Asian Journal of Science and Applied Technology
ISSN: 2249-0698 (P) Vol.10 No.1, 2021, pp.41-45
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ajsat-2021.10.1.2817

Association Rules Based Opinion Mining for E-Learning and Electronic Literature Reading

Shubham Dubey[1] and Ispany Marton[2]
[1&2]Faculty of Informatics, University of Debrecen, Hungary
E-mail: shubhamdubey1312@gmail.com, ispanymartn@inf.unideb.hu

(Received 12 April 2021; Revised 7 May 2021; Accepted 20 May 2021; Available online 27 May 2021)

***Abstract* - The revolution in Information technology has transformed our academics and projected it in the new dimension. ICT supported learning is best facilitating learners' needs. With some pros there are some cons as well which are affiliated with e-learning. This study is analyzing the opinions of respondents towards e-learning and electronic literature reading. To analyze the response authors employed Apriori algorithm of association rule mining approach. The support and confidence values or the analysis were kept high (≥85) to achieve good quality of association among such factors. Study finds stress generated while adopting e-learning and e-reading has higher association with content availability in web resources. This very inferable that the wide content range results more surfing and searching and resulting stress. Study also found that the distractions caused by resource divergence are strongly associated with stress and problems faced in eyes. The study is giving significant contribution towards policy making and analysis of factors that affects the learners' choices and problems the most.**

***Keywords:* ICT Supported Learning, Apriori Algorithms, Association Rule, Content Availability, Resource Divergence, Electronic Reading**

I. INTRODUCTION

The world is changing rapidly so is academics. There are several influencing factors which have had significant impact on pushing the classroom learning towards virtual mode of leering. Several researchers have identified the driving factors which motivates the learner to follow virtual mode of learning. The changing need of learners and innovation have given a new definition to the academic. ICT supported learning has dome tremendously well in the past half of the decade across the world (Ebner, M., 2007). Considering the advantages in one hand none can deny with the hitches and problems affiliated with and because of ICT supported learning. The problems emerged when implementing electronic learning are related to resources, technical knowledge, monotonic class environment, eyes stress, mental stress generated by screens etc (Chua, B. B., & Dyson, L. E., 2004, December). These issues must be treated well so that the feasibility and acceptance of ICT supported learning can be maximized. Educational data mining has become very popular domain in the recent past (Romero, C., & Ventura, S, 2010; Romero, C., *et al.,* 2010). Several analytics generated by learning management systems give crucial and accurate information about the teachers and students (Castro, F. *et al.,* 2007). Because of these facilities the interest and motivation of learners can be assessed timely. This definitely reduces the scope of dropouts. Various analytical methods are used to infer the knowledge in from the raw data; Association rule mining is one of them. This discusses the level of association among several item sets (Yabing, J., 2013). The item set is basically the set of actors which are taken in consideration during analysis. Apriority algorithm is one of the famous algorithms, which uses the frequent item set to generate the association rules for certain range of support and confidence values (Singh, J *et al.,* 2013). This study is applying Apriority algorithm to mine useful information and factors affiliated with e-learning and electronic literature reading.

II. LITERATURE REVIEW

The demands of current academic era have catalyzed the e-learning allot. This transformation has given an objective for innovation and scope of betterment in academics. Of course there are some goods and bads in every innovation so is with ICT supported learning as well. Around 129 studies were classified on the basis of learning styles. Fuzzy network model, Neural network model and association mining and Bayesian model were used to classify problems in e-learning (Khamparia, A. *et al.,* 2020). Popup issues, distraction, technical and some physical problems were also identified as crucial factors affecting learners' motivation (Dubey, S., & Piroska, B., 2019). From previous research works, one can conclude that e-learning is not appropriate for the learners without the self-discipline. The learners need more training as teachers to achieve excellence.There are other difficulties as academic honesty, punctuality unauthenticated resources etc (Wong, D, 2007, McPherson, M. A. *et al.,* 2008). In other hand there are some goods as well of e-learning like content availability, anytime any where execution of study, flexibility of time, variety of resources, polymorphic content and other audio-video advantages (Arkorful, V., & Abaidoo, N. 2015). Regardless all pros and cons none can deny with the acceptance of e-learning; in fact, the future of learning will be virtual learning (Welsh, E. T. *et al.,* 2003). In a study the factors affecting readers' motivation and quality of e-library are associated together using association rules mining method (Dubey, S., Sharma, P., & Mária, B., 2020). If we will be able to find the relation between the factors which are dominantly affecting the electronic literature reading and e-learning, then definitely we would be able to solve them efficiently. This article is identifying the association among such actors by applying Association rule mining approach of data mining.

III. METHODOLOGY

A. Data Collection and Research Setup

Present study analyzing the opinions of e-learners and e-literature readers according to their perception and preference about E-literature. The data was collected through survey using google form and had 12 questions. The questions were choice based (close end). The URL of google form was spread and it was open for response for 10 Days (Feb. 26 to March 6, 2020). Participation in the study was completely anonymous and by choice of participants, they were requested to share their contribution in research. 110 Respondents were asked about their experience with both electronic reading and e-learning. Considering motivation behind the reading to hurdles in continuation of the electronic reading (Dubey, S., Piroska, B. 2019). The factors are also covering the suggestions about the time of electronic reading, the stress made because of the electronic literature reading, frequency of following e-reading in a single day. Study is done using explanatory framework as it is guided by research questions.

B. Apriori Algorithm of Association Rule Mining

Association rules mining is a data mining method that utilizesnumber of occurrence of itemsets in established or executed set of transactions and generates the association rules among interested frequent itemsets (Han, J., Kamber, M., & Pei, J., 2011). During the process of Association rule mining Confidence and Support are two main parameters which play crucial role. Support denotes the extent to which the relationship exists in the data or relative percentage of together occurrence of an item in dataset. In other hand confidence represents the likely chance of two or more items that have a trend to occur together in a set of transactions. One can say, it is the ratio of the frequency of togetherness over the identical frequency of the picked item. Suppose we have a set of four transitions which are having {A, B, C, D, E, F} item sets here set of transactions are below.

R1: A, B, C
R2: A, C
R3: A, D
R4: B, E, F

In Table I, Row first is the transaction number and the second row represents the transactions and items appeared in it.

TABLE I FREQUENCY TABLE FOR ITEMS

Item	{A}	{B}	{C}	{D}	{E}	{F}
Frequency	3	2	2	1	1	1

1. *Support:* The proportion at which the items occurred together in the total number of transactions. We denote it by ‘s’.
   For example, s(A, C) = (Occurrence of A and C together) / (Total number of Transactions) 2/4=0.5
2. *Confidence:* Confidence for an association A→C will be the ratio of occurrence of A and C together over the Occurrence of A. It is relative quantity and is denoted by ‘c’.

Here c(A→C) = (Occurrence of A and C together) / (Occurrence of A) ⇒ 2/3 = 66.66

If the necessary condition for rules generation is to keep support 50% at least and confidence also must be 50% at least then obligatory minimum frequency for filtering will be derived from the formula below:

(Total transaction * (Minimum support value / 100) = 4*(50/100) = 2

Hence a revised table III will be obtained which has only item sets having at least 2 frequencies.

TABLE II REVISED FREQUENCY TABLE

Item	{A}	{B}	{C}
Frequency	3	2	2

Later than finding individual frequency up to the minimum support level there is a need of paringthe items with similar minimum support at least. Refer table IV.

TABLE III PAIR OF ITEM SETS WITH FREQUENCY

Item	{A,B}	{A,C}	{B,C}
Frequency	1	2	1

Considering the table III it is desired that the least frequency should be 2 or more than that. Which means we just consider {A, C} item set in the subsequent iteration and ignore {A, B} and (B, C}in table IV.

TABLE IV PAIR WITH MINIMUM SUPPORT

Item	{A,C}
Frequency	2

This is the optimal level of the table here association rules will be calculated. Table V is demonstrating the quantified relation between item sets.

TABLE V ASSOCIATION RULES GENERATED FOR THE GIVEN ITEM SET

Rule	**Support**	**Confidence**	**Percentage of Confidence**
A→C	2	2/3=0.66	66%
C→A	2	2/2=1	100%

In the first column rule between items is given the first row is for A implies C for which support is 2 but since the individual frequency of P is 3 hence confidence is 66%. In the second column, both A and C occurred twice and C individually occurs 2 times. Thus the Percentage of confidence for Rule C→A is 100%.

C. Analysis

The analysis was done in two stages; first is the preprocessing of the data gathered from respondents. Second phase is the association rules mining, which is actually post phase of analysis of the preprocessed data.

1. Pre-Processing and Modeling of Data

The Factors (F's) are transformed as questions in the survey as mentioned below:

F1 Do you follow e-learning/ reading more than 6 hours daily?
F2 Does electronic resource pool has wider content range?
F3 popups makes you struggle while following electronic reading?
F4 E-learning and electronic reading suffer with the problem of not verified content?
F5 Does E-reading facilitate your study habits?
F6 Does electronic reading has more distractions?
F7 Are you not able to continue reading for long time while following E-reading?
F8 Does screen affect stress level in electronic reading?
F9 Do you think electronic reading generates stress?
F10 Do you think 1 you need a teacher's explanation while electronic reading?
F11 Do you feel that E-learning or E-content have high retention/attention-grabbing?
F12 Are you satisfied with your E- learning content?

These all the questions had three level Likert scale possible choices of answers which are

A1 Very True
A2 Somewhat True
A3 Not True

Here during the pre-processing stage of data these options were replaced with quantitative values. Replacement scheme was as: 1 if Very True or Somewhat True and 0 if not true. Then if there was a '1' in a column we have replaced the 1 by the column name. Keep in consideration that the column names are F1, F2, F3…..…F12. Hence for example if in a row the first column is of F1 and any candidate chooses true or somewhat true in F1 then these very true has been replaced by 1 and again this one is replaced with F1. Similar replacements were taken place for F2, F3....F12 and for Candidate 1 to 110. All 'Not True/s' are replaced with 0 and then by a blank as table VI.

TABLE VI TRANSFORMED DATA COLLECTED FROM SURVEY

R1		F2	F3	F4	F5	F6	F7	F8	F9	F1		
R2		F2	F3	F4	F5	F6	F7	F8	F9	F1		F12
R3		F2	F3	F4	F5	F6		F8	F9	F1	F11	F12
.		F2	F3		F5	F6		F8	F9		F11	F12
.	F1	F2	F3	F4	F5	F6		F8		F1	F11	F12
.	F1	F2	F3	F4	F5	F6	F7	F8	F9	F1	F11	F12
.		F2	F3	F4		F6	F7	F8	F9	F1	F11	
.		F2		F4	F5	F6		F8	F9	F1	F11	F12
.												
R109		F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
R109		F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12

Here R1, R2, R3...R93 are participants of survey and F1, F2...F11 are factors related to e-learning taken in consideration for research. In order to generate association rules, there is a need for saving this transformed data as CSV file.

2. Finding Association Rules

Emphasis is to find the association among the factors using opinions. In the study, the minimum value of support was kept 80%. This indicates that any item set having its frequency more than 88 is selected for analysis. In the benchmark data set F1, F3, F4, F7, F10, F11, F12 have their frequency 19, 77, 73, 64, 69, 66 and 79 respectively, which is less than 80%. Another measure is confidence and its threshold value is also 80%. It means two or more items with joint frequency 88 or more than that are participating in the analysis. If the manual explanation of the work is discussed, then it can be treated as an iterative approach for rules extraction. In the first iteration of threshold mining, 7 items were ignored. F1, F3, F4, F7, F10, F11 and F12 have individual frequencies less than 88. Thus other {F2, F5, F6, F8, F9} are legitimatefrequent item set and are passed to the second iteration. In the second iteration mining of item, pairs will be done hence all pairs of item sets are checked for optimal value of support. Such item sets areas Pair ≈ (Fi,Fj) where i={1,2,3,4,5,6,7,8,9,10,11,12}–{1,3,4,7,10,11,12} and j={1,2,3,4,5,6,7,8,9,10,11,12} – {1,3,4,7,10,11,12}. Thus using this manual approach one can get all possible combinations of frequent item set and can get optimal frequency combinations of item sets. Generation of association rules was done by executing software named as apriori. This was executed for several combinations of support and confidence as figure 1.

```
E:\My PhD\Research During PhD\electronic reading assosiation rules\Apriori>apriori -tr -c85 -s85 input.txt 8585.txt
apriori - find frequent item sets with the apriori algorithm
version 6.22 (2016.04.09)        (c) 1996-2016   Christian Borgelt
reading input.txt ... [11 item(s), 102 transaction(s)] done [0.00s].
building transaction tree ... [52 node(s)] done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing 8585.txt ... [22 rule(s)] done [0.00s].

E:\My PhD\Research During PhD\electronic reading assosiation rules\Apriori>apriori -tr -c90 -s85 input.txt 9085.txt
apriori - find frequent item sets with the apriori algorithm
version 6.22 (2016.04.09)        (c) 1996-2016   Christian Borgelt
reading input.txt ... [11 item(s), 102 transaction(s)] done [0.01s].
building transaction tree ... [42 node(s)] done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing 9085.txt ... [15 rule(s)] done [0.00s].

E:\My PhD\Research During PhD\electronic reading assosiation rules\Apriori>apriori -tr -c95 -s85 input.txt 9585.txt
apriori - find frequent item sets with the apriori algorithm
version 6.22 (2016.04.09)        (c) 1996-2016   Christian Borgelt
reading input.txt ... [11 item(s), 102 transaction(s)] done [0.00s].
building transaction tree ... [20 node(s)] done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing 9585.txt ... [4 rule(s)] done [0.00s].
```

Fig. 1 Demonstration of Association rule generation

One can see in figure 1 for support 85 and confidence 85 the value of association rules generated is 22. For the value of support 85 and confidence 90 we have 15 rules and then for s=85 and c=95 we have 4 rules. This is how the APRIORI tool extracts association rules. With minimum support 85% and confidence 85% the analysis took place.

IV. RESULTS AND DISCUSSION

The rules derived from the set of data shows the association among factors that affects e-learning and electronic literature reading. Number of association rules generated for certain value of confidence and support are mentioned below in table VII.

TABLE VII NUMBER OF ASSOCIATION RULES DERIVED FOR THE RANGE OF CONFIDENCE AND SUPPORT

Confidence	80	85	90	95	100	80	85	90	95	80	85	90	95	80	85	90	95
Support	80	80	80	80	80	85	85	85	85	90	90	90	90	95	95	95	95
Number of Rules	50	30	21	7	0	34	22	15	4	18	13	9	2	9	7	5	1

Here purpose is to optimize the value of Confidence and Support and achieve as many as possible rules. Table VII, shows a range of 'c' and 's' both are ranging 80≤ c and s≤100. Here values of 'c' and 's' explain the quality of association rules. Consideration of all association rules from Table VII will generate needless and irrelevant sets of association rules. Thus to avoid association rules explosion researchers have selected higher values of Support and Confidence. Here in Table V all the rules having S ≥ 90 and C ≥ 85 are taken for further analysis.

TABLE VIII SET OF MINED VALID ASSOCIATION RULES

Association Rules
F9 →F6 (85.2941, 93.1034)
F2 →F6 (85.2941, 94.2529)
F8 →F6 (85.2941, 95.4023)
F8 →F9 F2 (85.2941, 95.4023)
F2 →F9 F8 (86.2745, 94.3182)
F9 →F2 F8 (89.2157, 91.2088)
F2 →F9 (90.1961, 94.5652)
F9 →F2 (94.1176, 90.625)
F8 →F9 (90.1961, 95.6522)
F9 →F8 (95.098, 90.7216)
F8 →F2 (94.1176, 94.7917)
F2 →F8 (95.098, 93.8144)

Observation of table VIII gives us idea about derived association rules along with their support as well as confidence greater or equals to 85. In F8→F2, F2 and F8 share the highest association with 94.1176 support value and 94.7917 confidence values in case of reverse association the value of support and confidence are 95.098 and 93.8144. For the given range of confidence and support the least associated item sets are F9 →F6. One can observer that F8 →F9 F2, F2 →F9 F8 and F9 →F2 F8 have two elements of in the right side which means item appearing in the left have association with the frequent item setof pairs. In F8 →F9 F2, F8 has association with F9 F2 with 85.2941 support and 95.4023 confidence. The interpretation of the results is discussed in the conclusion section.

V. CONCLUSION

The study has analyzed learners' opinion about e-learning and electronic literature reading. The present work confirmed the findings about how the factors affecting e-learning and e-literature readings are associated together. Study results from table I shows that the wider content availability of study content (F2) generates stress (F8). Also the inverse relation for the same is second highest association noticed in analysis. Also the findings support the statement that the electronic content reading affects stress which is widely inferred in previous studies. Also we can say that wider content range generates stress because of more surfing in web and that makes us keep using screen for a long. This is confirmed by F2 →F9 F8 (86.2745, 94.3182). Distraction generated through reading and e-learning has good association with stress generated (F8 →F6). Similarly wide content availability has good association with distraction factors (F2 →F6).The study gives significant results about the factors affecting learners' and readers' motivation towards electronic reading and electronic learning. The association among factors are quantified and discussed in brief. The study will be helpful for the future researchers and policy makers who are planning to achieve significant transformation in e-academic.

REFERENCES

[1] A. Khamparia and B. Pandey, "Association of learning styles with different e-learning problems: A Systematic Review and classification", Education and Information Technologies, Vol. 25, No. 2, pp. 1303-1331, 2020.

[2] S. Dubey and B. Piroska, "Exploration of Factors Affecting Learners' Motivation in E-learning", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* Vol. 5, No. 2, pp. 1269-1275, 2019.

[3] M. Ebner, "E-Learning 2.0= e-Learning 1.0+ Web 2.0?", In The Second International Conference on Availability, Reliability and Security (ARES'07), *IEEE,* pp. 1235-1239, April 2007.

[4] D. Wong, "A critical literature review on e-learning limitations", *Journal for the Advancement of Science and Arts,* Vol. 2, No. 1, pp. 55-62, 2007.

[5] F. Castro, A.Vellido, A. Nebot and F. Mugica, "Applying data mining techniques to e-learning problems" In Evolution of teaching and learning paradigms in intelligent environment, *Springer,* Berlin, Heidelberg, pp. 183-221, 2007.

[6] E. T. Welsh, C. R. Wanberg, K. G. Brown and M. J. Simmering, "E-learning: emerging uses, empirical results and future directions", *International Journal of Training and Development,* Vol. 7, No. 4, pp. 245-258, 2003.

[7] V. Arkorful and N. Abaidoo, "The role of e-learning, advantages and disadvantages of its adoption in higher education", *International Journal of Instructional Technology and Distance Learning,* Vol. 12, No. 1, pp. 29-42, 2015.

[8] S. Dubey, P. Sharma and B. Mária, "Quality assurance of national digital library of University of Debrecen: An extensive association rules based study using ISO/IEC25010", *COLLNET Journal of Scientometrics and Information Management,* Vol. 14, No. 1, pp. 119-133, 2020.

[9] J. Han, M. Kamber and J. Pei, "Data mining concepts and techniques third edition", *The Morgan Kaufmann Series in Data Management Systems,* Vol. 5, No. 4, pp. 83-124, 2011.

[10] J. Yabing, "Research of an improved apriori algorithm in data mining association rules", *International Journal of Computer and Communication Engineering,* Vol. 2, No. 1, pp. 25, 2013.

[11] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* Vol. 40, No. 6, pp. 601-618, 2010.

[12] C. Romero, S. Ventura, M. Pechenizkiy and R. S. Baker, (Eds.). *Handbook of educational data mining.* CRC press, 2010.

[13] B. B. Chua and L. E. Dyson, "Applying the ISO 9126 model to the evaluation of an e-learning system", *In Proc. of ASCILITE,* pp. 5-8, December 2004.

[14] M. A. McPherson and J. M. Nunes, "Critical issues for e-learning delivery: what may seem obvious is not always put into practice", *Journal of computer assisted learning,* Vol. 24, No. 5, pp. 433-445, 2008.

[15] J. Singh, H. Rama and D. J. Sodhi, "Improving efficiency of apriori algorithm using transaction reduction", *International Journal of Scientific and Research Publications,* Vol. 3, No. 1, pp. 1-4, 2013.