

# Application of Data Mining in Adaptive and Intelligent Tutoring Systems: A Review

Rajitha Koppisetty,

*B. Tech., Department of Computer Engineering,  
MPSTME, NMIMS University,  
Vile Parle West, Mumbai, India*

**Abstract** –Educational Data Mining focuses on the application of data mining techniques to educational data, in order to better understand students' learning behavior. Intelligent tutoring systems and Adaptive learning systems are computer systems that provide immediate and customized instruction or feedback to learners. The paper reviews the stakeholders in an educational system and the impact on each party involved. The paper studies in detail the algorithms Classification and Prediction, Clustering, and Association Rule Mining. The study concludes by providing a glimpse into potential of Educational Data Mining in terms of requirements and future scope.

**Keywords** – Educational Data Mining, Learning Management System, Classification, Clustering, Association Rule Mining, Student performance analysis

## I. INTRODUCTION

**D**ATA Mining deals with the extraction of information and patterns within a repository of data. It is an interdisciplinary field of computer science that analyses a large set of data to maximize or minimize an objective function. The applications of data mining in the field of education aim to solve educational problems. Educational Data Mining refers to the use of data mining techniques to educational data in order to analyze students' behavior and performance and resolve relevant educational research issues [1]. With the rise in competition and demand for quality education, allocation of educational resources to provide maximum benefit to the students is important. Educational Data Mining, in its results can aid all the stakeholders in an educational system to make informed choices regarding education, talent management, allocation of resources, etc.

## II. KEY FACTORS OF EDUCATIONAL DATA MINING

Educational Data Mining has the following key factors [1]:

### A. Objective

Depending on the varied stakeholders, the applications of educational data mined may have diverse uses. The results may be used not only to improve educational techniques and guide students in providing a better understanding of

the curricula but also to analyse the trends in education and predict the future demand. Educational institutions can use these findings to enhance the decision making process, improve success rate, analyse student behaviour, assist instructors, etc. [2]

### B. Data

With the evolution of new education systems such as E-learning, ITS, AHES, etc., the data generated is vast and varied. In a traditional educational institution, the student data may include GPA, subject wise grades, entrance exam results, etc. On the other hand, in a web learning environment, the data may be everything from the amount of time the student takes to study a particular subject to the student's aptitude in each module. Data mining techniques may vary depending on the type of data to be mined.

### C. Techniques

Traditional data mining techniques include classification, clustering, association rule mining, etc. Depending upon the data and the results required, the techniques are chosen. Not all techniques can be used on one data set without specific adaptation. For example, grouping of students into particular classes required clustering techniques, prediction of future results of a student in terms of academic performance uses classification, etc.

The following sections discuss in detail the following techniques:

- a. Classification and Prediction
- b. Association Rule Mining
- c. Clustering

The evolution of educational systems has resulted in various education systems being developed. Broadly, the education systems can be divided as follows:

### A. Offline/Traditional Education Systems

These education systems include schools and universities. The mode of instruction is a teacher imparting knowledge to a set of students. The student data generated is relatively less and consists of details such as student performance per subject over the years and cumulative performance. These systems are rapidly undergoing changes with the incursion of demand for quality education.

### B. E-learning/Web-learning System

These systems cater to students who for a variety of reasons cannot avail the use of traditional educational systems. E-learning systems make use of internet applications in order to impart knowledge to a large set of students irrespective of their location. This form of education is void of an instructor and largely depends on the student himself/herself. E-learning modules are highly instructional in nature as opposed to systems which promote cognitive and creative thinking.

### C. ITS/AEHS Systems

Intelligent tutoring systems (ITS) and Adaptive learning systems are computer systems that aim to provide immediate and customized instruction or feedback to learners, usually without intervention from a human teacher. They aim to reduce the dependency of students on teachers. An adaptive learning system focuses on adapting to the needs of a particular student. The system is customized to the needs of each student based on statistical analysis of student progress, based on which the system makes decisions to alter the teaching pathway adopted. Many intelligent tutoring systems today employ adaptive principles in operation, and hence the distinction between the two systems has been practically exterminated. While these systems are becoming more and more popular, they are highly expensive to build [3].

## III. ROLE OF STAKEHOLDERS

An education system consists of a variety of stakeholders as listed below [1]:

### A. Learners/Students

These are the primary users in an educational setting. In e-learning and ITS systems the progress of the student is monitored closely for every action taken. Students can be given immediate feedback on their performance and provided with suggestions for further study.

### B. Educators

Educators can use the results mined for adjusting curriculum and managing the teaching methodologies in a classroom. Predicting student performance and analysing repeated mistakes also helps instructors in making informed educational decisions.

### C. Educational researchers

Various educational research questions relevant today can be answered using data mining techniques. Researchers can analyse trends in education and predict future requirements to meet the rising demands.

### D. Universities/Administrators

Educational data mining is useful to universities and administrators in terms of allocation of resources. Universities can use the data analyzed to decide changes in course curriculum and create new course modules depending upon student aptitude or requirement. For example, monetary funds can be directed to facilities in a prioritized manner to improve efficiency of the institute as a whole.

## IV. APPLICATIONS OF DATA MINING IN ITS

### A. Analysis and visualization of data

Visualization of data allows for easy understanding of raw data. Data mined from an ITS consists of the time and frequency of access of each web page. This can be used in evaluating student performance per topic or structuring course material.

### B. Providing feedback to the instructors

Feedback is required for the instructors to make informed decisions regarding tackling the curriculum and course structure. Association rule mining can provide interesting relationships for analysis. Clustering, classification and association rule mining have often been employed to provide an indicator of the progress among students.

### C. Providing recommendations for students

An ITS system is capable of providing student with recommendations regarding the course material. This can be done by analysing the type of data studied by the user and providing suggestions based on the interest of the user or to maximize the user output. The most common methods to provide such recommendations are clustering, neural networks and decision trees.

### D. Predicting student performance

Predicting student performance enables the system to re-allocate timing and educational resources, provide suggestions, etc. to optimize the result. Classification and prediction algorithms such as Naïve Bayes algorithm, decision trees, etc. are used most often in such a scenario.

### E. Grouping Students

It is beneficial to group students into particular classes and cater to each class in a specific manner. This allows for separation of students and teaching each group in a particular manner. Such grouping of student is possible with the use of clustering algorithms such as k-means algorithm or estimation-maximization algorithm.

### F. Planning and Scheduling

Structuring of the course contents in a time efficient manner to yield maximum results is essential. Data mining the users past data and history can provide an efficient way to schedule and plan a course to cater to the student and bring out maximum efficiency. Classification, categorization and visualization are the widely used methods to obtain objectives such as academic planning and creating meaningful outcome typologies. Decision trees can also be used to maximize the course completion rate for a particular course.

## V. CLASSIFICATION AND PREDICTION

One of the primary goals of Educational Data Mining is to be able to predict the student performance. [4] Two forms of data analysis are used for extraction of information from classes and predicting trends in said data. [5]

Classification is a technique that is used to create predetermined classes of data that later has objects assigned

to them. For example, classification can be used to assign labels to each student entry in a database such as “Pass”, “Fail”, etc. Prediction means making an assumption based on the likelihood of an event occurring. In educational data mining it is used to determine the probability of a situation occurring with respect to an education scenario. For example, it is possible to predict if a student will pass or fail in a particular subject based on a set of characteristics such as past performance, student aptitude, GPA, etc. Both Prediction and Classification techniques work in two broad steps:

- i. Building a classifier model using a sample (assumed) data set.
- ii. Applying the classifier model to the required data set.

#### A. Naïve Bayes Algorithm

Naïve Bayes is a probabilistic classifier that uses the statistical method called Bayes theorem. This is one of the simplest forms of the Bayesian Networks utilized and has very satisfactory accuracy and sensitivity rates. [6] It considers two or more characteristics and calculates the probability of the event occurring given these characteristic's values. It provides a way of calculating the probability  $P(c | x)$  from  $P(c)$ ,  $P(x)$  and  $P(x | c)$ . The algorithm assumes that the effect of predictor  $x$  on  $c$  is independent of the values of the other predictors. The relation between them is [15] [4]:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

where,  $P(c|x)$  is the posterior probability,  $P(x|c)$  is the likelihood,  $P(c)$  is the class prior probability and  $P(x)$  is the predictor prior probability.

For example, if the question is to determine the likely hood of a particular student passing in a subject, say math, given the following attributes:

- i. GPA of the student (G)
- ii. Aptitude in math (M)
- iii. Work ethic of the student (H)
- iv. Prior performance in the subject

The probability of a student passing the subject given values  $G=g$ ,  $M=m$ ,  $H=h$  will be

$$F(g, m, h) = \frac{P(m, g, h, pass)}{P(g, m, h, fail)}$$

If the value is greater than 1, the classifier can predict pass. Otherwise the student is assumed to have failed. [4]

It should be noted that the algorithm assumes that all the characteristics considered are independent of one another. For example, if we were to determine the probability of a tomato ripening by considering the characteristics such as season, color and firmness, Naïve Bayes assumes that each of these characteristics are independent. However in many cases this is not true. Here, color and firmness have a relative dependence on one another.

Despite some disadvantages, Naïve Bayes is one of the most efficient algorithms in terms of simplicity and

calculations incurred. It manages to assign classes to data with a large number of objects with ease and efficiency.

#### B. Decision Tree Algorithm:

A decision tree works on the principle of “divide and conquer”. A decision tree works builds a classification model based on a tree structure. At the time of breaking down the data set into smaller subsets, the tree is incrementally constructed in order to build a classifier that takes decisions by following the branches of the tree to a leaf node. The leaf nodes (or terminating nodes) are the final decision or classifiers assigned to any object that participates in the process of a decision tree. The major advantage of a decision tree is the fact that it can handle both numerical and categorical data. It does not require a prior knowledge setting.

The tree follows a series of deciding rules (or questions generally in the “if, else” format) that run through several decision branches. These decision parameters are dependent upon the characteristics of the classes that are generated. For this purpose, the decision tree is first induced on a training set and then tested on the required data set.

The core algorithm followed in the paper for building decision trees called ID3 by J. R. Quinlan which employs a top-down, greedy search through the space of possible branches with no backtracking.

The ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has entropy of one. [13]

Entropy is an effective method to calculate the impurity degree. Among the other methods that can be used are Gini Index and Classification Error. [4] The paper has utilized the Entropy method to calculate the impurity degree.

For a classifier to be assigned to a particular object in the data set, a characteristic is considered. For example, if we are to find out if a particular student will pass in a given exam, we need to calculate the probability  $P(\text{pass})$  with two possible classifiers – pass and fail. The characteristics or attributes under consideration would be aptitude, attendance, GPA, etc. The decision tree is built calculating the following [13]:

- i. Entropy using the frequency table of one attribute:

$$E(T) = \sum_{i=1}^c -p_i \log_2 p_i$$

- ii. Entropy using the frequency table of two attributes:

$$E(T, X) = \sum_{c \in X} P(c)E(c)$$

- iii. Information Gain to determine the best attribute in for a particular node:

$$Gain(T, X) = E(T) - E(T, X)$$

This is continued for each branch with entropy that is not equal to zero. Each new attribute is selected and the partitioning is repeated to every non-terminal descendent node.

The algorithm stops when:

- i. Every attributes has been included in the path
- ii. The entropy of a leaf node is zero.

These algorithms were compared in order to evaluate their performance. It was found that the ID3 Decision Tree was 98% accurate as compared to Naïve Bayes' 94% and hence was the better choice to predict student performance [4].

## VI. ASSOCIATION RULE MINING

Association Rule Mining is a technique of data mining that searches for relationships or associations between data. These relationships are not based on the functional dependencies of the data but the repeated occurrence of pairs of data items [7]. Association Rule Mining employs principles of statistics in order to make relevant connections between the data elements of a database. In the educational scenario, data mining unearths patterns that influence student behavior, for example – if a student spends 'x' time on a particular web page, it is likely that the student is strong in 'm' subject.

The preliminary idea of Association Rule Mining is explained below [8]–

- i. I is the set of items in a data base having m number of items such that  $I = \{I_1, I_2, I_3, \dots, I_m\}$ .
- ii. Each transaction T performed on the database has an id (TID), for example – T1, T2, etc. Each transaction contains a set of items such that  $T \subseteq I$
- iii. If A be a set of items, then transaction T is said to contain A, if and only if  $A \subseteq T$ .

An implication can be drawn of the form  $A \Rightarrow B$ , i.e. if A, then B where B is a set of separate data items in T. Such an implication is called an association rule and it is bound by [8]–

- i.  $A \subseteq I$ ,
- ii.  $B \subseteq I$
- iii.  $A \cap B = \emptyset$

The association rule is used to determine the number of times B occurs given that A has occurred in a transaction, thus defining a relation between the two. In order to measure the interestingness of a rule, to select the most relevant, the following parameters are used if D is the set of all task relevant data [9] –

- i. Support – the rule  $A \Rightarrow B$  holds with support s if s% of transactions in D contains AUB.
- ii. Confidence – the rule  $A \Rightarrow B$  holds with confidence c if c% of the transactions in D that contain A also contain B.

The support and confidence for all rules are created initially. A threshold value is decided, and all rules having the support or confidence value below this measure are ignored.

### A. Apriori Algorithm

The Apriori algorithm is an influential algorithm for mining frequent item sets for Boolean Association Rules. The algorithm uses the values obtained for k elements to explore the (k+1) level of elements, hence employing a level-wise search.

The Apriori algorithm follows one important property – All non-empty subsets consisting of frequent item sets must be frequent. For example – if {AB} is a frequent subset, then {A} and {B} must also be frequent. [9]

The algorithm is fairly simple that relies on the following concepts [7]–

- i. The candidate and frequent item sets are defined as  $C_k$  and  $L_k$  respectively.
- ii. The initial (or previous) frequent item set  $L_{k-1}$  is computed.
- iii.  $C_k$  is calculated as the Cartesian product of  $L_{k-1}$  with itself.
- iv. Any item set having a value less than the required threshold value (or support) is removed and not considered for the following steps.
- v. This process is repeated iteratively till an association between the item sets is formed that satisfies the support and confidence requirements.

The generation of the candidate item set is expensive with regards to both time and space.

### B. FP-Growth Algorithm:

The FP-Growth algorithm uses a divide-and-conquer strategy to approach the database.

The FP –Growth algorithm follows two basic steps –

- i. Build a compact data structure called the FP-tree. This is built using two passes over the data set.
- ii. Extract frequent item sets directly from the FP-tree.

The tree structure has the following defining parameters [7]–

- i. Root node having item-prefix sub-trees as children.
- ii. Each node in the sub-tree consists of –
  - a. Item name
  - b. Count
  - c. Node link
- iii. Frequent item header table consisting of –
  - a. Item name
  - b. Head of the node link

The FP-tree construction follows a two-pass system [10] –

- i. Pass 1:
  - a. Scan the data set to find the support.
  - b. Reject the irrelevant data items.
  - c. Sort in decreasing order of support.
- ii. Pass 2:
  - a. Read each transaction once and map the path starting from a root node having null value.
  - b. If a common prefix (or item) occurs, increment the counter and assign pointers accordingly.

Figure 1 depicts the construction of an FP-tree.

The FP-growth algorithm extracts the frequent item sets from the above tree. This is done by simply following the path of each relevant leaf node to the root.

FP-tree construction proceeds in the following steps [7]–

**Input:** A transactional database DB and a minimum support threshold.

**Output:** Its frequent pattern tree, FP-tree.

- a. Scan the transaction database DB once. Collect the set of frequent items F and their supports. Sort F in support descending order as L, the list of frequent items.
- b. Create the root of an FP-tree, T, and label it as “root”. For each transaction Trans in DB do the following.
- c. Select and sort the frequent items in Trans according to the order of L. Let the sorted frequent item list in Trans be [p | P], where p is the first element and P is the remaining list. Insert item into tree.
- d. To insert an item, If T has a child N such that N.item-name = p.item-name, then increment N’s count by 1. Otherwise, create a new node N, and let its count be 1, its parent link be linked to T, and its node-link be linked to the nodes with the same item-name via the node-link structure. If P is nonempty, insert recursively.

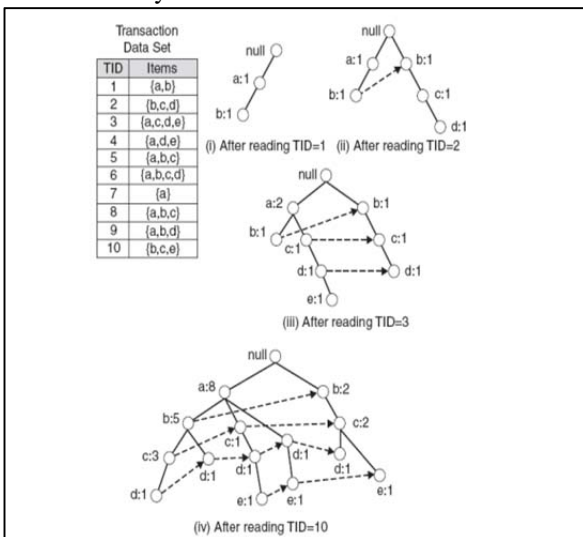


Fig.1. Construction of FP-Growth Tree [14]

The FP-Growth algorithm is a bottom-up approach and does not require the generation of candidate item sets.

On comparing the two algorithms, it was found that the FP-Growth algorithm was more space and time efficient.

## VII. CLUSTERING

In ITS, it is not possible to create a new model for every student. Grouping similar users based on some characteristics and studying the group as a whole is practical. Clustering is the method of grouping objects that belong to the same class based on some inherent similarities.

Clustering is a descriptive task that seeks to identify homogenous groups of objects based on the values of their attributes. [11] All the objects within a cluster can be treated as a single object. Cluster analysis involves first

partitioning the set of data into groups based on data similarity and then assigning them to label groups. The advantage of clustering over classification is that, it is adaptable to changes and it is easy to single out the key characteristic that distinguishes one group from the other. [12]

Clustering methods are varied and include various methods such as:

- i. Partitioning method
- ii. Hierarchical method
- iii. Density-based method
- iv. Grid based method
- v. Model based method
- vi. Constraint based method

Clustering algorithms can follow one or more of the above mentioned methods. Broadly, clustering methods can be classified as follows:

- i. Small data base clustering algorithms:
  - a. K-means algorithm
  - b. Fuzzy C-means clustering algorithm
- ii. Large database clustering algorithms:
  - a. Hierarchical clustering

### A. K-Means Algorithm:

K-means algorithm is a partition based algorithm. It attempts to decompose the data set into a number of partitions and then assign the data object to a particular partition based on how similar the data items are in a particular cluster. An attempt is made iteratively to reduce the measure of dissimilarity of the objects within a cluster while at the same time, increasing the difference between the objects belonging to two separate clusters.

K-means algorithm is the simplest unsupervised learning algorithms. The procedure follows the creation of k clusters from a data set of n elements. Each cluster is associated with one centroid – in total, there exist k centroids. The centroids are now placed in several locations throughout the dataset. A change in the position of the centroid gives varying results. Ideally, the centroids are placed as far away from each other as possible. After the initial creation of centroids is successful, each element in the data set is associated with one centroid based on a certain criteria that needs to be satisfied. All elements or objects within the data set are assigned a centroid. Once no elements are remaining, the early/initial grouping phase has been accomplished. For the following steps, k new centroids are calculated for the clusters resulting from the previous step. Once these new centroids are generated, the elements are reallocated to the new centroids. This loop of generation of centroids and assignment of elements is created. The positions of the centroids change per iteration. The loop is terminated once there is no change in the position of the centroid.

The algorithm aims at minimizing the objective function, in this case the squared error function which is:

$$J = \sum_{j=1}^k \sum_{i=1}^n |x_i^{(j)} - c_j|^2$$

Where,

- i.  $|x_i^{(j)} - c_j|^2$  Is the chosen distance measure between a data point  $x_i^{(j)}$  and  $c_j$ .
- ii. The cluster center  $c_j$  is an indicator of the distance of the  $n$  data points from their respective cluster centers.

The algorithm is composed of the following steps [12]:

- i. Place  $K$  points into the space represented by the objects that are being clustered. These points represent the initial clusters.
- ii. Assign each object to the group that has the closest centroid.
- iii. When all objects have been assigned, recalculate the positions of  $K$  centroids.
- iv. Repeat steps ii and iii until the centroids no longer change positions. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. The k-means algorithm can be run multiple times to reduce this effect.

K-means is a simple algorithm that has been adapted to many problem domains. In the educational domain, k-means algorithm has been proved to be popular for determining the various clusters of students by assigning labels to each student group. For example, if the dataset containing student performance per semester, per subject is subjected to the k-means algorithm it is possible to obtain a dataset partitioned based upon performance labels such as "Poor", "Excellent", "Average", etc. In an ITS System, it proves beneficial to classify students based on these labels so as to allow the system to adapt to teaching methodologies based upon the students' aptitude. Additionally, academic material recommendations can be made based upon the cluster to which a student has been assigned.

#### B. Fuzzy C-means Algorithm:

Fuzzy C-means clustering is an improvement on the K-means clustering algorithm. It allows a particular data element to be assigned to two or more clusters. Given a set of data, the problem of clustering lies in finding cluster centers that can characterize relevant classes of  $X$  with accuracy. The fuzzy C-means algorithm replaces the stringent requirement placed by K-means with a weaker, logically relevant requirement. This algorithm makes use of the basic principles of fuzzy logic.

The Fuzzy C-means method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m |x_i - c_j|^2$$

Where,

- i.  $m$  is any real number greater than 1
- ii.  $u_{ij}$  is the degree of membership of  $x_i$  in the cluster  $j$
- iii.  $x_i$  is the  $i$ th of  $d$ -dimensional measured data
- iv.  $c_j$  is the  $d$ -dimension center of the cluster
- v.  $|\ast|$  is any norm expressing the similarity between any measured data and the center.
- vi.  $1 \leq m \leq \infty$

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above. The algorithm stops when the following becomes true –

$$\max_{ij} \{ |u_{ij}^{(k+1)} - u_{ij}^{(k)}| \} < \epsilon$$

Where,

- i.  $\epsilon$  is a termination criterion between 0 and 1
- ii.  $K$  is the iteration steps.

This procedure converges to a local minimum  $J_m$ . The algorithm is similar to the K-means algorithm.

Fuzzy C-means algorithm is a much more efficient and not stringent compared to the K-means algorithm with respect to the constraints employed on the data.

#### C. Hierarchical Clustering:

Hierarchical clustering aims to build a hierarchy of clusters such that the relationship between the individual members is depicted and the clusters are merged based on the similarities in the characteristics of the clusters.

There are two types of hierarchical clustering algorithms [12]:

- i. Agglomerative – this is a "bottom up" approach which begins by examining a single cluster. The similar clusters found based on evaluation of certain characteristics are then merged to give larger clusters. This process finally stops when the difference between the members of adjoining clusters is irreconcilable to one cluster.
- ii. Divisive – this is a "top-down" approach which starts with the analysis of individual clusters. Each cluster is examined to determine if it can be broken down further into smaller clusters with a finite number of members. The algorithm continues the derivative procedure till the clusters obtained are unit sized, i.e., they cannot be broken down further to form meaningful clusters of data.

Given a set of  $N$  items to be clustered, and an  $(N \times N)$  similarity matrix, the basic process of hierarchical clustering using agglomerative method is –

- i. Start by assigning each item to its own cluster. If  $N$  items exist, there are now  $N$  clusters, each containing just one item. Let the distances (similarities) between the clusters equal the

distances (similarities) between the items they contain.

- ii. Find the closest (most similar) pair of clusters and merge them into a single cluster, so that there are now N-1 clusters.
- iii. Compute distances (similarities) between the new cluster and each of the old clusters.
- iv. Repeat steps ii and iii until all items are clustered into a single cluster of size N.

The distances between the clusters can be computed using different methods.

#### VIII. ADVANTAGES AND DISADVANTAGES

The advantages and disadvantages of the various algorithms discussed in the previous sections are detailed in Table 1.

	ALGORITHM	ADVANTAGES	DISADVANTAGES
CLASSIFICATION AND PREDICTION	Naïve Bayes	It requires a small amount of training data to estimate the parameters necessary for classification.	It assumes that all the characteristics considered are independent of one another.
		Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.	Drawing its roots from decision theory, many a times it is not possible to calculate with certainty the class to which a particular object must be assigned.
	Decision Tree	Understandable prediction rules are created from the training dataset.	In case of small samples, the data may be over-fitted or over-classified.
		Finding the leaf nodes enables the test data to be pruned, thus reducing the number of tests.	Only one attribute at a time is tested for decision making.
		The tree is built quickly.	
ASSOCIATION RULE MINING	Apriori	It uses large item set property.	It assumes the transaction database to be memory resident.
		The algorithm is easily parallelized.	Multiple scanning of the database.
		It is easy to implement.	
	FP-Growth	It is faster in terms of computation.	It is expensive in terms of memory, i.e., it requires a large memory space.
		It "compresses" the data set.	
		It requires only two passes over the dataset thus eliminating the need for a candidate item set generation	
CLUSTERING	K-MEANS ALGORITHM	It is fast, robust and easy to understand.	The algorithm fails for a non-linear dataset.
		The results are optimal when the data set is distinct or well separated.	It is unable to give optimal results for categorical data, i.e., when the mean is not defined.
		Works best for large number of variables, when k is small.	It does not handle noisy data and outliers well.
		Produces constricted clusters	Fixed number of clusters can make it difficult to forecast what K should be.
			Difficulty in comparing the quality of clusters formed
	FUZZY C-MEANS	It is an unsupervised learning technique.	It requires a long computational time.
		It always converges for the given values.	It is sensitive to noise.
	HIERARCHIAL	No prior information regarding the clusters is required.	It has a time complexity of $O(n^2 \log n)$ .
		The algorithm is efficient and easy to implement.	It is sensitive to noise and outliers.
		Provides good result visualizations	The algorithm fails to correct any previous error, if any.

Table 1 Advantages and disadvantages [13] [4] [12]

#### IX. CONCLUSION AND FUTURE WORK

Educational Data Mining is a new paradigm that could possibly change the pace of educational progress in our country. The study of this field will prove not only lucrative in the long run but revolutionary in terms of student experience, which will stand to redefine the student-teacher relationship and boost student productivity. Educational Data Mining has a long way to go in terms of providing functioning systems at a reasonable cost and providing quantifiable results in terms of positive impact on the student education. The research conducted by scholars all over the world so far has been paramount in developing the initial systems for the same.

It is essential to bring out utmost productivity in these upcoming Adaptive and Intelligent Tutoring Systems. Maximum productivity can be obtained through the utilization of algorithms in a time and space efficient fashion. For this purpose, the algorithms in each of the following techniques – Classification, Clustering and Association Rule Mining have been studied along with their expected impact on an educational dataset. The algorithms studied were basic with major scope for improvements with regards to computational cost and time complexity. Improving upon these algorithms is expected to be the focus of any future study. Though ITS Systems have drawbacks in terms of being strictly instructive and expensive to implement, it is expected that with the increase in innovation, in the near future full implementation of these systems will become a possibility.

#### ACKNOWLEDGMENT

This research paper is made possible through the help and support from my teachers and family. I would like to sincerely thank my parents for their guidance and support – both emotional and financial. I would also like to extend my gratitude to my teachers who have pushed me to challenge myself and encouraged me to excel. The product of this research paper would not be possible without all of them.

#### REFERENCES

- [1] Romero, Cristóbal, and Sebastián Ventura. "Educational data mining: a review of the state of the art." *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, IEEE Transactions on 40.6 (2010): 601-618.
- [2] Tair, M. M. T., and Alaa M. El-Halees. "Mining educational data to improve students' performance: A case study." *International Journal of Information and Communication Technology Research* 2.2 (2012).
- [3] Ghadirli, Hossein Movafegh, and Maryam Rastgarpour. "A Web-based Adaptive and Intelligent Tutor by Expert Systems." *Advances in Computing and Information Technology*. Springer Berlin Heidelberg, 2013. 87-95.
- [4] B.Nithyasri, K. Nandhini, Dr. E.Chandra. "Classification Techniques in Education Domain". (IJCS) International Journal on Computer Science and Engineering (2010) Vol. 02, No. 05, 1679-1684.
- [5] Jantan, H., Hamdan, A. R., & Othman, Z. A. (2010). "Classification and prediction of academic talent using data mining techniques". *Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 491-500). Springer Berlin Heidelberg.
- [6] Devikala, D., Kamalra, N. "Data Mining Approaches on Detection of Academic Failure and Dropout: A brief Survey". (IJCTT) International Journal of Computer Trends and Technology (2014). Vol. 14, No.3.
- [7] Vanitha, K. "Evaluating the Performance of Association Rule Mining Algorithms." *Journal of Global Research in Computer Science* 2.6 (2011): 101-103.
- [8] Kumar, Varun, and Anupama Chadha. "Mining association rules in student's assessment data." *International Journal of Computer Science Issues* 9.5 (2012): 211-216.
- [9] Angeline, D. Magdalene Delighta. "Association Rule Generation for Student Performance Analysis using Apriori Algorithm." *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)* 1.1 (2013): p12-16.
- [10] Tan, Pang-Ning, and Michael Steinbach. "Vipin Kumar, Introduction to Data Mining." (2006).
- [11] Varghese, Bindhya M., et al. "Clustering Student Data to Characterize Performance Patterns." *IJACSA International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence* (2011).
- [12] Sharmila, Mishra R.C., "Performance Evaluation of Clustering Algorithms", (IJETT) International Journal of Engineering Trends and Technology. (2013) Vol. 4, Issue 7.
- [13] Khurana K, Sharma S. "A Comparative Analysis of Association Rules Mining Algorithms." *International Journal of Scientific and Research Publications* (2013). Volume 3, Issue 5.
- [14] <http://www.saedsayad.com/>