



Application of Data Mining Techniques in Education: A Review

*Muhammad Muhammad Suleiman¹, Mukhtar Ibrahim Bello², Shilpa³, and Muhammed Kuliya⁴.

¹School of Rural Technology and Entrepreneurship Development,
Rano, Kano State Polytechnic, Kano Nigeria, muhddkd@gmail.com

²Department of Computer Science, School of Technology, Kano State Polytechnic,
Kano Nigeria, mibello2@gmail.com

³School of Computer Application, Lovely Professional University,
Phagwara, India shilpa.21819@lpu.co.in

⁴Department of Computer Science, Federal Polytechnic,
Bauchi, Nigeria, muhammedkuliya@gmail.com

Abstract

Data mining is the process that analyzes large data to find fresh and unknown information that increases industry productivity. In the field of trying, to discover novel and potentially suitable data, DM also be called KDD (Knowledge Discovery in Databases). Currently, DM has been introduced with educational settings, which is called EDM, it is an area of systematic analysis focused on the improvement of discovery approaches in the exclusive types of facts from academic situations and using this approach to effectively understood learners and the settings wherein they learn. EDM is emerging that focuses on analyzing educational data to develop models for improving experiences and efficiency in teaching and learning. Growing popularity in DM and the system of education is transforming educational data mining into a modern, rising research culture. Educational data mining involves removing hidden knowledge from vast educational datasets utilizing techniques and resources such as sorting, decision tree, clustering algorithms, etc. to create novel methods of knowledge exploration from educational databases, which is used for educational assessment which decision-making.

Keywords: EDM, Clustering, Knowledge Discovery in Databases (KDD), WEKA, Algorithms, Prediction

1. Introduction

The technique of innovation in various areas has led to a large amount of data being stored in a variety of formats, such as records, documents, archives, photos, sound, recordings, logical data, and several new information groups. To be more efficient, the information obtained from various applications needs an effective methodology for extracting information from massive stores. Knowledge Discovery in Databases (KDD) (Baker 2011), Regularly alluded to as information mining, targets finding valuable data from huge information assortments. DM fundamental capacities are the utilization of various strategies and calculations to find and concentrate examples of putting away data. DM and Knowledge Discoveries applications have a rich focus in dynamic and have

become a basic segment in various associations (Alshareef et al. 2020). DM techniques were presented into novel arenas of competences in Machine Learning, Statistics, Databases, Artificial Intelligence, computations, Pattern Reorganization, etc. (Brijesh Kumar Baradwaj 2011); (Hapter n.d.).

Current research is increasingly at the use of EDM. This recent emerging arena, called EDM is concerned with rising methods for discovering understanding from background information about instruction. EDM uses several approaches, including Decision Trees, Naïve Bayes, Neural Networks, K- Nearest Neighbor, and numerous others. You can discover other classes of knowledge using other tools, such as classifications, association rules, and clustering (Romero et al. 2008). The awareness revealed can be

utilized to predict student enrollment in a specific course, to alienate the traditional teaching model in the classroom, to detect unfair means used in online examinations, to identify unique value systems in learner's result sheets, and to anticipate student performance, etc. (Brijesh Kumar Baradwaj 2011); (Manjarres, Sandoval, and Suárez 2018).

Data Mining in the field of computer science that seeks to identify various potential factors and patterns that will help make decisions. Data mining focuses on various fields such as machine learning databases and machine intelligence (known as artificial intelligence) which are very rapidly developing fields (Brijesh Kumar Baradwaj 2011); (Hapter n.d.); (Manjarres et al. 2018); (Ashraf 2014). One of the key goals in all educational settings is to guarantee those study mechanisms allow learners to comprehend their learning paths. This is where DME or EDM for its English acronym is of fundamental value to colleges and universities and to all organizations that offer multiple learning processes. (Manjarres et al. 2018).

Given the immense contribution that technology has made to all professions nowadays, it could be stated that all stages of life are changing, in particular the way we interact and the obstacles that have fallen to net support and the scale of computing that has been acquired and stored in different information technology. It is not unfair to say that the world is changing disappointingly with all this. Decision-making in this transition is an action that is not anticipated, particularly when such decisions are crucial to the future of establishments. (Brijesh Kumar Baradwaj 2011); (Manjarres et al. 2018).

An Educational Data Mining system needs to focus on collecting, archiving, and analyzing student learning and evaluation data. An Educational Data Mining system in the current scenario is a very new and very small academic field. EDM has grown out of existing disciplines as with all new fields and is spreading to overlap with new ones. Many of the researchers who shape EDM comes from the community of the Intelligent Tutoring System (ITS), where ready access to large amounts of educational data makes EDM a logical way forward. EDM research shares some commonalities with the community of Artificial Intelligence in Education (AIED) (Brijesh Kumar Baradwaj 2011). The analysis carried out in the EDM research is often related to psychometric and educational statistics techniques. EDM is poised to revolutionize or at least improve and extend the predictive approaches used in education by taking the findings of decades of data mining and machine learning work to fruition (Brijesh Kumar Baradwaj 2011); (Hapter n.d.).

2. Research objectives

- a) To identify the potential for application of DM in educational settings.

- b) Identifying data mining techniques in the education sector.
- c) identifying which data mining technique is more appropriate in education

3. Review of related literature

In this text we have investigated those papers wherein, in any event, one information mining procedure has been utilized to break down or take care of an issue related with an instructive domain, for which the text establishes in indexed lists made in Scopus.com were evaluated and GoogleScholar.com (Manjarres et al. 2018) related with "Instructive Data Mining" that had, at any rate, one arrangement that was not from a similar creator and that had a decent working structure. Those works that didn't consider the meaning of which procedure they had utilized were not considered by their substance (Manjarres et al. 2018).

Verma et. al. (2012) recommended utilizing DM methods, for example, grouping, choice tree, and relationship to improve the exhibition of understudies in advanced education. The staff may utilize the yield of the procedure, as a specialist to introduce the most suitable courses to understudies. Instructive course organizers can likewise utilize it to have further developed techniques on understudy course arranging (Srivastava and Srivastava 2013). He and Wu (2013) look at the educating and learning forms understudy conduct by utilizing the Live Video Streaming (LVS) procedures of information mining and content mining. They found that understudies were utilizing visit messages to pass on positive feelings, negative feelings, and social help articulations. There is no positive relationship(Srivastava and Srivastava 2013).

By applying white box arrangement strategies, Carlos, Cristobal, and Sebastian examined the information of 670 center school understudies and anticipated the drop-out and disappointment rate with a precision of over 90% (Walte et al. 2014). Jing talks about different potential applications including advertising, graduated class raising money, endurance investigation, constancy, and numerous others in the absolute first papers on applying Data Mining procedures to Luan Higher Education (Walte et al. 2014). Pandey and Pal (2011) endeavor to make sense of who is a specialist teacher working with understudies by stepping through a psychometric examination that changes subjective factors into quantitative and further use of affiliation rules Kabakchieva applied chosen information-digging calculations for grouping on Bulgarian college test information uncovered that the forecast outcomes are not surprising and that the classifiers work (Brijesh Kumar Baradwaj 2011); (Walte et al. 2014).

Naeimeh Delavari, Somnuk Phon-Amnuaisuk (2008), spoke about different applications of EDM. They

studied significant applications, i.e. predicting the performance of students, which aids institutes predict the final grade of students and their success in the course (Kaur 2015). M.ohammed M. Abu Tair and Alaa M. El-Hales were aiming to use EDM to advance the routine of the student. In the case they followed, a student knowledge archive of approximately 15 years was created, and data mining strategies were applied to it to discover interaction, clustering, and outlier identification. The paper demonstrated how the use of data mining could boost student efficiency. Naive Bayesian classification techniques and Rule Induction were adopted for classification. K means and outer detection methods were also used to form clusters, including distance and density-based approaches (Walte et al. 2014).

Dr. Mohd Maqsood Ali (2013) has addressed applications such as admission control, predicting monitoring of students, and predicting learner success in the area of education with specific methods. This method supports to identify learners and their output (Kaur 2015). M. Santhi Swaroop, K. Venkat Raju (2013) has given various applications to different DM applications, for example, social insurance DM applications, advertise bin investigation DM, current training framework DM applications, and CRM information mining applications. They said that this audit would assist analysts in concentrating on DM issues (Kaur 2015).

Kita Jain, Vishal Srivastava (2013) described the concept of data mining and analyzed that DM is presently a unique and novel area of research and that ANN itself is very suitable for analyzing DM problems due to its improved robustness, agile self-organization, parallel computing, distributed storage and a high degree of fault-tolerant. They found that these methodologies are increasingly focused on implementations in manufacturing, education, and sciences (Kaur 2015).

Ms. M. S. Mythili, Dr. A. R. Mohamed Shanavas (2014) provided student output analyzes using various algorithms (J48, decision tree & random forest). They establish that random forest efficiency is better than that of other algorithms used in the post-implementation process. Random forest algorithm has the highest precision to measure student achievement (Kaur 2015). Dr. Varun Kumar, Anupama Chadha (2011) has researched numerous data mining techniques in education such as Syllabus Management, Identification of Cheating in Electronic Test, and student success review. They researched various methods used in such applications (Kaur 2015).

4. The Concept DM

DM is a practice of collecting valuable information from vast quantities of data and trends (Srivastava and Srivastava 2013). It is also termed the method of information creation, the study of information/patterns, or the mining of computer knowledge (Manjarres et al. 2018);(Kaur 2015). The key aim of this research method is to take data from a dataset and turn it into a structure that is accessible and useful for future use. Data Mining is used by analyzing the data existing in the databases to solve problems (Manjarres et al. 2018);(Kaur 2015). Data mining is an essential procedure where different smart techniques are applied to extract useful patterns. Five (5) key elements of data mining Include:

- a) Extract, transform and load transaction data into the storage of data.
- b) Storage and then control of these data in a multidimensional relational database.
- c) Offer access to the data to IT practitioners and market analysts.
- d) Data processing using computer tools.
- e) Way to present the data in a meaningful form, such as a table or a chart; (Hapter n.d.); (Manjarres et al. 2018); (Kaur 2015).

5. EDM (Educational Data Mining)

With the advancement in 2008 of the annual International Conference on EDM and Journal of EDM, EDM has grown as a solid research field. The International EDM Society conducts the EDM Annual Conference and publishes the EDM Journal (Follow 2020);

"Educational data mining is an emerging discipline, concerned with developing methods to explore unique and increasingly large-scale data from educational settings and using these methods to better understand students and the settings in which they are learning" (International EDM Society, 2011) (Follow 2020); (Algarni 2016).

According to the International Society for EDM (2011), relevant data is often provided at different levels of hierarchy in any learning context, which cannot be calculated but it must be verified by the characteristics derived from the data. It is also important to take into account resources such as cost, sequence, and context in the study of data on education. For example, learners' learning patterns (involvement of learners, duration of login, amount of conversations, and type of questions sent to an instructor) can be evaluated along with their final grades.(Follow 2020); (Algarni 2016).

5.2 EDM vs Learning Analytics

Researchers may not have a specific definition for utilizing EDM methods since EDM is strongly related to the Learning Analytics (LA) analysis sector (Follow 2020).

“Learning analytics is tracking, gathering, analyzing and reporting data on learners and their experiences to identify and improve learning and the situations in which it happens (International Learning Analytics and Intelligence Conference et al . 2016)” (Follow 2020).

Learning Analysis (LA) is another viewpoint, which is getting progressively popular. LA is on the dual properties: that LA handles pre - set up facts in a structure that a PC can deal with, and that its strategies can be adequate for taking care of large data sets, which would not deal with it physically. (Ferguson 2012) Research people group in the LA and EDM have alternate points of view on the best way to acquire and investigate educational data (Berland et al., 2014).

Studies in LA study use added human-led strategies of experimentation, rely primarily on relational processes and knowledge structures, and instead explore opportunities to educate and inspire teachers and learners; for example, educating a teacher on whether a single student suffers so that the teacher may approach the learner and engage actively to promote the learning of that student. But in the other side, EDM specialists center more around utilizing computerized strategies for disclosure inside educational data, demonstrating explicit builds and their relationships, applications in mechanized adjustment, for example, supporting the experience of a student by distinguishing educational programming and altering it naturally to customize the experience of the student (Corbett & Anderson, 1995; Baker et al. 2006; Arroyo et al., 2007 and Berland et al. 2014) referred to in (Follow 2020).

5.3 Goals for using EDM

In pedagogical contexts, the introduction of new machinery and public records repositories increases the data quantity and diverges data excellence. Mobile devices, for example, can permit the scientists to detention learner collaboration in more features (Srivastava and Srivastava 2013); (Follow 2020); (Algarni 2016). Below are the main targets set for using EDM:

- a) Predicting behaviors of the learner by improving models of the students. Modeling characterizes and categorizes the features of a student, or states that represent the knowledge, awareness, inspiration, motivation, meta-cognition, and attitudes of the student.
- b) The creation or development of concept systems in the area of science. There are concept models of the materials being taught, for example, and

models that illustrate the interrelationships of knowledge inside space.

- c) To research the most successful pedagogical resources for the learning of students that can be accomplished across learning programs.
- d) Establishing scientific proof to help or articulate instructive speculations, structures, and educational wonders to distinguish center persuasive parts of learning with the goal that better learning frameworks can be planned (Srivastava and Srivastava 2013); (Follow 2020).

5.4 EDM stakeholders

Relevant stakeholder groups review educational information from a range of contexts, conform to their mission, priorities, and criteria for using EDM (Hanna, 2004). Below are the four (4) main stakeholder groups according to their goals for the use of EDM.(Follow 2020):

- a) **Students:** Suggestion or Optimization of personal learning patterns, learning experiences, and learning materials.
- b) **Educators:** Analyzing the learning behaviors of the students, getting the most supporting teaching as well as predicting studies to increase the efficiency of teaching.
- c) **Developers/Researchers:** Evaluation of study materials, enhancement of learning schemes, and evaluation of effective DM techniques.
- d) **Institutions:** Improve Processes of decision-making in educational institutions in terms of productivity and expenditure, such as enrolment and allocation of financial capital (Follow 2020).

5.5 EDM Application

Numerous studies were created in the EDM area. A system has been suggested by Alexandro and Georgios cited in (Algarni 2016) for probing learner performances in online instruction videos.

The proposed system comprised of catching data on student execution, structuring a data model to store action data, and making modules for checking and picturing conduct concerning the student utilizing caught data. Scientists depended on a large portion of the understudies to watch recordings in the days paving the way to assessments or due date for the task. What is more, delays and continues were watched for the most part in recordings related to a task. Another lamentation was that the creator didn't examine what affected the conduct of the student or why a few understudies would not watch online recordings (Algarni 2016).

In other examinations, he built a model utilizing data mining philosophies to anticipate which understudies are probably going to drop out in a college program during their first year. That review utilized the classification calculation, Naive Bayes, to develop the forecast model dependent on the present data. The

framework result was promising in recognizing understudies who required particular regard to lessen the dropout rate. Leila Dadkhahan cited in (Algarni 2016) attempted to explain what was required for understudy achievement in foundations of advanced education to-dropout figures. Therefore, the use of data mining strategies added to higher maintenance and graduation levels for graduates (Alshareef et al. 2016).

5.6 Data Mining Applications in Education

A wide number of educational data mining applications have been identified, as has been reflected throughout the whole chapter. This segment explores four (4) application areas that have established careful attention inside the segment. (Baker 2011):

- a) **Predicting Student Performance:** Predicting Student Performance: In expectation of understudy results, we foresee the obscure estimation of a variable characterizing the understudy. The qualities for the most part anticipated in the instruction division are the presentation of understudies, their imprints, information, or score. The strategy of classification is used to consolidate singular things dependent on quantitative qualities or dependent on preparing a set of recently labeled things. The exhibition expectation of students is extremely popular utilization of DM in the training area (Bala and Ojha 2012). Various procedures and models are used to foresee the presentation of understudies, for example, choice trees, neural systems, rule-based frameworks, Bayesian systems, and so on. This analysis is helpful to someone in foreseeing the exhibition of the understudy for example foreseeing the accomplishment of the understudy in a course and anticipating the last grade of the learner dependent on highlights taken from logged data. (Hapter n.d.); (Kaur 2015); (Algarni 2016); (Alshareef et al. 2016).

Numerous regression methods are used to forecast student marks, for example, straight relapse to conjecture understudy scholastic yield, step-wise direct relapse to anticipate understudy time spent on a web site, and numerous direct relapse to classify factors that help foresee school accomplishment and anticipate test results in separation learning (Kaur 2015); (Algarni 2016).

- b) **Visualization and Analysis of Data:** It is utilized to highpoint useful data and to help make decisions. For example, in the education field, it may be useful for course managers and educators to evaluate the use details and the behaviors of the students during the course and get a quick understanding of the learning of a pupil. Visualization and statics knowledge are the two

main methods to use for this project. Statistical analysis of instructional data will provide educators with information such as when students enter and leave, the most important sites students access, how many updates of eLearning content, how many different types of sites are browsed, and how much time these various pages take to be browsed (Kaur 2015); (Algarni 2016).

It also gives data on reports on scheduled and daily user tendencies, usage premises, how considerably material learners determination analyze and the sequence in which they study topics, study activity patterns, positioning, and activity sequencing. Visualization allows the use of graphic tools to help people interpret and evaluate results. There are numerous studies related to the visualization of various educational data, such as hourly, daily, seasonal, and annual user behavior patterns on online forums (Kaur 2015); (Algarni 2016).

- c) **Grouping Students:** In this situation, member classes are formed according to their unique characteristics, specific characteristics, etc. The instructor/developer will use these clusters/groups of students to create a customized learning framework that can facilitate successful community learning. Classification and clustering are the DM strategies used for this mission. Specific clustering algorithms used for grouping students include hierarchical clustering, K-means, and model-based clustering (Kaur 2015); (Algarni 2016); (Alshareef et al. 2016).

A clustering algorithm is characterized by large generalized sequence data, which help to find groups of students with similar learning attributes, such as hierarchical clustering algorithms, which are used in intelligent e-learning systems to a group of learners according to their learning style (Kaur 2015); (Alshareef et al. 2016).

- d) **Enrollment Management:** In tertiary education, it is frequently used to explain well-planned strategies and ways to shape college enrollment to meet planned objectives. It is a concept of organization and a systematic set of activities designed to enable educational institutions to exert more influence on student enrollments. These activities also require maintenance programs, publicity strategies, granting financial assistance, and acceptance strategies. (Kaur 2015); (Algarni 2016); (Alshareef et al. 2016).
- e) **Predicting Profiles of the Students':** DM can allow management to classify students' socioeconomic, regional, and psycho-graphic characteristics based on the details the students offer at the time of entry. The neural networking

technique can be used to identify students of different categories (Kaur 2015).

- f) **User Modelling:** It includes what a learner knows, what is the user interface, what is the attitude and motivation of a learner, and how satisfied users are with online education. EDM may be used to model user awareness, user behavior, and user interaction (Kaur 2015).
- g) **Planning and Scheduling:** It is utilized to improve the conventional educational cycle through preparing potential classes, course preparation, resource distribution planning, and assists in enrollment, and advisory processes, curriculum creation, etc. Classification, categorization, estimation, and visualization are diverse DM techniques utilized for this job. Decision trees, relation analysis, and decision forests were used in course scheduling to analyze the enrollment preferences and the completion of courses in extension courses. Educational training programs were planned with bunch analysis, neural networks, and decision trees of back-propagation to find a correlation among the training classifications. Bayesian models and Decision trees were suggested to help managing organizations investigate the likely effects of changes in recruitment, admissions, and courses (Kaur 2015); (Alshareef et al. 2016).

- h) **Organization of Syllabus:** Currently, syllabus organization is affected by various considerations such as associated, competitive, or partnering university systems, instructor capacity, professional judgment, and expertise. This organizational approach can not automatically promote the cognitive ability of the students optimally (Kaur 2015); (Bala and Ojha 2012). Exploration of topics and their interactions will specifically assist in improved syllabus structure and offer insights into current educational curricula. One of the data mining applications is to classify similar topics in the syllabus of education programs in a broad educational institution (Kaur 2015).
- i) **Detecting Cheating in Online Examination:** Often electronic reviews during a day are done directly over the Web and if a swindle happens then one of the simple challenges to overcome knows: who is there? Students however the latest industry and media controversies indicate that it has been a normal occurrence, should not only lie. Techniques for data mining can recommend models that can help establishments detect and avoid cheats in web-based assessments. The models created to use data consisting of various student identities, tension situations produced by online assessments, and typical strategies used by students to cheat and achieve a better grade on such examinations. (Kaur 2015).

6. EDM Methodology Categories

Table 1: EDM Methodology categories (Hapter n.d.); (Algarni 2016); (Walte et al. 2014)

Categories	Key Applications	Objectives
Prediction	Identify students at risk. Understand educational outcomes for students	Develop a model for predicting certain variables that are based on other variables. The variables of the predictors that be unchanged or derived from the data collection.
Clustering	Find parallels between students or classes, and variations. Categorized new actions for the teachers.	Assign similar amounts of data to various groups depending on the system characteristics. The number of clusters will differ based on the model and clustering method objectives.
Relationship Mining	Consider the connection between rates of parent education and dropping-out pupils. Discovery of curriculum correlations in sequences of courses; figuring out which pedagogic approaches contribute to more effective/robust learning.	Extract the relationship in Datasets between two or more variables
Discoveries with Models	Discoveries of associations across student activities and attributes of students or social variables; study problem review through a broad range of backgrounds.	This suggests a framework of a phenomenon that uses clustering, visualization, or knowledge engineering as an element of a further detailed prediction prototype or mining collaboration.
Data Distillation for Human Judgement	Human pattern recognition in learning outcomes, actions, or collaboration; marking of expectation model data for utilization in future development.	This model's main objective is to discover a novel way for researchers to accurately recognize or define features in the data.

7. Data used through EDM

EDM provides a good view of, and deeper understanding of, learners and their learning patterns. DM methods were used to analyze educational data and to address educational issues. Like other extraction methods using DM techniques, EDM removes important, interoperable, valuable, and novel knowledge from the educational statistics. Nevertheless, EDM is primarily worried about new methods of exploring special data forms in educational settings. These approaches are used to increase awareness of educational events, teachers, and the environment in which they are studying. The development of computational approaches to combining data and philosophy will contribute to improving the excellence of T&L practices (Algarni 2016). The through the usage of knowledge in education systems has made accessible a tremendous data. EDM provides an amount of useful knowledge. Hence, the main data source utilized in EDM to data can be classified as follows;

- 7.1 Online schooling, also recognized as formal education, is when a face-to-face interaction passes information to learners. Common approaches such as evaluation and questionnaires may capture the data. It tests students' cognitive capacities and decides how they think. Therefore, the data can be applied to statistical techniques and psychometrics.
- 7.2 Intelligent Tutoring Systems (ITS) and Adaptive Educational Hypermedia Systems (AEHS) aim to configure the students' data based on user's profiles. Therefore, it is essential to apply data mining techniques to create profiles of the user. Then the data produced by that program may aid in more research (Algarni 2016).
- 7.3 LMS and E-learning offer content, training, communication, and monitoring resources that allow students to study on their own. DM strategies can be extended to the data contained in the repositories by the programs.

On the basis of the three categories identified by Romero et al. quoted in (Algarni 2016), we may group EDM study by category of data used: formal education, online education (e-learning), LMS, smart tutoring systems, adaptive educational systems, evaluation questionnaires, quality of documents, and so on. (Algarni 2016).

8. EDM Techniques

- 8.1 **Clustering Algorithm:** It is a separation of information into identical groupings of data. Clustering plays an important part in DM areas such as text mining, data recovery, and

systematic data exploration, online analysis, application of spatial DB, marketing, and more (Alshareef et al. 2016).

The clustering of data is a technique of unverified and arithmetical analysis of data. It is utilized to divide the same data into a consistent community of elementary learners it is utilized to work on a broad data collection to find secret trend and interaction allows making fast and effective decisions. Cluster analysis is used to segment a broad data collection into subsets called clusters. Each cluster is an array of data items similar to each other. We are positioned within the identical cluster but vary in certain clusters from artifacts. In Clustering, algorithms follow are used in education mining. (Srivastava and Srivastava 2013); (Alshareef et al. 2016).

- 8.2 **Classification:** It is the analysis of the data method that abstracts models, which characterize vital classes of data. This method also uses algorithms for the classification of decision-trees. The method for classifying data includes studying and classifying. In the learning system, the classification algorithm analyzes the training data sets. Reference datasets are utilized in grouping to assess the consistency of the classification laws. If the precision is suitable, the rules may extend to the latest data tuples (Srivastava and Srivastava 2013); (Bala and Ojha 2012); (Alshareef et al. 2016); (Johina and Kamra 2015).

- 8.3 **K-Mean Clustering Algorithm:** It is among the popular Data Mining clustering algorithms. K-Means is a non-hierarchical clustering approach intended to separate data into one or more clusters. This approach splits data into clusters such that data with the same functionality are clustered into one group and data with various traits is clustered into a cluster (Alshareef et al. 2016).

K-Means Clustering (KMC) proposes partitioning items into K-clusters in which each object belongs to the cluster with the nearest mean. Using this process, accurately K distinct clusters with the greatest possible characteristics have been developed. Initially the greatest quantity of clusters k leading to the highest (distance is uncertain and must be calculated from the results). The purpose of K-Means clustering is to reduce the role of squared error or total variance in the intracluster (Alshareef et al. 2016).

Let $X = \{x_1, x_2, x_3, \dots, x_n\}$ be the set of data points and $V = \{v_1, v_2, \dots, v_c\}$ be the set of centers.

- a) Choose 'c' cluster middles randomly.
- b) Compute the distance matrix and the cluster center.
- c) Assign a piece of data to a cluster middle where the distance from the cluster is a lower limit of entire cluster middles.
- d) Recalculate the novel cluster middle by using the following:

$$v_i = \left(\frac{1}{c_i} \right) \sum_{j=1}^{c_i} x_j$$

Where 'ci' refers to the data points in the cluster

- e) Recompute the gap (distance) among each data set as well as the new cluster centers, which have been obtained.
- f) If no data point has also been reassigned then pause, otherwise reiterate move from step (3) (Alshareef et al. 2016).

8.4 ID3 Technique - Terms utilized in ID3 Procedure:

- a) Establish the attribute of Classification
- b) Calculate Entropy Classification.
- c) Measure the information gain by using the grouping attribute.
- d) Choose the element with the biggest gain to become the next level in the tree (starting from the root node).
- e) Remove the Node Attribute, create a reduced table (Alshareef et al. 2016).

8.5 Dimensionality Reduction Techniques:

Dimensional reduction is the most critical tool for removing unnecessary attributes and vibration and can be further categorized through the isolation and selection system of software. The well-known algorithm COBWEB creates the student clusters. The clusters of students are established dependent on the credits earned from the term granted. As previously specified, the group utility is defined as the attribute's weighted distance utility (i.e. credits). To mine the output data of the learners, the data mining classification techniques such as – Decision Tree- Random Tree & J48 Classification Models were developed using the WEKA tool with 10 Cross-Validation Fold (Alshareef et al. 2016).

9. Benefits of DM In education

The students' usage of DM methods is to get better resources for their job counseling. Educational data mining will enable the student as well as the management to improve their educational standard. DM with a student, which implies it, includes the student-specific details such as name, age, class, course, address, etc. (Priyadarshini and Ray 2017). It

also helps the students learn better and improve their educational process. The DM technique can aid in student academic achievement improvements. It also enhances Web-based education systems (Priyadarshini and Ray 2017)(Chopra et al. 2018).

10. Future research

As we know, there are a variety of classification and clustering strategies that are used for knowledge of data mining. Many algorithms have performance and computation time issues from all such implementations. Some algorithms have a very long processing time and a low accuracy, which is a big problem. With the development of educational applications, algorithms with greater accuracy, and less processing time need to be established more strongly. There is a lot of space in this field to develop new algorithms. Investigators are doing a lot of work in this field.

11. Conclusion

The expanded use of innovation in education produces a great deal of daily data, and has been a priority for a lot of academics around the globe; the area of EDM is moving exponentially and has the benefit of incorporating innovative techniques and algorithms built in various fields of Machine Learning and DM. As mentioned above, various scholars' explored specific applications of DM in the area of education, for instance, student success analysis, electronic test hacking, preparation and scheduling, student grouping and enrolment management, etc.

12. Acknowledgment

First of all, I share my heartfelt debt of appreciation to the Almighty God, who always helps me in my endeavours. I would like to thank my teachers for their support and encouragement. Then I want to say thank you to my family and friends. I am grateful to all those who helped me at every stage in my life, in one way or another.

12 References

- Algarni, Abdulmohsen. 2016. "Data Mining in Education." *International Journal of Advanced Computer Science and Applications (IJACSA)* 7(6):58–77.
- Alshareef, Fatima, Hosam Alhakami, Tahani Alsubait, and Abdullah Baz. 2016. "Educational Data Mining Applications and Techniques." *International Journal of Latest Trends in Engineering and Technology (IJLTET)* 7(2 JULY):484–88.
- Alshareef, Fatima, Hosam Alhakami, Tahani Alsubait, and Abdullah Baz. 2020. "Educational Data Mining Applications and Techniques." *International Journal of Advanced Computer*

- Science and Applications* 11(4):2928–30.
- Ashraf, Imran. 2014. "Data Mining Algorithms and Their Applications in Education Data Mining." *International Journal of Advance Research in Computer Science and Management Studies* 2(JULY 7).
- Baker, Ryan S. J. 2011. "Data Mining for Education." *Encyclopedia of Data Warehousing and Mining*.
- Bala, Manoj, and D. B. Ojha. 2012. "Study of Applications of Data Mining Techniques in Education." *International Journal of Research in Science And Technology* 1(4).
- Brijesh Kumar Baradwaj, Saurabh Pal. 2011. "Mining Educational Data to Analyze Students' Performance." *International Journal of Advanced Computer Science and Applications* 2(6):59–63.
- Chopra, Shivangi, Lukasz Golab, T. Judene Pretti, and Andrew Toulis. 2018. "Using Data Mining Methods for Research in Co-Operative Education." *International Journal of Work-Integrated Learning* 19(3):297–310.
- Follow, Jesse Tetsuya. 2020. "Why Is Educational Data Mining Important in the Research?" *Educational Data Mining Is an Emerging Discipline , Concerned with Developing Methods for Exploring the Unique and Increasingly Large-Scale Data Obtained from Educational Settings , and Uses Those Methods to Better Understand Students and the Settings In* (Baker 2014):1–8.
- Hapter, C. n.d. "ANALYSIS OF EDUCATIONAL DATA MINING." *Book CHAPTER-3* 39–59.
- Johina, and Vikas Kamra. 2015. "A Review: Data Mining Technique Used In Education Sector." *International Journal of Computer Science and Information Technologies* 6(3):2928–30.
- Kaur, Hardeep. 2015. "A Review of Applications of Data Mining in the Field of Education." *International Journal of Advanced Research in Computer and Communication Engineering* 4(4):409–12.
- Manjarres, Andrés Villanueva, Luis Gabriel Moreno Sandoval, and Martha Janneth Salinas Suárez. 2018. "Data Mining Techniques Applied in Educational Environments: Literature Review." *Digital Education Review* (33):235–66.
- Priyadarshini, Nandita, and Mitrabinda Ray. 2017. "A Review: Data Mining Techniques in Education Academia." *CEUR Workshop Proceedings* 18(19).
- Romero, Cristóbal, Sebastián Ventura, Pedro G. Espejo, and César Hervás. 2008. "Data Mining Algorithms to Classify Students." *Educational Data Mining 2008 - 1st International Conference on Educational Data Mining, Proceedings* 8–17.
- Srivastava, Jaya, and Dr Abhay Kumar Srivastava. 2013. "Data Mining in Education Sector: A Review." *Special Conference Issue: National Conference on Cloud Computing & Big Data* 184–90.
- Walte, Dipesh, Hari Reddy, Vivek Ugale, and Amol Unwane. 2014. "Overview of Algorithm ED Mining for Higher Education : An Application Perspective." *Nternational Journal of Engineering Research & Technology (IJERT)* 3(2):521–23.

Author's Biography



About the Author:

Is currently a Lecturer I in the School of Rural Technology and Entrepreneurship Development, Rano, Kano State Polytechnic, Nigeria. He obtained a Nigerian Certificate in Education (NCE) in 2004 from Federal College of Education (Technical.), Bichi, and International Diploma and Advance Diploma in 2012 and 2013 from Informatics Academy, Singapore. He also has undergone Bachelor of Education (Business

Education) from Ahmad Bello University, Zaria and Bachelor of Information Technology from International University of East Africa, Kampala, Uganda in 2012 and 2014 respectively, currently a graduate of Master of Science in Information Technology in the year 2020 from Lovely Professional University, Phagwara, India. He is a member of several professional bodies such as Teachers' Registration Council of Nigeria (TRCN), Indian Science Congress, Kolkata (ISC), Association of Business Educators of Nigeria (ABEN), Business Educators Association in Vocational Education (BEAVE). His research interests' areas are ICT, Education, Data Mining, Information Security and Cloud Computing. The author published articles in several Journals and attended conferences in both National and International respectively.