A NOVEL APPROACH TO MAKING EFFICIENT E-LEARNING PROCESS IN THE USING BIG DATA

Renu[1], Girish Kumar[2] and Kanika Sharma[3]

Research Scholar Computer Science and Engineering
Lovely Professional University, Phagwara, India.
renudhiman2725@gmail.com
Assistant Professor, School of Computer Applications
Lovely Professional University, Phagwara, India.
girishvansh@gmail.com
Assistant Professor, School of Computer Applications
Lovely Professional University, Phagwara, India.
kanikasharma2177@gmail.com

ABSTRACT- *E-learning and Big Data are correlated with each other. Online learning is E-learning. As the quantity of data increases day by day and it becomes huge and giant. A few years ago,data was measured in gigabytes then in terabytes and nowadays it is measured in dozens of petabytes.There is a bombardment of Different types of data like Facebook, Twitter, biological data, industrialdata, the Internet of things, mobile phones, cameras, audio, video, educational data, and sensor data. Such type of data sets is coming under one name is known as Big Data. In our learning system, there are many topics and subjects, which are not readable and students are not interested to studysuch type of topics in a particular subject. In this paper, there is a description of an improved Apriori algorithm with the help of which, frequently data itemsets can find out, by using the Association RuleMining algorithm, within less time rather than Existing method. With the help of getting frequentitemsets, can get some output in which topics are searched and studied by users and students. As a result, universities and learning systems can apply these topics in the syllabus.*

KEYWORDS*: -* *E-learning, Big Data, Data Mining, Association rule mining, Apriori.*
DOI: 10.48047/ecb/2023.12.si8.597

1.0 INTRODUCTION

Before the internet era learning was done in traditional learning. One of the aims of this exploration is to integrate the concept of "assessment for learning" into a learning technology to better fit student learning capabilities [2]. Furthermore, recognizing and reacting to learners' preferred delivery styles to improve student performance and increase their engagement in learning materials is another aim of this study. Before the internet era, there were many types of distance education modeslike programs on televisions recorded videos/audio or manuals. Instructors were solving the problems by phone or email [1]. Now system is dramatically changed by technology. Most institution provides online courses. As result the universities provide offer of b-learning (hybrid classroom and online learning) and e-learning (pure online learning). There are many positive aspects of e-learning, like good communication between teachers and students. Students can enhance their learning skills. Test evaluations, technical skill development, and individual and collaborative activities are promoted by e-learning. It reduces gap between theory and practical. There are many e-learning management systems, like the VARK model**.** The full form of VARK is visual(v), aural (a), read/write(r), and kinaesthetic (k). This model provides a good learning method to the students. This is an adaptive tutoring system used to teach the Java programming language. Iweaver is an adaptive tutoring system used to teach the Java programming language. Research

and Practice in Technology Enhanced Learning styles in an adaptive e-learning Environment Protus is used for teaching the Java programming language. AEHS-LS is an adaptive e-learning system. It is used to teach the java scripting language. Learning Management System (LMS) is a way to provide e-learning for example Moodle, Sakai, ILIAS and Newton and Dream Box learning platforms. Educational data mining (EDM) and learning analytics (LA) are two research fields to analyze this data. The data mining is process for knowledge discovery by extracting the meaningful information and patterns from big data. Association rule mining is one of the techniques used in data mining. It is used for uncovering relationship in data and frequent item set data mining. As the size of big data is increases day to day, as well as need of highly efficient and highly scalable data mining tools is increases. Big data includes multidimensional and multilevel data. There are many parallel and distributed approaches that can use for low dimensional data but can't give better result for high dimensional data. So, by applying association rule detection technique of Map Reduce, which is based on interestingness measures. Get a new technique for easy and efficient transaction of data. This approach can enhance efficiency of e-learning process in future using big data. Let us go through all terms regarding this approach.

1.2 BIG DATA

As quantity of data increases day by day and it becomes huge and giant. Few years ago, data was measured in gigabytes then in terabytes and now days it measured in dozens of petabytes. There is bombardment of Different types of data like Face book, Twitter, biological data, industrial data, internet of things, mobile phones, camera, audio, video, educational data, and sensor data [3]. Such type of data sets is coming under one name is known as Big Data, that is very large in size and complex. The term big data first time was used 1990 and credit was given to John Mashy to make it popular. The traditional data processing applications software are used to deal with big data. Many tools can be applied on big data to capture, manage, and process data. There are many challenges for big data such as data capturing, data storage, data processing, data analysis, search, transition, sharing, transfer, visualization, query, updating data and security. It is necessary to explain 7Vs for description of big data because introduction of big data is incomplete without explanation of 7v's. these 7v's of big data are volume, velocity, variety, veracity, veracity, visualisation, vision. Characteristics of big data can be defined on the basis of these 7v's [4].

1.3 DATA MINING

Data Mining is the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable pattern in data with the wide use of databases and the explosive growth in their sizes. Data mining refers to extracting or "mining" knowledge from large amounts of data. Data mining is the search for the relationships and global patterns that exist in large databases but are hidden among large amounts of data [5]. The essential process of Knowledge Discovery is the conversion of data into knowledge in order to aid in decision-making, referred to as data mining. The knowledge Discovery process consists of an iterative sequence of data cleaning, data integration, data selection, data mining pattern recognition, and knowledge presentation. Data mining functionsinclude clustering, classification, prediction, and associations. One of the most important data mining applications is that of mining association rules. Association rules, first introduced in 1993,are used to identify relationships among a set of items in databases [6].

APRIORI ALGORITHM

The Apriori algorithm is a classical algorithm in data mining. It is used for mining frequent item sets and relevant association rules. It is devised to operate on data groups of candidatesare tested against the data [7]. The algorithm terminates when no further successful extensions are found. Apriori algorithm and hash tree structure to count candidate itemset containing a lot

of transactions, for instance, items brought by customers in a store [8]. It is very important for effective Market Basket Analysis and it helps the customers in purchasing their itemswith more ease which increases the sales of the markets. It has also been used in the field of healthcare for the detection of adverse drug reactions.

- **Support and confidence**

The support of an itemset, is the proportion of transaction in the database in which the item X appears. It signifies the popularity of an itemset. Confidence is an indication of how often the rule has been found to be true [9].

- **Association rules**

Association rule learning is a prominent and a well-explored method for determining relations among variables in large databases. When we go grocery shopping, we often have a standard list of things to buy. Each shopper has a distinctive list, depending on one's needs and preferences [10].

2.0 METHODOLOGY

Classical Apriori algorithm of association rules is proposed. Apriori algorithm is a step-wise search, applied to search for the next item set from the previous item set. The L1 is the frequent-1 item set which is used to compute the next frequent-2 item set which is L2. Thus, LK is used until it cannot find LK-1. The proposed method will use the transposition of the dataset and then find out the frequent item sets using Apriori algorithm steps. The Apriori property is available which includes that all sub-item sets having items, in frequent item sets must also be frequent. The algorithm acquires two steps to find out the frequent items and association rules. In the proposed technique, one new value will define, called as the items present in the transition. Improvement will be done in the classical algorithm. To do that, this technique will be further enhanced by modifying the minimum support calculation formula in the transpose technique. When in any particular transition the number of itemsbelow the threshold value then from the dataset that transition will be deleted. This approach will reduce the scans over the database, which will further reduce the time to generate the rules.

2.1 ALGORITHM FOR PROPOSED TECHNIQUE

Parameters: Set up the following parameters to reach the objective:

- Dataset is required to implement the new technique.
- Minimum support calculation formula is used in a different manner.
- Apriori property is considered to generate the frequent item sets.

Transpose of Dataset: Dataset is transposed; thus, implementation is done by applying algorithmon transactions in place of items (As in Apriori algorithm items are used to get frequent results). L1= (x0, y0, x1, y1)

Transpose (0, 0, M, M)

Where, M denotes matrix.

Minimum support calculation formula: To calculate the support for each candidate item set, minimum support calculation formula is used.

Finding the frequent item sets: Use new minimum support calculation by dividing the average number of transactions with total number of transactions. Then, apply the apriori algorithm on the transposed database.

LI =find_frequent_I-itemsets(D);

Ck = apriori _gen (Lk-l, minsup), // generation of frequent item sets using Apriori

```
for each transaction tED {
C1 = subset (Ck, t);
candidates
for each candidate C E C1
c.count+ +; //increment the count of all candidates
 }
Return L1;
```

3.0 IMPLEMENTATION

Implementation of proposed method is done in Python programming language. Python is an interpreted high-level programming language. Anaconda Navigator application is used to write the implementation code. In Anaconda Navigator, Spyder 3.2.4 development environment is selected to read the dataset attributes. In first phase of proposed method, generate data sets. Data sets can be of any type like academic dataset, market dataset, dataset of corporate sector. With the help of proposed method, Association rule can be generated in less time rather than existing method.
in case of existing method, there is one item per iteration.

4.0 RESULTS AND DISCUSSION

The proposed algorithm is based on E-Learning dataset. As per my proposed topic entitled with "A novel approach to make efficient E- Learning process in future using big data" this is clearly mentioned that it comes under the Big Data domain. There are so many scholars have been published their research in this field but still there is a lot of work pending. As we know this is very use full for the high dimensional dataset, some researchers have defined that this is area of research for the future scope. Till the time most of the work have been done on the low dimensional dataset. No doubt there are some researcher who had revealed their work on high dimensional data set as well, but the respective results have been taken in maximum time. Proposed dissertation work is picked from this idea to solve the problem for high dimensional dataset with high performance and within deduced time. The outcomes will be produce the better accuracy and reduce the access time. in figure 4.1, there is execution of existing method. Apriori algorithm is used for execution. There is class file, variable explorer, file explorer, association rule generation of given data and execution time is shown.

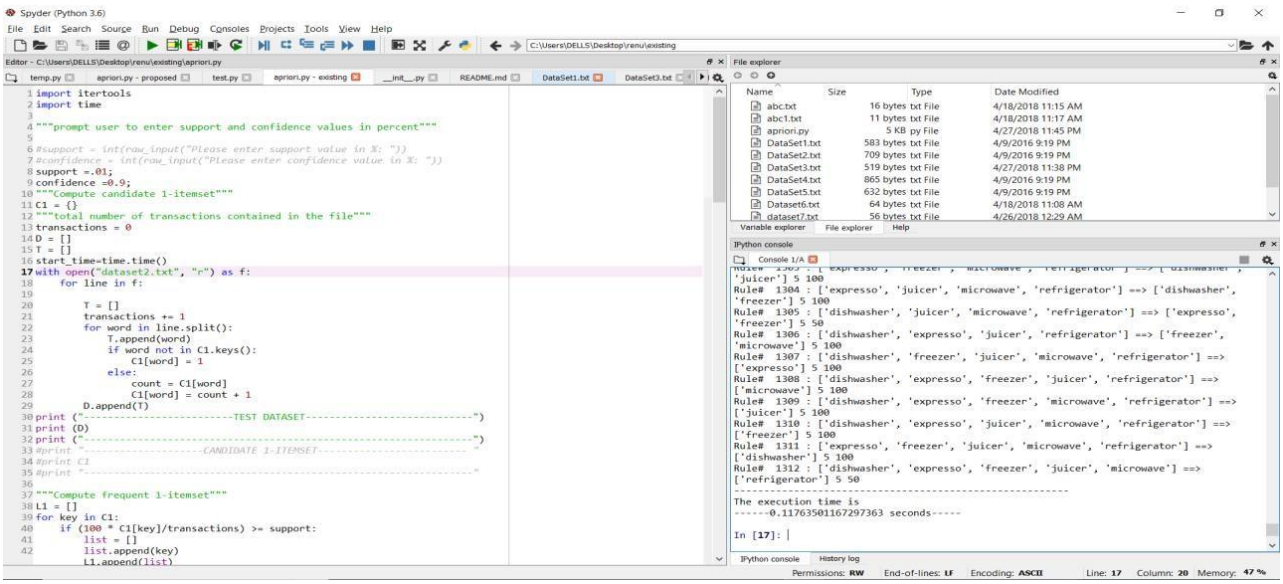

Figure 4.1: Implementation of existing method

Figure-4.2, shows the association rule generation, with the help of which we can find the frequent itemset in existing method. In this case execution time is 0.117 seconds.

Figure 4.2: Association rules and execution time of existing method

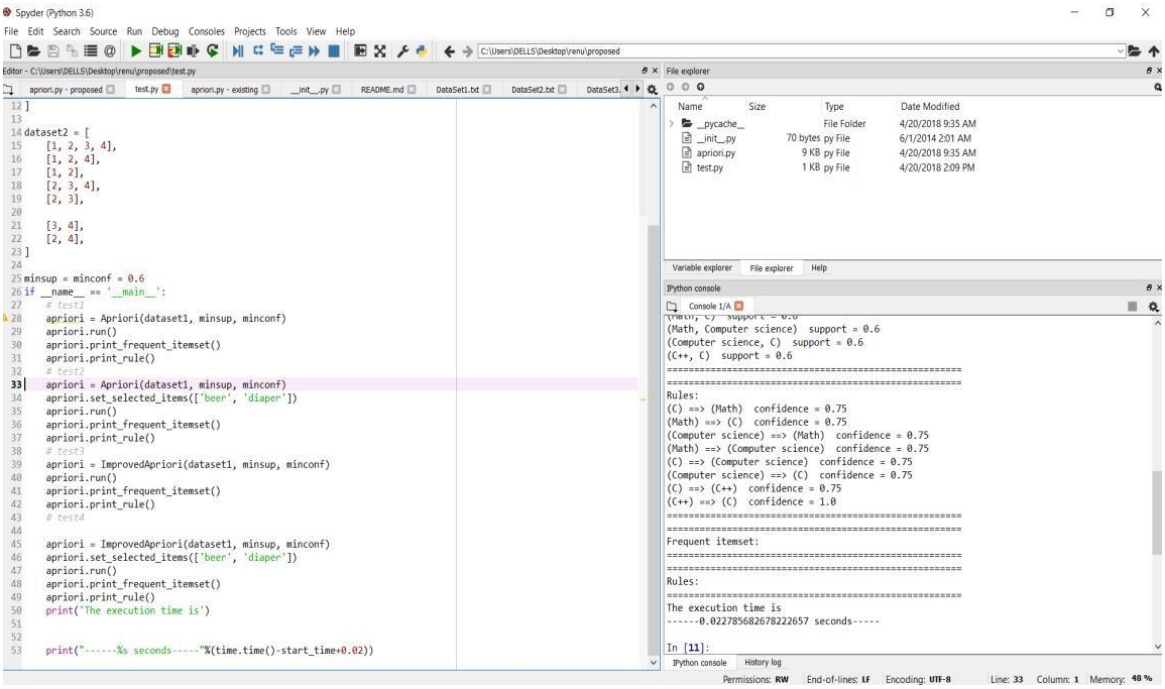

Figure 4.3: Implementation of proposed method

Figure 4.3, shows the execution of apriori.py file of proposed method. Apriori.py file contain the definition of class and code for execution.

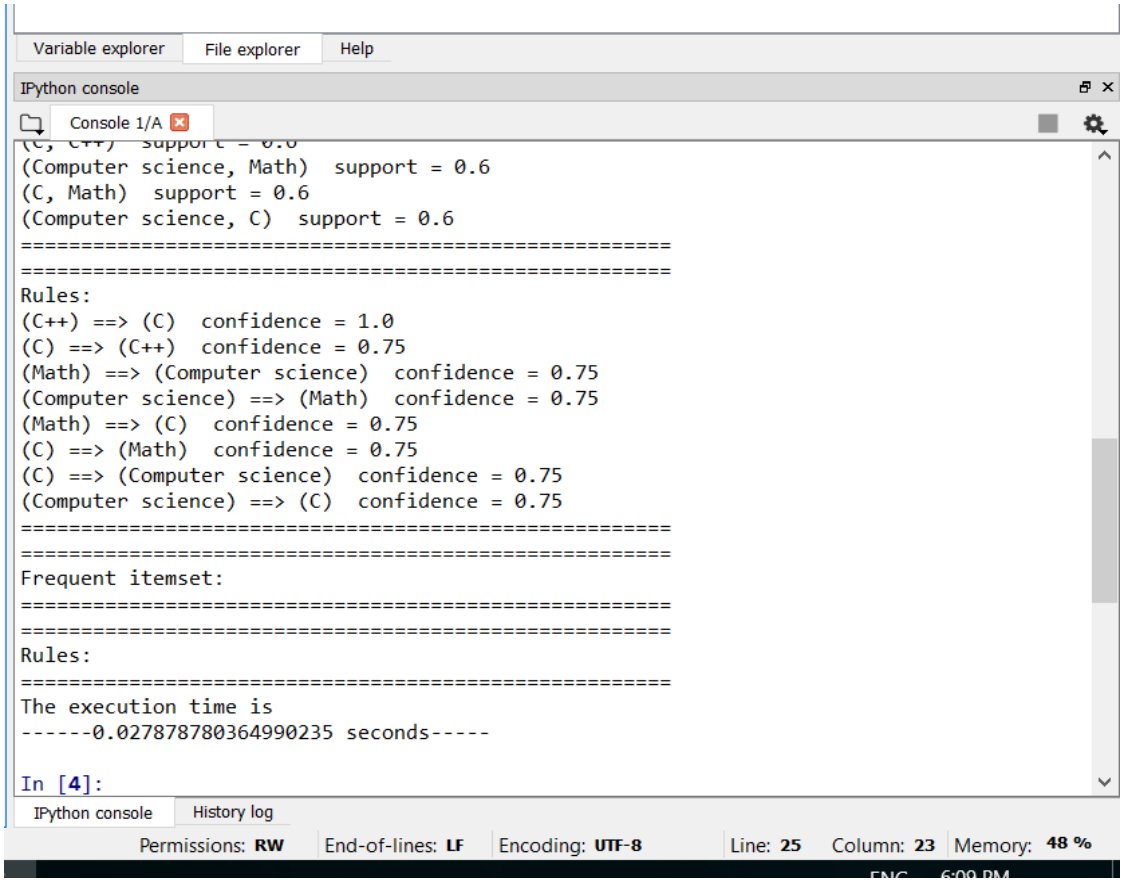

Figure 4.4: Association rule and execution time of proposed method

5.0 COMPARISONS AND IMPROVEMENT

Comparisons of present study with the base paper is shown below;

Table 5.1: Comparison table on different data sets

Different datasets	Execution time of existing method (In seconds)	Execution time of proposed method (In seconds)
Dataset 1	0.117	0.027
Dataset 2	0.102	0.024
Dataset 3	0.181	0.025
Dataset 4	0.242	0.028

Table 5.1 shows the comparison table on different data sets. Table shown four types of different datasets like dataset1, dataset 2, dataset 3, dataset 4. These datasets are different in terms of types of items and numbers of items.

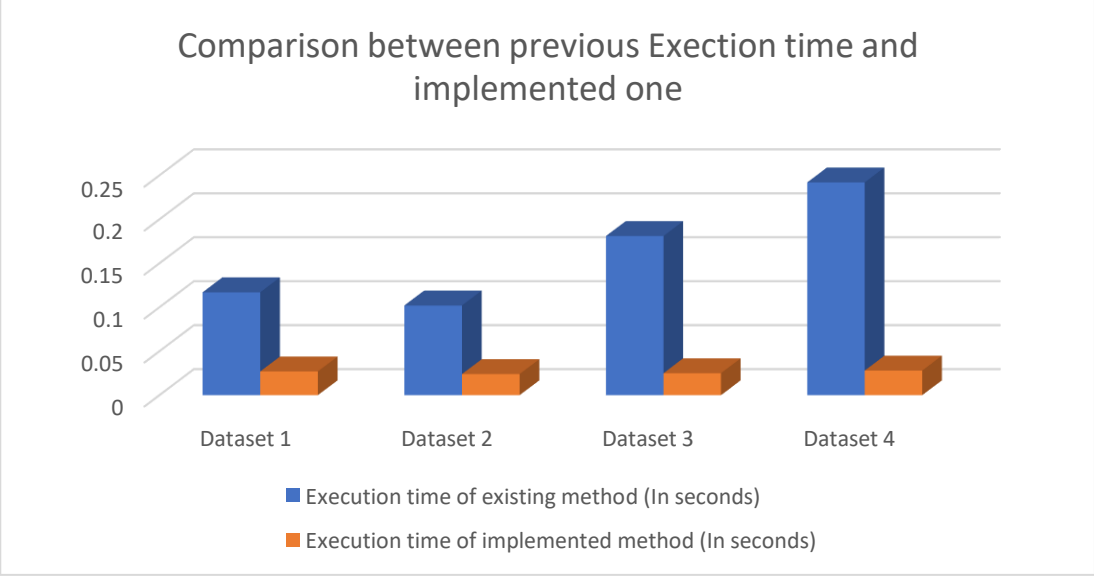

Figure 5.1: Execution time comparison

Figure 5.1, shows the comparison between previous execution time and implemented one. This graph is generated on the basis of values shown in table 6.1.

6.0 CONCLUSION

Educational institutions are generating huge volumes of data through the admission process, evaluation and teaching learning. The field of education is gaining insight and is obtaining actionable data from large chunks of varied data known as Big Data. With the advent of e- learning provide in many universities, the amount of data available to all the stake holders of the educational system is enormous. The Big Data paradigms are needed in current world to add value to the processes of educational institutions. The motive of this proposed method is to make better e-learning management system by using big data and apriori algorithm. Proposed method helps to generate association rules to pick frequent itemset. This method takes less time to generate association rules rather than existing system. So, in e-learning management system, proposed method helps to take less time, to pick a frequent itemset from big data.

7.0 FUTURE SCOPE

As the big data is very complex then proposed method can help to generate association rules in less time with the help of association rule mining algorithm. Execution time can be reduced in future. Accuracy level can be increased by using some of the algorithms of data mining. Execution time is also reduced which helps the individual to save their time.

REFERENCES

[1] R. M. Ijtihadie, J.-M. Poggi, C. Tuleau-Malot, and N. Villa-Vialaneix, “Random Forests for Big Data,” Big Data., vol. 1, pp. 1–19, 2013

[2] A. Ashraf, “Incremental development & revolutions of E-learning software systems in education sector: a case study approach,” Human-centric Comput. Inf. Sci., vol. 3, no. 1, p. 8, 2015

[3] A. Monreale. Montresor, and Y. Velegrakis, “BLADYG: A Graph Processing Framework for Large Dynamic Graphs,” Big Data Res., vol. 9, pp. 9–17, 2017

[4] D. J. Prajapati, S. Garg, and N. C. Chauhan, “MapReduce Based Multilevel Consistent and Inconsistent Association Rule Detection from Big Data Using Interestingness Measures,” Big Data Res., vol. 9, pp. 18–27, 2017

[5] Priyanka, Vinod Kumar Sharma," Apriori algorithm for mining frequent itemsets-A review", International Journal of Computer Application and Engineering Technology, Vol. 3, Issue 3, pp. 232-236, 2014

[6] Damla Oguz and Belgin Ergenc, " incremental itemset mining based on matrix apriori

algrithm", international journal of Computer Application and Engineering Technology, Vol. 4, pp. 226-230, 2015

[7] K. Rameshkumar," A noval algorithm for association rule mining from data with incomplete and missing value," ictact journal on soft computing, Vol.1, issue 4, 2011

[8] Jia Guo, Jing-yi Ren and Yu-jing Zhang," Comparative study on algorithm for mining association rules based on data mining", International Conference on Computational science and engineering, 2015

[9] S. Sagiroglu and D. Sinanc, "Big Data : A Review," 2013 Internal Confference Collaboration Technology System, pp. 42–47, 2013.

[10] A. Katal, S. Rinzivillo, F. Pratesi, F. Giannotti, and D. Pedreschi, "Privacy-by-design in big data analytics and social mining," EPJ Data Sci., vol. 3, no. 1, pp. 1–26, 2017