

## DFAM: A DISTRIBUTED FEEDBACK ANALYSIS MECHANISM FOR KNOWLEDGE BASED EDUCATIONAL BIG DATA

Rashidah F. Olanrewaju<sup>a</sup>, Burhan Ul Islam Khan<sup>a\*</sup>, Roohie Naaz  
Mir<sup>b</sup>, Asifa Mehraj Baba<sup>c</sup>, Farhat Anwar<sup>a</sup>

<sup>a</sup>Department of Electrical & Computer Engineering, Kulliyah of  
Engineering, International Islamic University, Malaysia

<sup>b</sup>Department of Computer Science & Engineering, National  
Institute of Technology, Kashmir

<sup>c</sup>Department of Electronics & Communication Engineering,  
Islamic University of Science and Technology, Kashmir

### Article history

Received

26 November 2015

Received in revised form

14 January 2016

Accepted

10 October 2016

\*Corresponding author  
burhan.iium@gmail.com

### Abstract

There is almost digitization of the entire educational system, as there is an abundant of digital materials available. The educational system is not left out from the global standardization as well as authorities are imposing certain standard at local and global level. As a result, these data which are getting generated in the context of educational also complies all the basic characteristics of the Big Data such as volume in terms of size, and others velocity, variety etc. In order to store, search and process an open source project, Apache Hadoop has been conceptualized, whereas it lags the application specific needs especially in the field of education to enhance the teaching and learning processes. In this paper, an architectural model is illustrated to demonstrate the existing eco-system and a proposed model for provisioning the enhanced teaching-learning mechanism, so that it can be adopted to enhance the intelligence into mechanism of educational framework.

Keywords: Big data; data analytics; data mining; educational data; knowledge discovery

© 2016 Penerbit UTM Press. All rights reserved

## 1.0 INTRODUCTION

The enormous generation of data in the present time demands a huge amount of parallel and distributed computing capacity such as storage and processing resources for managing a huge amount of data. Cloud computing concept has come into this situation as a very prominent technology which provides a better resolution with the aspect of not only storage services facilities but also Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). The cloud computing infrastructure has been adopted by enterprise level as well as individuals in a large scale. The cloud computing applications are not only limited to some specific domain such as finance, healthcare, e-governance, defense etc, but it has also elaborated its application specific domains into educational institutions. As there are many countries in the world which have come across a situation where educational systems meet a massive change as

students do not bother about the closed classroom trainings, they are more interested into online learning systems which have been introduced in last decade. There are so many online learning platforms for example coursera, udemy, MOOC (Massive Open Online Course) which have got positive responses from the educational aspects [1] [2].

Now days it can be seen that the concept associated with online learning systems has completely changed as 10 years back various E-learning applications only used a unidirectional interaction form of live streaming but the recent communication scenario associated with online learning systems has changed. It allows both the users as well as instructor to communicate in a bidirectional form. The procedure also evolved into formulation of virtual platforms. There are so many existing virtual platforms which ensure a massive exchange of educational big data in a secure and efficient transmission activity. The virtualization of cloud platforms invites so many challenges such as handling a huge amount of heterogeneous data (e.g.,

data associated with different formats) and it also includes accumulating streams of transactional data which can be further preceded for data analytics. Data analytics do the knowledge based data extraction from a huge amount of data. Many existing studies also reviewed some of the present data mining techniques which are unable to handle a massive dimension of data generated by virtualized e-schools. It can also be seen that the existing data mining techniques are only applicable for processing a very little amount of data whereas the massive growth of unstructured data cannot be handled by the conventional relational databases thus non applicability of customized data analytics can happen. Storing the above mentioned data into an existing traditional database also has become very challenging. As the demands for E-learning from various online web applications such as (MOOC and other online applications) requires a knowledge based data analytics where the educational data are also needed to be extracted for some specific useful information regarding the respective domain [3]. An efficient pathway is required to implement the proposed framework in a real-time system which expands the future research direction of predictive modelling, behaviour detection, validation and relationship in between various data mining models.

In the present scenario, it can be seen that different types of tools and frameworks which are used for

Hadoop ecosystem are either proprietary of any enterprises or an open source platform designed by a company, which facilitates all types of validation requirement. In this paper, a novel approach which is termed as DFAM: A Distributed Feedback Analysis Mechanism for Knowledge based Educational Big Data has been introduced.

### 1.1 Overview of E-Educational System

The tremendous growth in IT industries has resulted in new achievements over the cloud computing. The 2014 statistical analysis has highlighted that the educational investment had increased to 30.48 billion. The 68% of the investment covers research and development areas which cover over 200 or more cities and nearly 31 provinces. It has been observed that the research areas cover various institutions of teaching, more than 2000 people containing 1,800 curriculum units, and many shared resources. An investigation says that many universities and colleges have installed the networks in their campus, digital libraries and many educational development infrastructures. There is a school link project in order to provide information regarding primary and secondary level of education. Many organizations have completed structural educational or data information till now but there is a research focus over the development of e-learning for education based on cloud platform.

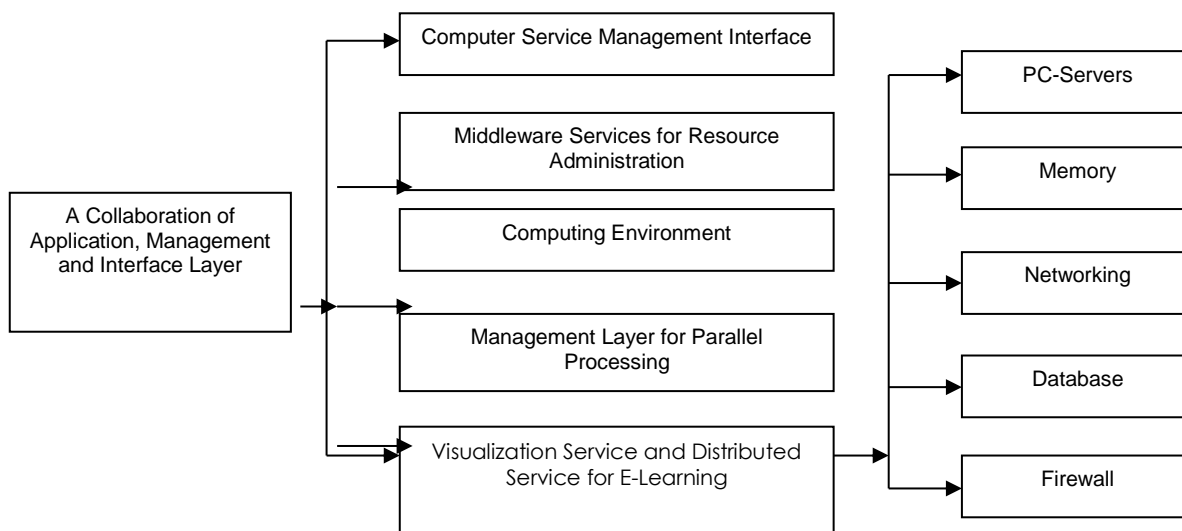


Figure 1 Basic e-educational cloud infrastructure

The study conducted by Nir [4] highlights that 21st century people are more interested in using online E-learning applications; thus there are certain factors (e.g., cost effectiveness, maintainability, usability) which affect the live classroom based training program. It has been found in many industries that if an employee wants to attend the online classes and also he wants to remain up-to-date to his work schedule, it may affect the learning cost. The huge demand of learning efficiency & flexibility in many organizations increases

the association of e-learning courses with the corporate sectors. E-learning can be very much efficient for optimizing time management and it can also provide skilful motivation and knowledge to the employees. It can also be seen that various companies offer E-learning platforms for other companies for their educational and training services [4].

The future research direction in the field of data mining invites various technical challenges such as in the field of access control, infrastructures and

bandwidth utilizations. The concept of E-learning in educational systems ensures an efficient distributed and parallel computing, high speed corporate networks, also high speed cloud and virtualized platforms which can be shared collaboratively and worldwide for providing quality training services to the people 24 hours a day. These virtualized platforms for online educational systems can be utilized for enhancing the business growth of various training institutes worldwide. It will also encourage people for cloud computing oriented distance education in future. The concept will be implemented in terms of very small configurations between very few group settings. The existing studies highlight that students are more interested and encouraged for taking online sessions rather than classroom based traditional course valuation. It is also found that online students maintain more peer to peer contacts with other users during a class work. The users also enjoy it more, spend more time on class work, can be able to understand the material better, and the students who are being taught in an online E-learning platform have performed, on average, 20% better than the students who were taught in the traditional classroom oriented teaching programs. The schematic diagram has been represented using Figure 1.

## 1.2 Related Work

This section introduces some of the existing prior studies and their techniques which are being applied in modern E-learning systems. The best way to review all the existing e-learning systems and their effectiveness is to look into various survey works done by previous researchers.

Rani in [5] has examined different existing methods for using the instruction principles in distributed computing. A study towards information investigation for instructive streams was examined by Macfadyen *et al.* [6]. Tulasi *et al.* [7] have introduced a dialog on part of Big Data examination in advanced education utilizing learning administration framework. A study in comparable bearing has been done by Pocatilu *et al.* [8] and Ashraf *et al.* [9] as well where the creators have introduced positive effects of utilizing distributed computing as a part of e-learning arrangements. A theoretical discussion associated with different customary e-learning frameworks for distributed computing has been done by Sultan in [10] relating to the e-learning framework. Drigas *et al.* [11] have examined some new instruments and existing difficulties of Big Data. Nasr *et al.* [12] have introduced an intelligent apparatus that can be utilized for accomplishing adaptable e-learning environment in distributed computing.

Prativa *et al.* [13] have discussed about different issues connected with Big Data mining. A one of a kind method of getting to an asset utilizing connected information administration over learning framework has been advanced via Carreon *et al.* [14]. Fardoum *et al.* [15] have highlighted different strategies for instructive frameworks in cloud particularly concentrating on

advancement of eLearnXML Platform. Dong *et al.* [16] have displayed an E-learning biological community situated in distributed computing structural engineering.

Masud *et al.* [17] have exhibited an E-learning framework structural planning in view of distributed computing. A factual estimation of E-learning procedures additionally has been presented.

Fernandez *et al.* [18] have displayed the present structure of e-learning utilizations of Cloud Computing. Nuzzo *et al.* [19] presented an investigation carried out on the information of issue 1 of GAW 15. The fundamental target is to recognize which of the SNPs can be viewed as the most educational to anticipate the individual quality expression profiles. After a first stride of quality expression bunching, every individual has been named with a "phenotypic" class, reporter to an arrangement of comparative expression profiles. The issue can in this way be dealt with as a grouping one. In this setting, the most pertinent issue is the high disparity between the quantity of cases and the quantity of characteristics, with the goal that a component choice step is required.

The investigation of Chen *et al.* [20] introduced a bibliometric investigation of basic BI&A productions, specialists, and exploration points taking into account over 10 years of related scholarly and industry distributions. Finally, the six articles that included this uncommon issue are presented and described regarding the proposed BI&A research structure.

Goebel and Gruenwald [21] gave a diagram of normal knowledge based information extraction methods and ways to unravel these assignments. They likewise proposed a component order plot that can be utilized to study learning and information mining programming where the investigation of Aghabozorgi *et al.* [22] led a systematic study which is situated to the difficulties and examination with enormous instructive information included with revealing or using so as to separate information from substantial information sets diverse instructive information mining methodologies and strategies. Ngo *et al.* [23] portrayed an outline and usage for a foundation that will unite the devices and the information to give access to scientists in the field of advanced education institutional examination. The framework will incorporate combination and length of time for information from distinctive sources, implanted factual environment, elite computational back-end, and extensibility for future Big Data and unstructured information.

Daniel [24] distinguished contemporary difficulties confronting establishments of advanced education systems and investigates the capability of Big Data in tending to these difficulties. The paper then blueprints various open doors and difficulties connected with the usage of Big Data in the setting of advanced education. The paper finishes up by sketching out future headings identifying with the improvement and usage of an institutional venture on Big Data.

Moniruzzaman *et al.* [25] have proposed a study to talk about the sources and difficulties of enormous

information furthermore contrasting the different NoSQL databases taking care of huge information.

It is found in the existing survey papers that there is a huge availability of implementation studies which talk about the design perspective associated with E-learning systems. A significant research gap highlights that a lot of implantation strategies on E-learning systems are very much repetitive in nature and they only focus on maintaining data storage and various application features. There are fewer implementation studies which discuss about analyzing the data. Moreover, very few explored about Big Data. It should be pointed out that there is a huge difference between normal data and Big Data. Insertion of a massive amount of normal data into existing Relational Database Management Systems (RDMS) is quite easy and it can also be analyzed using existing data mining techniques. But Big Data is generally considered as a stream of information having qualities e.g., information heterogeneity, substantial volume, instability, data velocity and so forth. Inferable from such qualities of existing instructive information, it is neither conceivable to store in social database framework nor is appropriate to be subjected to fantastic mining calculations. Thus, there is an unmistakable need of a procedure that can address such issues in instructive enormous information. The following area will show the proposed framework that is implied for performing information extraction from instructive information.

## 2.0 METHODOLOGY

The proposed study aims to present a novel and efficient technique which is termed as A Distributed Feedback Analysis Mechanism (DFAM) for knowledge based educational Big Data which can be applicable for the enhancement of the concepts associated with E-learning mechanisms in cloud platform. The design perspective of the proposed system also includes the feedback analysis of different type of sentiments given by the respective users of the proposed system. The design principle of the proposed DFAM includes collaborative and distributed characteristics where collaborative characteristics mean to evolve the transmission and processing of a huge stream of heterogeneous data such as structured, unstructured and semi structured data. The distributed scenario of the proposed technique offers a better, flexible and platform independent communication for processing of huge amount of data; it also provides much pervasiveness to acquire efficient data transmission. The proposed system uses the same design pattern followed by software engineering. The performance metrics of the proposed system will be evaluated with respect to performance parameters such as CPU elapse time and data time in an experimental prototyped test bed. The comparative analysis shows the effectiveness of the proposed DFAM mechanism as compared to the existing techniques e.g., MOOC (Massive Open Online Course). The hidden agenda for

working on the concept of DFAM is the improvement of teaching as well as learning experiences in the field of e-learning systems. The proposed system which is denoted as DFAM has been developed with the use of various pre-processing tools and elements which can be further utilized for systematic achievement of huge heterogeneous data associated with ubiquitous and distributed environment of E-learning systems. A collaborative interface is designed for this purpose which can be accessed by three types of actor's viz. i) student ii) instructor, and iii) policymaker. The significant contributions of DFAM are i) to provide exposure to multidisciplinary domain e.g., big data, semantic, and data analytics techniques on educational data, ii) to formulate a novel text-mining technique of emotional and sentiment analysis, iii) to identify the attitude of actors (positive/negative). The architectural design paradigm of the proposed DFAM is depicted in Figure 2.

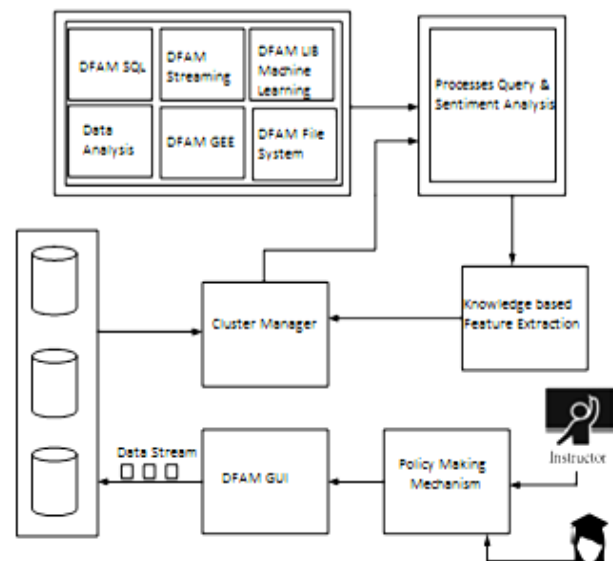


Figure 2 Architecture of DFAM

The above figure illustrates the architecture of the proposed system which includes two different types of modules which are student and instructor module respectively. The student and instructor both will be considered as users of the proposed system. Firstly, the student will be connected with a policy making framework which will lead him into a graphical user interface of the proposed system. The data streams are being flowed into databases and then the cluster manager module executes the flow of data streams and processes the queries and sentiments for the analytics. In Figure 2, the graphical representation of the proposed system also highlights another module where six different types of sub modules are present. The respective sub modules of the proposed system are given below.

**DFAM SQL:** This sub module can process different types of SQL queries.

**DFAM Streaming:** It also controls the flow of data streams in a distributed and collaborative manner. The



computing resources associated with this evolve parallel computing scenario for ensuring proper time synchronization in a request and response paradigm.

**DFAM LIB Machine Learning:** This module introduces a new machine learning concept for the classification and clustering of heterogeneous datasets.

**Data Analytics:** This module will do the data analysis or sentiment analysis with respect to knowledge based information extraction from heterogeneous data streams.

**DFAM GEE:** DFAM General Execution Engine is a different framework for execution of different DFAM file in heterogeneous distributed files systems.

**DFAM File System:** The proposed system has been implemented using its own operating system which is termed as DFAM File System. The proposed system can be evaluated irrespective of any type of application or environmental scenario.

### 3.0 RESULTS AND DISCUSSION

The proposed model can be implemented in a real time scenario using three different types of modules namely Client profiling, Suggestive inputs and study of feedback based on the sentiments of the clients. The proposed system first initializes a dynamic registration form for retrieving the scholar information. After performing the registration process, a graphical user interface of the proposed system will be provided. The above mentioned module is facilitated with the privileges of giving or expressing individual sentiments. The above module also includes proper validation of the users. The data type which has been taken into consideration is heterogeneous in nature; also, the user profiling has been done using a traditional RDBMS. The conversation of data is considered here in text format. The proposed system has been designed and evaluated using Java platform on general 32-bit system consisting of above 1,00,000 student clients who are solely recognized to use DFAM thereby occupying around 130 petabytes of unformatted data. In this section, mathematical models of the proposed DFAM have been evaluated and it also illustrates the further manifestation of the model for implementing DFAM in a mathematical model. There are two different types of modules which have been developed for prototyping the mathematical modelling. One sub-section gives the complete detailing of the proposed DFAM model and another section illustrates the prototyping of DFAM model. Both the sub-sections are discussed below.

#### 3.1 Development of DFAM Model:

In this portion, the proposed DFAM model has been evaluated in a theoretical aspect. The proposed system includes a stream of educational heterogeneous data which is represented by  $H_D$  and considers a set of attributes which is denoted as  $\delta$ . The proposed system also generated a relation and mathematically expressed that by  $a \in \delta$  where  $a$  is a single attribute

and belongs to a set of attributes. The system also tags the attributes as  $T_a$ . A connecting component  $\gamma_m$  has been designed which maps the components of  $H_D$  with the components of  $\gamma_m$ . The proposed mathematical model for designing DFAM is denoted by the equation 1.

$$D = \{H_D, \delta, \{T_a | a \in \delta\}, \{\gamma_m | \gamma_m\}\} \quad (1)$$

In the next level, an equation will be established which will establish a formula for which  $\Omega$  signifies the meaningful occurrence of text data in a DataStream. The proposed mathematical model is defined by

$$f(\Omega) = \{b \in H_D, | b |= \Omega\} \quad (2)$$

Hence, here  $\Omega$  is defined as a set of attributes which are having the respective properties mentioned in the equation 2. The proposed system uses a lemma concept which has been applied for analysis of sentimental feedbacks which can be either in a structured, unstructured or semi-structured form. Different types of Lemma modules are explained as follows.

**Lemma -1:** In this method, a sentiment which is denoted by  $(\Omega, f(\Omega))$  can be denoted with another sub sentiment which is  $(\kappa, f(\kappa))$ , it can be a super sentiment of  $(\Omega, f(\Omega))$ , if it satisfy one condition which is  $f(\Omega) \subseteq f(\kappa)$ .

**Lemma -2:** The concept of Lemma-2 defines that  $(\Omega, f(\Omega))$  which is a sentiment can be denoted by a proper non empty sentiment if there is no non empty sentiment present in the set of  $(\Omega, f(\Omega))$ .

**Lemma -3:** Two types of sentiments which are  $(\Omega, f(\Omega))$  and  $(\kappa, f(\kappa))$  are considered as two disjoint sets if they satisfy the condition  $f(\Omega) \cap f(\kappa) = Null$  else it can be said that the sentiments belonging to the above mentioned sets have no-empty relationship. Hence, the proposed system applies a valid approach which is based on a lattice-based approach and can be applicable for detecting positive and negative responses. The following equations define how positive and negative responses can be calculated.

+ve responses are calculated by

$$(\Omega, f(\Omega)) \cup (\kappa, f(\kappa)) = (\Omega \in \beta, f(\Omega) \cup f(\lambda)) \quad (3)$$

-ve responses are calculated by

$$(\Omega, f(\Omega)) \cap (\tau, f(\tau)) = (\Omega \in \tau, f(\Omega) \cap f(\tau)) \quad (4)$$

As discussed in the formulation above, DFAM can be utilized for building a particular knowledge based text extraction from emotions just by utilizing the traditional data-mining approach in  $H_D$  as a section of investigating process. DFAM constructs a rule  $\Omega \rightarrow \tau$ , where  $\Omega$  and  $\tau$  are intentions of two sentiments.

#### 3.2 Implementing DFAM Model in Prototype:

The proposed system has been evaluated with Java and server based scripts. The proposed DFAM based system uses only a single client interface and that GUI

can be accessible by different type of users such as:- 1. Guide 2. Student and 3. Policy maker. Access to different modules is secured by standard cryptographic technique which is based on secure authentication mechanism.

The proposed system detects and captures high dimensional data with respect to huge data streams. DFAM executes some different modules which check the occurrence of a particular word which carries

meaningful information and performs the above mentioned strategies and lemma methods. The occurrence of the particular term is searched from the streams of a document. Input data is feedback shared by the clients involved which is in the form of text. In the end, the system implementation of DFAM logic performs the retrieval of knowledge in the form of positive as well as negative emotions.

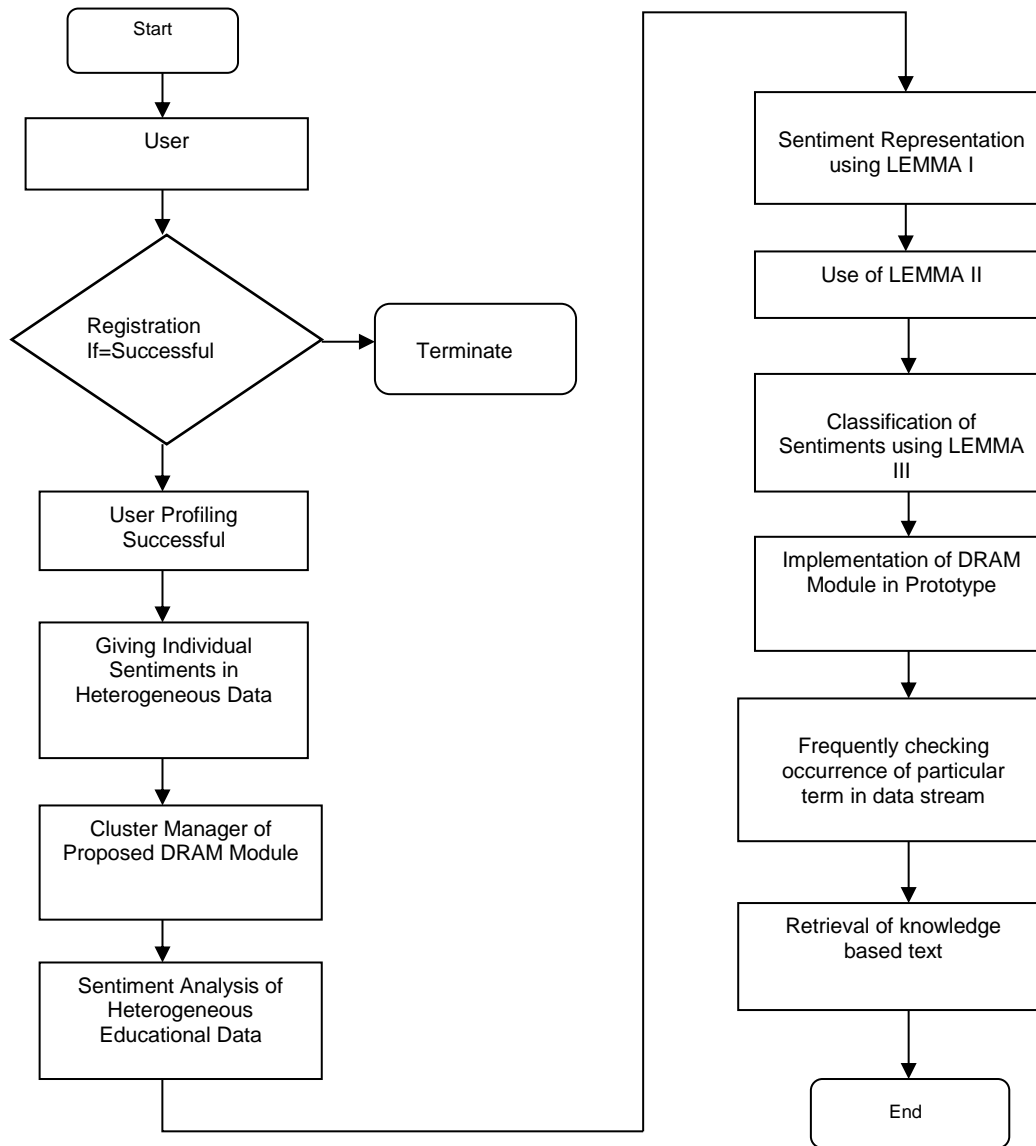


Figure 3 Flow Diagram of the Implementation

### 3.3 Study Outcomes

This section illustrates the experimental outcomes of the proposed system DFAM where CPU processing time with respect to Data Size is taken as performance parameters of the proposed system. The experimental prototyping is done in a real-time scenario whereas the

computation time associated with processing of huge stream data has been evaluated by the proposed system. Normally it can be seen that for processing of huge data streams, potential memory is required. The experimental outcomes have been anticipated to show how much time is required to process a massive amount of data by the proposed DFAM systems.

CPU processing time and respective data size have been taken into consideration for the measurement of performance metrics of the proposed DFAM technique. A test bed has been built for analyzing the required processing time with respect to memory size as huge memory is required to process a flow of data stream.

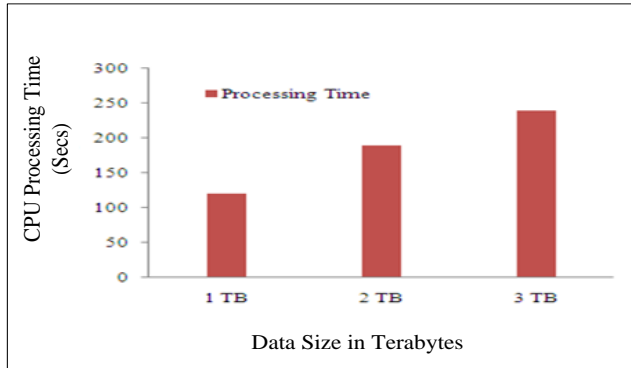


Figure 4 Outcome of DRAM (Processing Time)

For the better understanding of the effectiveness associated with the proposed system, it can be anticipated by the above Figure 4 that very little time will be required to process a heavy stream of data. The experimental prototype has been developed in order to optimize the Computational time. The proposed DFAM ensures a very low computational complexity. The proposed model uses student, instructor and policy makers as prime user modules. A real time exchange of educational big data has been considered in order to set up the complete experimental prototype. CPU processing time with respect to multiple machines are also taken into consideration for checking that how much time parallel and distributed processing units are consuming for analyzing data as well as sentiments of distributed machines. Three types of experimental data sets of 1 to 3 terabytes have been processed to generate the desired outcomes. In Figure 4, it can be seen a graphical representation has been given and outcome also approximated to the round values. It was seen in various existing systems and their respective comparative analysis ensures that Apache Spark is found to be one of the faster processing tools which can manage and analyze a massive data stream very efficiently. Many existing studies also talk about its effectiveness in terms of memory management and parallel processing. It has been successfully installed in a very robust environment and it is also capable of processing 1000 TB of data in 234 minutes with respect to 190 nodes. A similar kind of experimental analysis is performed with DFAM and the respective outcomes are presented in Figure 4. The above figure also showcases that the proposed system outperforms the existing Apache Spark methodology. The prime reason behind the inefficiency of Apache Spark is that it uses existing Hadoop ecosystems and map reduces programs for the development of their proposed

concept, hence the proposed algorithm became computationally time consuming as it also adopts the complex design principles of software frameworks. An insight into the design specification of the proposed DFAM systems reveals that it is a simple enhancement of the existing data mining tool WEKA3 and it got sufficient potential and capability for processing and analyzing a massive amount of heterogeneous data very efficiently. The prime contribution of this proposed study is that even in the absence of conventional big data analytics tool such as Hadoop, Map Reduce, Cassandra etc., the proposed system can be able to perform better in terms of computation and efficient memory usage.

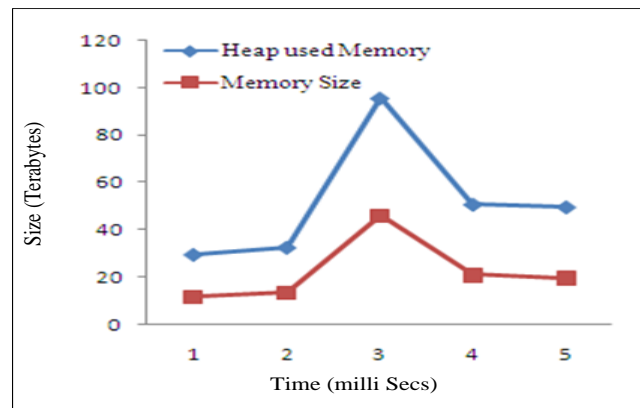


Figure 5 Memory usages by proposed framework

A closer look into the above represented graphical depiction which is Figure 5 can showcase that performance analysis of the proposed DFAM system also has been done with respect to memory consumption, where it can be anticipated that how much heap is used by the memory within a particular interval of time. The experiment has been performed and outcomes are approximated to the round values. It also shows that the proposed DFAM system is much effective than the existing Apache Spark.

## 4.0 CONCLUSION

The planned performance evaluation of DFAM adopts experiential characters. In this study, a complete interaction collaborative application has been designed, where user profiling is done based on the user data collected and after that a semantic analysis of required discussion forum educational data has been evaluated. The experimental prototyping uses an algorithmic approach of mining and the lemma modules which have been designed in DFAM. In the section VI, the experimental outcomes of the proposed system has been highlighted which ensures that the proposed DFAM algorithm is very much effective as compare to the existing Apache spark which uses the distributed Hadoop ecosystem as well as various Map

Reduce programs. The comparative analysis of the proposed system evaluate that DFAM is much more efficient as compare to the existing Apache Spark system in terms of processing time as well as memory consumption. The proposed system also utilizes the WEKA3 experimental test bed for the processing of massive heterogeneous Big Data stream. It can be seen that the experimental prototyping has been done in a real time scenario and it uses an acceptable amount of data which are generated by various distributed nodes. The proposed algorithm also ensures the cost effectiveness as well as very less commutation time as compare to existing traditional E-learning techniques. The future extent of the proposed framework incorporates treatment of heterogeneous information in a disseminated situation; additionally the proposed framework would be multifold, one it sets out an establishment of taking care of Educational Big Data, particularly unstructured content information, second it takes into account sentiment analysis which is exceptionally novel methodology in instructive framework. The proposed study will be helpful for the further continuation of research in Data Mining and Big Data analytics.

## References

- [1] Devadass, R., L. Haldurai., G. Vidya, R. Gokilavani, and T. Madhubala. 2015. Data Visualization Techniques on Partially Purified Plant Extracts against *Aedes aegypti* (Culicidae: Diptera). *IJARCSSE*, 5(1)
- [2] Domenico, T. 2011. Cloud Computing and Software Agents: Towards Cloud Intelligent Services. In *WOA*, 11: 2-6
- [3] Kim, S., S. Su-Mi, and Y. Yong-Ik. 2011. Smart Learning Services Based on Smart Cloud Computing. *Sensors*, 11 (8): 7835-7850
- [4] Nir, K. 2010. Cloud Computing in Developing Economies. *Institute of Electrical and Electronics Engineers Computer*, 43(10): 47-55
- [5] Rani, U. 2013. Effective Ways Cloud Computing can Contribute to Education Success. *Advanced Computing*, 4(4)
- [6] Macfadyen, L.P., D. Shane, P. Abelardo, and G. Dragan. 2014. Embracing Big Data in Complex Educational Systems: The Learning Analytics Imperative and the Policy Challenge. *Research & Practice in Assessment*, 9(2): 17-28
- [7] Tulasi, B. 2013. Significance of Big Data and Analytics in Higher Education. *International Journal of Computer Applications*, 68(14)
- [8] Paul, P., A. Felician, and V. Marius. 2009. Using Cloud Computing for E-Learning Systems. In *Proceedings of the 8th WSEAS International Conference on Data Networks, Communications, Computers (Dncoco'09)*, 7-9
- [9] Ahmed, A., M.E. Hazem, M., E-Razek, M.A. Samir, E. Yehia, and M. Nikos. 2014. Enhancing Big Data Processing in Educational Systems. *Research & Practice in Assessment*, 17-28
- [10] Sultan, N. 2010. Cloud Computing for Education: A New Dawn. *International Journal of Information Management*, 30(2): 109-116
- [11] Drigas, A.S., and L. Panagiotis. 2014. The Use of Big Data in Education. *International Journal of Computer Science Issues (IJCSI)*, 11(5): 58.
- [12] Nasr, M. and S. Ouf. 2011. An Ecosystem in E-Learning using Cloud Computing as Platform and Web2.0, *The Research Bulletin of Jordan ACM* 2.134-140.
- [13] Pratiba, D., and G. Shobha. 2014. Educational Big Data Mining Approach in Cloud: Reviewing the Trend. *International Journal of Computer Applications*, 92(13): 43-50
- [14] Gutierrez-Carreón, G., T. Daradoumis and J. Jorba. 2015. Integrating Learning Services in the Cloud: An Approach that Benefits Both Systems and Learning. *Journal of Educational Technology & Society*, 18(1): 145-157
- [15] Fardoun, H.B., S. R. Lopez, D. M. Alghazzawi, and J. R. Castillo. 2012. Education System in the Cloud to Improve Student Communication in the Institutes of: C-Learn XML. *Procedia-Social and Behavioral Sciences*, 47:1762-1769
- [16] Dong, B., Q. Zheng, J. Yang, H. Li, and M. Qiao. 2009. An E-Learning Ecosystem Based on Cloud Computing Infrastructure. In *Advanced Learning Technologies, ICALT, Ninth IEEE International Conference*, 125-127
- [17] Masud, M.A.H. and X. Huang. 2012. An E-Learning System Architecture Based on Cloud Computing. *System*, 10(11)
- [18] Fernandez, A., D. Peralta, F. Herrera, and J. M. Benítez, 2012. An Overview of E-Learning in Cloud Computing. In *Workshop on Learning Technology for Education in Cloud (LTEC'12)*, Springer Berlin Heidelberg, 35-46
- [19] Nuzzo, A., C. Fuchsberger, and R. Bellazzi. Retrieved, 10<sup>th</sup> Oct, 2015. Data Mining Approaches to Forecast Molecular Phenotypes from SNPs Data.
- [20] Chen, H., R.H.L. Chiang, and V.C. Storey. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4): 1165-1188.
- [21] Goebel, M., and L. Gruenwald 1999. A Survey of Data Mining and Knowledge Discovery Software Tools. *ACM SIGKDD Explorations Newsletter*, 1(1): 20-33.
- [22] Aghabozorgi, S., H. Mahrooian, A. Dutt, T.Y. Wah, and T. Herawan. 2014. An Approachable Analytical Study on Big Educational Data Mining. In *Computational Science and Its Applications-ICCSA*. Springer International Publishing, 721-737
- [23] Ngo, L.B., P. Xuan, K. Ferguson, C. Marshall, J. McCann, Y. Zheng, and A.W. Apon 2012. An Infrastructure to Support Data Integration and Curation for Higher Educational Research. In *Proc. 8th IEEE International Conference on e-Science, Chicago*, 8-12
- [24] Daniel, B. 2015. Big Data and Analytics in Higher Education: Opportunities and Challenges. *British Journal of Educational Technology*, 46(5): 904-920.
- [25] Moniruzzaman, A.B.M., and S.A. Hossain. 2013. Nosql Database: New Era of Databases for Big Data Analytics-Classification, Characteristics and Comparison. arXiv preprint arXiv:1307-0191