

## Distance Learners' Attributes in Optimizing Learning Achievement Using Learning Analytics

Noraizan Amran<sup>1</sup>, Farrah Diana Saiful Bahry<sup>2</sup>, Halida Yu<sup>3</sup>

Faculty of Information Management, Universiti Teknologi MARA, Cawangan Selangor,  
Kampus Puncak Perdana, 40150 Shah Alam, Selangor, Malaysia.

Email: noraizan@uitm.edu.my<sup>1</sup>, farrahdiana@uitm.edu.my<sup>2</sup>, halida917@uitm.edu.my<sup>3</sup>

To Link this Article: <http://dx.doi.org/10.6007/IJARBSS/v11-i1/8157> DOI:10.6007/IJARBSS/v11-i1/8157

**Published Date:** 24 January 2021

### Abstract

The distance learning operations and activities from Learning Management System (LMS) and various information system provide huge educational data. Data mining can be used as a tool to extract the huge educational raw data and turn them into useful information. Educational Data Mining (EDM) is the utilization of data mining systems on educational data, to analyse insight information and to determine educational strategic direction. This paper aims (1) to propose distance learners' attributes in optimizing learning achievement using learning analytics and, (2) to identify the application areas of the attributes for Educational Data Mining and Learning Analytics. Data pre-processing is the first step in any data mining process and, Waikato Environment for Knowledge Analysis (WEKA) is one of the machine learning algorithms for educational data mining undertakings. It allows the transformation of raw educational data into a suitable format, ready to be used by a data mining algorithm for performing a specific educational data set. Therefore, a precise description of distance learners' attributes views was developed and gathered. A framework was designed to demonstrate the proposal of distance learners' attributes from distance learners' profile, distance learners' learning activities and distance learners' learning behaviour; and it will shape the improvement of Distance Learners' Learning Achievement. Hence, the specified educational dataset in the proposal can be adapted and visualize using any learning analytics tools for strategizing distance learning new strategies.

**Keywords:** Information Management, Data Mining, Information Science, Educational Data Set, Learning Analytics.

### Introduction

Today, the opportunity for lifelong learning lies there for everyone, anywhere and at any time. Modern technology makes learning becoming borderless, convenience and easy. This is demonstrated in the recent covid-19 pandemic crisis. Although the sudden outbreak has taken us by surprise, teaching and learning activities are still possible, conducted via various online platforms. Throughout the covid-19 crisis, school children all over the world are joining

other adult learners and mature students who are furthering their studies via technology supported distance learning programs.

This distance learning operations and activities provide huge educational data, which one of the main sources is the database from the Learning Management Systems (LMS) that is developed and used by the institution and distance learners' themselves. The information systems provide the data collection, storage, and retrieval; facilities the transformation of data into information and manages both data and information (Coronel and Moriss, 2019). Whichever type of educational data is taken from either students' use of interactive learning environments, computer-supported collaborative learning, or administrative data from schools and universities, it often has multiple levels of meaningful hierarchy, which often need to be determined by properties in the source data itself, rather than in advance. Issues of time, sequence, and context also play important roles in the study of educational data (Educational Data Mining, 2020). Educational data mining looks for new patterns in data and develops new algorithms and add/or new models, while learning analytics applies known predictive models in instructional systems (Mining, T. E. D., 2012).

Distance Learning or E-Learning Distance learning is defined as an organized educational process that bridges between the distance of the learners and the educator by the means of technology with minimum meetings (Fauzi & Iga, 2020). Distance learning, or also known as distance education, e-learning, and online learning, is a form of education in which the main elements include physical separation of teachers and students during instruction and the use of various technologies to facilitate student-teacher and student-student communication (Encyclopedia Britannica, 2020). Distance education is a planned learning experience or method of instruction characterized by quasi-permanent separation of the instructor and learner(s). Within a distance education system, information and communication are exchanged through print or electronic communications media (Keegan, 1980). UNESCO defines distance education as an educational process and system in which all or a significant proportion of the teaching is carried out by someone or something removed in space and time from the learner. In addition, to help teachers develop the characteristics of good teaching, distance learning programs will need to provide teachers with ongoing opportunities to improve their content knowledge, instructional skills, knowledge about how students learn, and understanding of learning from a student point of view (Burn, 2011).

Hence, this paper aims (1) to propose distance learners' attributes in optimizing learning achievement using learning analytics and, (2) to identify the application areas of the attributes for educational data mining and learning analytics.

### **Educational Data Mining**

Data mining can be used as a tool to extract huge educational raw data to become useful information. In the industry, data from the database can be analyzed and used as valuable information to the organization. In educational system, data mining techniques can search useful patterns and these useful patterns can be used for many purposes such as predicting student performance, grouping students and enrollment management (Saini, 2014). Data mining can be connected to many fields and business application, for example, retail deals, internet business, remote detecting, bio informatics and others (Ahmadi and Ahmad, 2013). Education is the most essential element to pick up learning and achievement, in this manner data mining is especially pertinent to educational areas. Thakur and Mahajan (2015) used

powerful data mining technology weka tool for the preprocessing, classification and analysis of institutional result of undergraduate students.

Educational Data Mining (EDM) is the utilization of data mining systems on educational data. The goal of EDM is to break down such information and to determine educational research issues. EDM manages growing new techniques to investigate educational information and utilize data mining strategies to better comprehend student learning environment. The EDM procedure changes raw information originating from education frameworks into useful data that could conceivably affect education research and practice. A predictive data mining model using classification-based algorithms also has being used to identify and display the slow learners among students (Kaur et al., 2015). Educational data analysis can provide an insight on what students know, what they should know, and what can be done to meet their academic needs. With appropriate analysis and interpretation of data, educators can make decisions that positively affect student outcomes (Lewis, 2010).

The data mining techniques can be grouped into classification, clustering, association analysis and decision tree (Saini, 2014). Classification is the required technique and mainstream approach. Classification is the way toward finding an arrangement of models or capacities that depict and recognize information classes or ideas, with the end goal of having the capacity to utilize the model to anticipate the class of articles whose class name is obscure. Unlike a classification model, the reason for expectation model is to decide the future result instead of the present conduct. Its yield can be absolute or numeric esteem. It is a regulated learning since the classes are resolved before analyzing the data. Clustering is best suitable for finding groups of similar data items. In clustering, an arrangement of information things is divided into an arrangement of classes with the end goal that comparative attributes things are gathered (Prabha, 2015). Association analysis is utilized to find relationships between attributes and items, for example, as the presence of one pattern implies the presence of another pattern (Ahmadi and Ahmad, 2013). Association rule is a mainstream method for market basket analysis since every conceivable mix of fascinating item grouping can be investigated while; decision tree is a tree structure like a flowchart, in which the rectangular boxes are known as the node. Each node represents to an arrangement of records from the original data set. Internal node is a node that has a child and leaf (terminal) node is nodes that do not have children. Root node is the highest node. The decision tree is utilized for finding the most ideal approach to recognize a class from another class (Saini, 2014). In data mining environments, there are a lot of approaches, application and techniques can be applied to the any different kind of situation and environment. The various kind of algorithm model can be used from the data mining are such as neural nets, genetic algorithms, k-nearest Neighbor, Naive Bayes, support vector machines, WEKA, decision trees, neural networks (Aher and Lobo, 2011), (Kaur et al., 2015). Educational data from Malaysian Student Information System (SIS) and School Examination Analysis System (SEAS) database also can be fully utilized either by the individual school to identify the factors which contribute to the academic performance, or by the State or Federal School Administrator which involves bigger volumes of data (Amran et al., 2016).

### **Weka Data Mining**

WEKA is considered as a point of interest framework in the historical backdrop of the data mining among machine learning research groups. The open-source clustering toolkit Weka is

used for analyzing the K-means algorithms, Hierarchical clustering and Density based clustering (Ayyoob, 2015). WEKA is a collection of machine learning algorithms for data mining undertakings. The algorithms can either be applied straightforwardly to a dataset or called from our own Java code. The WEKA workbench contains a collection of visualization tools and algorithms for data analysis, together with graphical user interfaces for simple access to this functionality. It is freely available software. It is portable & platform independent due to it is completely executed in the Java programming language and in this way keeps running on any platform. WEKA has a few standard data mining undertakings, data preprocessing, clustering, classification, association, visualization, and feature selection (Ahmadi and Abadi, 2013).

The WEKA GUI chooser dispatches the WEKA's graphical areas which has six buttons: Simple CLI, Explorer, Experimenter, Knowledge Flow, ARFFViewer, and Log (Aher and Lobo, 2011). The imperative elements behind the achievement of WEKA are it gives various algorithms to data mining and machine learning, open source and freely available, platform-independent, effortlessly useable by individuals who are not in data mining authorities, gives adaptable facilities to scripting tests and has stayed up with the latest, with new algorithms being included in research literature. The methods which are given by the WEKA are data preprocessing and visualization, attribute selection, classification (OneR, Decision trees), prediction (Nearest neighbor), model evaluation, clustering (K-means, Cobweb) and association rules (Joshi and Panchal, 2014). As indicated by Aher and Lobo (2011), in WEKA tools the Explorer interface has a few boards that offer access to the fundamental segments of the workbench which are the Preprocess panel imports the data from a database, a CSV file, ARFF and others, and preprocesses this data utilizing filtering algorithm which can be utilized to change the data from one format to other for instance numeric attributes into discrete ones.

Data mining will be viewed as most helpful in educational areas. Anticipating students' academic performance is of extraordinary concern to the education institutes. By applying data mining techniques and tools in expectation of students' performance is useful to recognize the capabilities of students, their interests, and shortcomings furthermore supportive to cluster the students based to their performance.

### Distance Learners' Data Pre-Processing

Data pre-processing is the first step in any data mining process. It allows transforming the available raw educational data into a suitable format ready to be used by a data mining algorithm for solving a specific educational problem. Pre-processing is always the necessary first step in any data mining process/application. This task is very important because the interestingness, usefulness and applicability of the obtained data mining models highly depend on the quality of the used data. Figure 1 shows the main data pre-processing steps adopted from Romero et al. (2014).

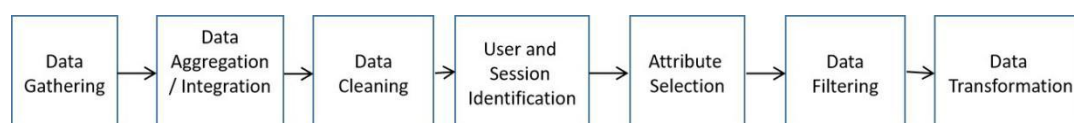


Figure 1: Data pre-processing steps

A precise description of end-user data views must be developed and gathered. It used to help to identify the main educational data elements. The information gathered been group together base on the information needs, users, and sources; and the data elements needed to produce the information (Amran et al., 2018) which called as data integration. A data cleaning was performed in order to standardize the data format, removing the irrelevant and fixing the incomplete data. In this study, attribute selection is the last step that been done to identified which attributes that significance as part of the data set or data element. Data filtering and data transformation will be used in future study when performing the real data analytic process.

This initial process is crucial in shaping the success of data analytic process. It also needs to align with higher institution learning objectives and directions. Districts and higher education institutions typically have much more data than they use to inform their actions. Part of the problem is that data reside in multiple systems in different formats. The development of standards for education information systems, software to facilitate data integration from multiple systems, and designing easy-to-use data dashboards on top of different data systems are all active areas of technology development. Data mining and analytics can be done on a small scale. In fact, starting with a small-scale application can be a strategy for building a receptive culture for data use and continuous improvement that can prepare a district to make the best use of more powerful, economical systems as they become available. Starting small can mean looking at data from assessments embedded in low-cost or open learning systems and correlating those data with student grades and achievement test scores (Mining, T. E. D., 2012).

### Proposed Distances Learners Attributes

A framework has been designed to demonstrate the proposal of distance learners' attributes from distance learners' profile, distance learners' learning activities and distance learners' learning behaviour in optimizing learning achievement using learning analytics as at Figure 2.

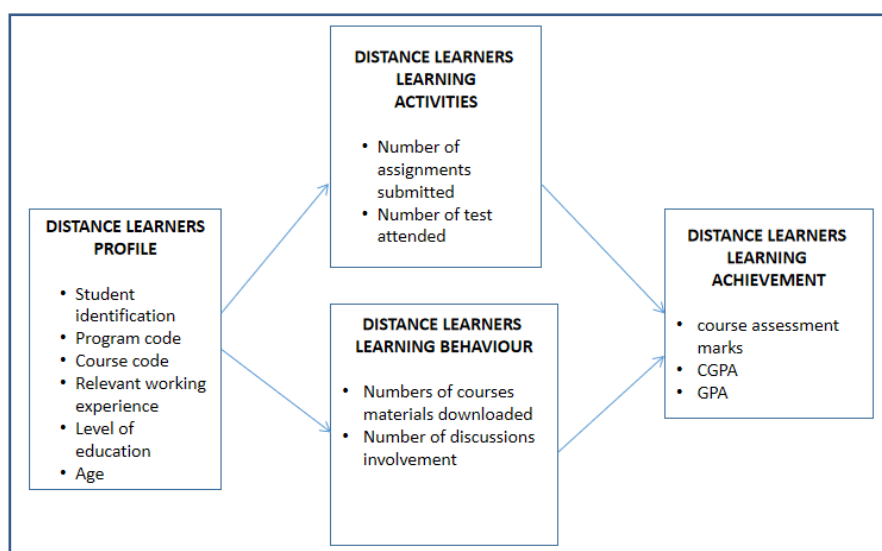


Figure 2: A Proposed Model on Distance Learners' Attributes

1. Distance Learners' Profile
  - a) Distance Learner Identification (Student Identification), the program code enrolled by the distance learner and the course code registered by the distance learner. The source of this data is the Student Information Management Systems (SIMS).
  - b) Distance Learners' relevant working experience for adult/working learners, namely the working industrial type, highest position level and years service in the industry. The source of this data may come from the Student Record/File itself.
  - c) Distance Learners' previous education, namely as educational level, and educational field. The source of this data may come from the Student Record/File itself.
  - d) Distance Learners' age while enrolled the programme. The source of this data may come from the Student Record/File itself.
  
2. Distance Learners' Learning Activities for the courses, which data from Learning Management System (LMS/uFUTURE/i-Class)
  - a) Number of assignments submitted
  - b) Number of tests attended
  
3. Distance Learners' Learning Behaviour for the courses, which data from Learning Management System (LMS/uFUTURE/i-Class)
  - a) Numbers of courses materials downloaded
  - b) Numbers of discussions involvement
  
4. Distance Learners' Learning Achievement for the courses, which data from Student Information Management System (SIMS/eRES)
  - a) Course assessment marks
  - b) Grade Point Average (GPA) is an academic performance from 0.00 - 4.00, calculated as an average of the GPA in the current semester.
  - c) Cumulative Grade Point Average (CGPA) is an academic performance from 0.00 - 4.00, calculated as an average of the GPA from all completed semesters.

The distance learners' educational entities, attributes and sources has been tabulated in Table 1.

Table 1

*Distance Learners' Educational Entities, Attributes and Sources*

Entities	Description	Attribute	Source
DISTANCE LEARNERS' PROFILE	Learners ID, program and course	<ul style="list-style-type: none"> <li>● Student Identification</li> <li>● Program Code</li> <li>● Course Code</li> </ul>	Student Information Management systems (SIMS)
	Relevant working experience	<ul style="list-style-type: none"> <li>● Industrial type</li> <li>● Position level</li> <li>● Years service</li> </ul>	Student Record / File
	Level of Education	<ul style="list-style-type: none"> <li>● Educational level</li> <li>● Educational field</li> </ul>	Student Record / File
	Age on registration	<ul style="list-style-type: none"> <li>● Student age</li> </ul>	Student Record / File
DISTANCE LEARNERS' LEARNING ACTIVITIES	Course assessment	<ul style="list-style-type: none"> <li>● Number of assignments submitted</li> <li>● Number of tests attended</li> </ul>	Learning Managements System (LMS/ uFUTURE/ i-CLASS)
DISTANCE LEARNERS' LEARNING BEHAVIOUR	Course materials	<ul style="list-style-type: none"> <li>● Numbers of courses materials downloaded</li> </ul>	Learning Managements System (LMS/ uFUTURE/ i-CLASS)
	Group discussion	<ul style="list-style-type: none"> <li>● Number of discussions involvement</li> </ul>	Learning Managements System (LMS/ uFUTURE/ i-CLASS)
DISTANCE LEARNERS' ACHIEVEMENT	Course assessment marks	<ul style="list-style-type: none"> <li>● Course assessment marks</li> <li>● Course grade</li> <li>● CGPA</li> <li>● GPA</li> </ul>	Student Information Management systems (SIMS/eRES)

Bienkowski et. Al (2012) presents broad areas of applications that are found in practice, such as (1) modeling of user knowledge, user behaviour, and user experience; (2) user profiling; (3) modeling of key concepts in a domain and modeling a domain's knowledge components, (4) and trend analysis. These areas represent the broad categories in which data mining and analytics can be applied to online activity, especially as it relates to learn online.

The application areas of the attributes for educational data mining and learning analytics have been identified and the model of application areas also have been developed as at Table 2.

Table 2

*Application Areas for Educational Data Mining and Learning Analytics*

<b>Application Area</b>	<b>Proposed Distance learning data analysis objectives</b>	<b>Types of Data Needed for Analysis</b>	<b>Source</b>
Distance Learners' Profiling and Achievement Modelling	Do relevant working experience assist the distance learners' studies and the achievement	<ul style="list-style-type: none"> <li>● Student identification</li> <li>● Program code</li> <li>● Course code</li> <li>● Industrial type</li> <li>● Position level</li> <li>● Years service</li> <li>● Course assessment marks</li> <li>● Course grade</li> <li>● CGPA</li> <li>● GPA</li> </ul>	<ul style="list-style-type: none"> <li>● Student Information Management systems (SIMS)</li> <li>● Student Record / File</li> <li>● Student Information Management systems (SIMS/eRES)</li> </ul>
	Do previous education assist the distance learners' studies and the achievement	<ul style="list-style-type: none"> <li>● Student identification</li> <li>● Program code</li> <li>● Course code</li> <li>● Educational level</li> <li>● Educational field</li> <li>● Course assessment marks</li> <li>● Course grade</li> <li>● CGPA</li> <li>● GPA</li> </ul>	
	Do age influence distances' learners studies and the achievement	<ul style="list-style-type: none"> <li>● Student identification</li> <li>● Program code</li> <li>● Course code</li> <li>● Student age</li> <li>● Course assessment marks</li> <li>● Course grade</li> <li>● CGPA</li> <li>● GPA</li> </ul>	
Distance Learners' Profiling, Learning Activities and Achievement Modelling	What do patterns of learners behaviour on course assessments mean for their	<ul style="list-style-type: none"> <li>● Student identification</li> <li>● Program code</li> <li>● Course code</li> <li>● Industrial type</li> <li>● Position level</li> <li>● Years service</li> <li>● Educational level</li> </ul>	<ul style="list-style-type: none"> <li>● Learning Managements System (LMS/ uFUTURE/ CLASS)</li> <li>● Student Information</li> </ul>



	learning and achievement	<ul style="list-style-type: none"> <li>● Educational field</li> <li>● Student age</li> <li>● Number of assignments submitted</li> <li>● Number of tests attended</li> <li>● Course assessment marks</li> <li>● Course grade</li> </ul>	Management systems (SIMS/eRES)
Distance Learners' Profiling, Learning behaviour and Achievement Modelling	What do patterns of learners behaviour on course materials mean for their learning and achievement	<ul style="list-style-type: none"> <li>● Student identification</li> <li>● Program code</li> <li>● Course code</li> <li>● Industrial type</li> <li>● Position level</li> <li>● Years service</li> <li>● Educational level</li> <li>● Educational field</li> <li>● Student age</li> <li>● Numbers of courses materials downloaded</li> <li>● Course assessment marks</li> <li>● Course grade</li> </ul>	<ul style="list-style-type: none"> <li>● Learning Managements System (LMS/ uFUTURE/ CLASS)</li> <li>● Student Information Management systems (SIMS/eRES)</li> </ul>
	What do patterns of learners behaviour on group discussions mean for their learning and achievement	<ul style="list-style-type: none"> <li>● Student identification</li> <li>● Program code</li> <li>● Course code</li> <li>● Industrial type</li> <li>● Position level</li> <li>● Years service</li> <li>● Educational level</li> <li>● Educational field</li> <li>● Student age</li> <li>● Number of discussions involvement</li> <li>● Course assessment marks</li> <li>● Course grade</li> </ul>	

**Distance Learners' Profiling and Achievement Modelling** can demonstrate the association analysis between the Distance Learners' previous working experience, previous education, age and the achievement. It may help the Institution Management in reviewing the Distance Learning Programme and the courses desirable to the potential students especially to the Adult/working Distance Learners. It may help the Institution Marketing and Promotion Department to prepare their business strategic plan.

**Distance Learners' profiling, Learning Activities and Achievement Modelling** can demonstrate the patterns of course assessment and the learning and achievement. This pattern can assist the lecturers to tailor educational opportunities to each student's level of need and ability. The student's achievement performance can be monitored and predicted. Potential issues also can be spotted early so that interventions can be provided to identify students at risk of failing a course or program of study. A predictive model can track the students' progress and make predictions about his/her future behaviours or performance, such a future course outcomes and dropouts.

**Distance Learners' Profiling, Learning Behaviour and Achievement Modelling** can demonstrate the learning behaviour on course materials and group discussions mean for their learning and achievement. Online learning systems log student data while downloading course materials and group discussions can provide data to lecturers to help them diagnose student learning issues. Does more online interaction between the students and the lecturers increased learning interest and improved the student's achievement performance? What sequence of topics is most effective for a specific student? What features of an online learning environment lead to better learning? What will predict student success?

### **Conclusion and Recommendation**

The need for quick and ready data is crucial in today's fast race society. In the education environment, the data mining strategy can be used to extract raw educational data from the readily available data in existing LMS databases to determine new learning preferences. By further extracting and manipulating existing data from LMS databases, education officers are able to get new information to cultivate new learning directions and strategies efficiently as they do not have to collect new data. Data mining can be used as a tool to extract the huge educational raw data to turn them into useful information.

With appropriate data mining analysis and interpretation of this distance learners' attributes from distance learners' profile, distance learners' learning activities and distance learners' learning behaviour; educators can recognize the capabilities of students, their interests and shortcomings and make decisions that positively optimizing students learning achievement using learning analytics. In addition of this, the data mining analysis from distance learners' attributes also can help the Learning Institution Management on reviewing the academic programme, marketing and promotion, and business strategic plan.

This study can be used as a guidance by Academic Administrators in learning institutions to fully utilize the Learning Management System (LMS) and other information systems related to distance learners to optimize their learning achievement. The framework of proposal distance learners' attributes and the model of application areas of the attributes for educational data mining and learning analytics which being developed in this study can be referred by the Academic Administrator to generate insight distance learning data analysis objectives.

Researchers and practitioners are recommended to conduct more studies on distance learner's data mining especially on adult's distance learners to further exploit its potential of acquiring more data and information.

### **Acknowledgement**

The authors thank the Research Management Institute, Universiti Teknologi MARA, for granting financial assistance to make this research possible.

## References

- Aher, S. B., & Lobo, L. M. R. J. (2011, March). Data mining in educational system using weka. In *International Conference on Emerging Technology Trends (ICETT)* (Vol. 3, pp. 20-25).
- Ahmadi, F., & Ahmad, M. S. (2013). Data mining in teacher evaluation system using WEKA. *International Journal of Computer Applications*, 63(10).
- Amran, N., Bahry, F. D. S., Anwar, N., & Yu, H. (2016). Educational Data Utilization by the School Administrators to Improve the Students' Academic Performance. *International Journal of Academic Research in Business and Social Sciences*, 6(11), 474-485.
- Amran, N., Mohamed, H., & Bahry, F. D. S. (2018). Developing Human Resource Training Management (HRTM) Conceptual Model Using Entity Relationship Diagram (ERD). *INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS AND SOCIAL SCIENCES*, 8(12).
- Ayyoob, M. P. (2015). Tools Pros and Cons of Clustering Algorithms using Weka Tools. In *International Journal of Computer Applications: National Conference on Advances in Computing Communication and Application*, pp-13-16.
- Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology*, 1, 1-57.
- Burns, M. (2011). Distance education for teacher training: Modes, models and methods. *Burns.–Washington: Education Development Center Inc.*
- Coronel, C., & Morris, S. (2019). 13rd Ed. Database Systems. Design, Implementation and Management. Cengage Learning
- Educational Data Mining. (2020). Retrieved at <https://educationaldatamining.org>
- Encyclopedia Britannica. (2020). Retrieved from <https://www.britannica.com/>
- Fauzi, M., & Iga, S. (2020, February). Enhance the Institution Reputation with Open Learning System-Based Distance Learning Development Prototype in Globalization Era. In *2nd International Conference on Islam, Science and Technology (ICONIST 2019)* (pp. 231-239). Atlantis Press.
- Joshi, A. P., & Panchal, R. K. (2014). Data mining for staff recruitment in education system using WEKA. *Int. J. Res. Comput. Sci. Manag*, 2(1).
- Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and prediction based data mining algorithms to predict slow learners in education sector. *Procedia Computer Science*, 57, 500-508.
- Keegan, D. J. (1980). On defining distance education. *Distance education*, 1(1), 13-36.
- Lewis, D., Madison-Harris, R., Muoneke, A., & Times, C. (2010). Using data to guide instruction and improve student learning. *SEDL Letter*, 22(2), 10-12.
- Mining, T. E. D. (2012, October). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. In *Proceedings of conference on advanced technology for education* (pp. 1-64).
- Prabha, S. L. (2015). Dr. AR Mohamed Shanavas. Application of Educational Data mining techniques in e-Learning A Case Study. *International Journal of Computer Science and Information Technologies*, 6(5), 4440-4443.
- Romero, C., Romero, J. R., & Ventura, S. (2014). A survey on pre-processing educational data. In *Educational data mining* (pp. 29-64). Springer, Cham.
- Saini, P. (2014). Building a Classification Model for Enrollment In Higher Educational Courses using Data Mining Techniques. *arXiv preprint arXiv:1405.3729*.

Thakur, R., & Mahajan, A. R. (2015). Preprocessing and Classification of Data Analysis in Institutional System using Weka. *International Journal of Computer Applications*, 112(6).