sampled students are provided in Table1.

Following an extensive review of relevant literature in the areas of Technology Acceptance Model (TAM) and information technology, a self-administered multi-item structured questionnaire was developed to collect data in relation to the research problem. Moreover, seven constructs, including: Perceived Usefulness (PU) measured by four items, attitude measured by three items, and intention to use SNSs measured by four items. All multi-item scales were adapted from Davis [8]. Whereas, Perceived Enjoyment (PE) assessed by four multi-item scale, Perceived Connectedness (PC) measured by three items, and Perceived Involvement (PI) measured by three items were adopted from Nysveen et al., [9]. Further, actual use of SNSs multi item scales for collaborative learning (7 items) and socializing (9 items) were adopted and modified from Saw et al., [10] and Li [11]. All research constructs were assessed on five-point Likert-type scales. In addition, some demographic items were included in the questionnaire.

Table 1: Survey sample characteristics

Characteristics	**%**
Gender	
Male	54.6 %
Female	45.4 %
Age	
18-22	44 %
22-30	35 %
30 and above	21 %
Favorite SNs	
Facebook	96.6%
YouTube	48.6%
Twitter	26%
Length of Usage of SNs	
Less than a year	6.8 %
1-2 year	14.6 %
2 year+	78.6 %
Frequency of Using SNs	
Several time a day	74 %
Once a day	15.6 %
Every few days/Once a week	10.4 %

IV. MINING FRAMEWORK

Data mining is a term coined to describe the process of shifting through large databases in search of interesting and previously unknown patterns. The accessibility and abundance of data today makes data mining a matter of considerable importance and necessity. The field of data mining provides the techniques and tools by which large quantities of data can be automatically analyzed. Data mining is a part of the overall process of Knowledge Discovery in Databases (KDD).

The following techniques used throughout this research as part of the triangulation of techniques used for validating the results.

1. Unsupervised Clustering Using K-Means.
2. Mining Supervised Classification using Decision Tree.
3. Rules induction using association rules.

A cluster analysis is a type of classification and analysis phase techniques within data mining frameworks. A major issue with cluster analysis is identifying the appropriate number of clusters. Following Lehmann (1979), initial guidelines and the given sample size of 158, the appropriate number of clusters for the available data falls in the range of two to five clusters. Hence, Hierarchical clustering was used to derive solutions within these ranges. "Ward's method was chosen to minimize the within-cluster differences and to avoid problems with "chaining" of the observations found in linkage methods" [12].

One of the most important phases of a Data Mining process (and one that is usually neglected) is that of data exploration through visualization methods.

Visualization feature is considered as one of the important tools for disseminating results in order to discover valid, novel and potentially useful patterns from this relatively highly dimensional and large amounts of data and make use of those patterns to come up with some rules, interpretation, and prediction. The analyzed data cover metrical scales for the computations in addition to nominal scales in the classification process to cover non-numerical values

A classification and clustering computation characteristics are analyzed and described. These characteristics are taken from different prospective cover size, shape, and average density. In addition to these unary features, also binary features or relations between the clusters used. These characteristics then help to identify clusters with similar characteristics, or even to identify objects. The concluded patterns may provide useful input for model-based interpretation.

Researchers mainly used RapidMiner Studio [13] as a data mining modeling and analysis tool. RapidMiner is a code-free modern analytics platform for data ingestion, data blending, predictive modeling, and deployment.

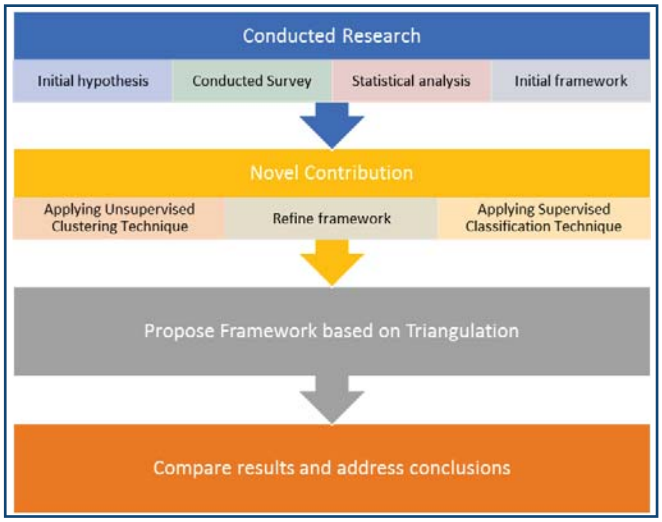

Figure 1: Research Stream Process with Triangulation

V. ANALYSIS AND RESULTS

An unsupervised clustering conducted as a first exploratory mining technique with four and five clusters.

Table 2: Cluster center of gravity (COG) – Four Clusters

Dimension	C1 (n=62)	C2 (n=78)	C3 (n=84)	C4 (n=76)
Socializing	3.37	4.15	3.58	4.36
Usefulness	2.88	3.91	3.44	4.45
Enjoyment	3.06	4.08	3.44	4.54
Attitude	2.99	4.19	3.79	4.46
Intention	2.99	4.13	3.61	4.57
Involvement	2.85	3.25	3.04	3.74
Connectedness	2.99	3.44	3.42	4.24
Learning	2.49	2.00	3.86	3.92

The clustering technique showed that cluster 4 respondents are considered as best users. The findings characterized them as the ones who have high level of perceived usefulness, perceived enjoyment, perceived connectedness and perceived involvement with SNS's. Further, this high level of perceptions has led to positive attitude followed by high level of intentions towards using SNSs in both collaborative learning as well as socializing. It is worthwhile noting that the clustering technique applied did not differentiate between undergraduate and postgraduate respondents.

Table 3: Cluster center of gravity (COG) – Five Clusters

Dimension	C1 (n=59)	C2 (n=51)	C3 (n=43)	C4 (n=70)	C5 (n=77)
Socializing	4.35	3.27	4.42	3.85	3.67
Usefulness	4.44	2.79	4.47	3.50	3.49
Enjoyment	4.52	3.03	4.62	3.60	3.50
Attitude	4.49	2.86	4.50	3.83	3.86
Intention	4.53	2.91	4.57	3.71	3.71
Involvement	3.88	2.78	3.26	3.20	3.05
Connection	4.41	2.91	3.81	3.20	3.46
Learning	4.08	2.93	2.49	1.90	3.93

As revealed no significant clusters centroids data raised from increasing number of clusters.

As shown in Figure 2 the resulted decision tree shows that the high-level using SNS's in collaborative learning mainly by learning construct, BSc, with high enjoyment, intention, while the low level using SNS's in collaborative learning Usefulness and Attitude.

To finalize the knowledge discovery process, another model is developed. It aims at identifying the extent of correlation of these features. It has as input the features extracted from the previous clustering model.

The Rules display the qualified association rules. The rule grid displays all qualified rules and their probabilities (correctness).

Table 4: Clusters general description summary

Cluster	Size of Evaluated Data	Percentage (Density)	Given Name	Description	CL Rank
Cluster 1	2,294	21%	Periodical User	Periodically use SNS's mainly for socializing (responding to others)	4
Cluster 2	2,886	26%	Socializing User	Usually use SNS's and initiating conversations mainly for socialization purposes	3
Cluster 3	3,108	28%	Frequent user	Use SNS's in common purposes including collaborative learning	2
Cluster 4	2,812	25%	High Frequent User	Use SNS's often efficiently with highest enjoyment and collaborative learning best candidate	1
Total	**11,100**	**100%**			

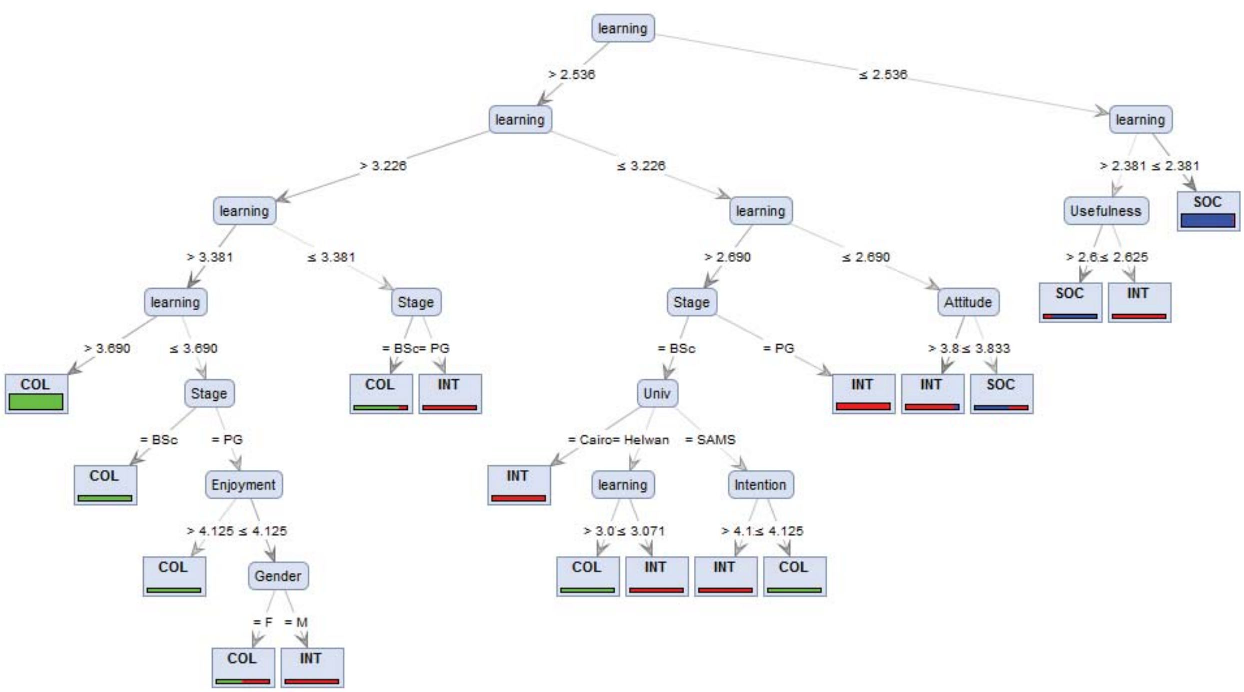

Figure 2: Decision tree induction for the classified data

The following Table 5 shows the set of rules produced by the model, which explains the power of relationship between different attributes.

The main contributors for socializing within inducted rules were mainly measured by the learning group with low values (lower than 2.38) then connection group, while the collaborative learning identified by mainly learning group with high values (higher than 3.2) then the intention group.

Table 5: Rules produced by the applied model.

	Classified as:		
Rules	Socializing	Collaborative	Neutral
if learning > 3.226 then COL	1	129	9
if learning ≤ 2.381 then SOC	83	0	1
if Univ = Cairo then INT	2	0	38
if Usefulness ≤ 3.250 then INT	1	1	10
if Usefulness > 4.375 and Intention ≤ 4.875 then INT	0	0	9
if Univ = Helwan then INT	0	2	5
if Connect > 3.167 then SOC	5	0	0
if Intention ≤ 4.375 then COL	0	3	0
else INT (0 / 0 / 0)	0	0	0

These results conforming to the decision tree results within the learning part while slightly changed in the socializing part.

VI. DISCUSSION

The clustering technique resulted in four clusters. The findings identified cluster 4 respondents as “the best user” of SNSs in both collaborative learning and socializing. Moreover, no differences were identified between undergraduate and postgraduate students.

The results drawn from the association rules technique underscored that the respondents have high level of intentions towards SNSs, which is reflected in their usage in collaborative learning. Yet, the findings did not show any significance between undergraduate and postgraduate with respect to the abovementioned results. Whereas, the Decision Tree technique emphasized that undergraduate students usage of SNS’s in collaborative learning was dependent on their level of perceived enjoyment as well as their intentions to use SNS’s.

In conclusion, the results of the three data mining techniques applied emphasized the following:

1) All three techniques underscored that intention towards SNS’s usage is positively associated with its use in collaborative learning. Likewise, partial support for this result was emphasized by Labib and Mostafa [2] that underscored statistical significant association between intention towards SNSs and collaborative learning among postgraduate students only.
2) 2) Both Decision Tree and clustering techniques identified those respondents Perceived Enjoyment is significantly related to collaborative learning.
3) Clustering and association techniques show insignificant differentiation between undergraduate and postgraduate respondents. Such result is consistent with the same study previously mentioned [2] where statistical results revealed insignificant differences between under and post graduate students in terms of collaborative learning and socializing.

VI. CONCLUSION AND FUTURE WORK

The triangulation of techniques led to reliable results.

The findings did not show significance for any of demographic attributes of respondents, or between undergraduate and postgraduate.

All three techniques addresses that SNSs’ usage is positively associated with its use in collaborative learning. The perceived enjoyment, learning, and intentions were the most significantly related constructs to collaborative learning.

Based on the previous results and discussion, a number of issues may be considered as future opportunities to be explored by interested researchers. They are the following

1. Compare between linear and non-linear techniques in the use of SNs in Collaborative learning.
2. Extend the sample used to cover larger demographic scale and to include more dimensions.
3. Use the output of the model to improve the e-learning practices used in education in MENA region.

REFERENCES

[1] K. Tarantino, J. McDonough and M. Hua, "Effects of Student Engagement with Social Media on Student Learning: A Review of Literature," *The Journal of Technology in Student Affairs*, 2013.

[2] N. Labib and R. Mostafa, "Determinants of Social Networks Usage in Collaborative Learning," in *International Conference on Communication, Management and Information Technology*, Prague, 2015.

[3] M. Adedoyin-Olowe, M. Gaber and F. Stahl, "A Survey of Data Mining Techniques for Social Network Analysis," 2014.

[4] S. E. H. D. G. Na Li, "Using Social Media for Collaborative Learning in Higher Education: A Case Study," in *5th International Conference on Advances in Computer-Human Interactions*, Valencia, Spain, 2012.

[5] P. D. J. M. K. S. V. S. Gamila Obadi, "Using Spectral Clustering for Finding Students' Patterns of Behavior in Social Networks," in *DATESO*, 2010.

[6] M. T. a. O. R. Z. Reihaneh Rabbany k., "Social Network Analysis and Mining to Support the Assessment of Online Student Participation," vol. 13, no. 2, 2011.

[7] D. Shin, "Analysis of online social networks: A cross national study," *Online Information Review*, vol. 34, no. 3, pp. 473-495, 2010.

[8] F. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319-40, 1989.

[9] H. Nysveen, P. Pedersen and H. Thorbjornsen, "Intentions to use mobile services: antecedents and cross-service comparisons," *Journal of the Academy of Marketing Science*, vol. 33, no. 3, pp. 330-46, 2005.

[10] G. Saw, W. Abbott, J. Donaghey and C. McDonald, "Social media for international students, it's not all about Facebook," *Library Management,* vol. 34, no. 3, pp. 156-174, 2013.

[11] D. Li, "Online social network acceptance: a social perspective," *Internet Research,* vol. 21, no. 5, pp. 562-580, 2011.

[12] J. J. ,. A. R. T. R. B. W. Hair, "Multivariate Data Analysis," 1998.

[13] RapidMiner, "RapidMiner Studio Manual," RapidMiner, London, 2014.