Modeling and Simulation of K-Means Clustering Learning Object Adaptability Model for Selecting Materials in E-Learning

Awoyelu I.O.
Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

Awosan O.A.
Department of Computer ScienceOsun State College of Education, Ilesa, Nigeria

Adagunodo E.R.
Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

ABSTRACT
The delivery of adaptive instructional materials to learners is a good way of achieving effectiveness in learning session. Fitting teaching material varies from one student to another based on students' knowledge level, that is, cognitive status and ability to learn. This study proposes a model for selecting appropriate learning materials to learners. The proposed model was formulated using the K–means Clustering Algorithm and represented using Unified Modelling Language (UML). The adaptive model was simulated using the K-Means Clustering Algorithm to train and test the adaptive model using the historical data collected from a developed e – Learning system for Undergraduate students taking Introduction to FORTRAN programming Language course. The evaluation results showed that the system had precision of 0.7143, 0.6667, 1.000 and 0.2942 for learning object 1, learning object 2, learning object 3 and learning object 4 respectively. Also, the recall results were 0.6250, 0.8889, 0.6667 and 0.3864 for learning object 1, learning object 2, learning object 3 and learning object 4 respectively. The system can be used to effectively and successfully assign learning materials to learners based on their cognitive level.

General Terms
Data Mining, Clustering

Keywords
Learning objects, Adaptive, Cognitive, K-means clustering, E-learning

1. INTRODUCTION
The trend in educational technology is towards providing instructional environment that will provide individualization in learning. This therefore leads to developing adaptive learning systems that provides learning materials to learners taking into consideration their learning requirements and understanding capability. Instructional approaches and techniques that are geared to meet these needs are called adaptive instructions [1].

According to [2], an adaptive instruction is generally characterized as an educational approach that incorporates alternative procedures and strategies for instruction and resource utilization. It is built in such a way that it permits students to navigate various routes and spent good time in learning.

Particularly, an adaptive instruction refers to a way of accommodating individual differences in learners with the aim of helping each one of them develop their individual skills required in the teaching and learning process..

Reference [3] describes three essential ingredients of adaptive instruction. The first ingredient is that adaptive instruction provides various alternatives for learning and there are many goals from which the leaner can choose. Secondly, it analyses the capabilities of an individual and adjust to the learner's particular cognitive level. Lastly, it strengthens an individual's ability and develop the skills necessary for success. The aim of adaptive e-learning is to provide the learner appropriate content according to his or her cognitive level at the right time.

In a learning system, students' knowledge level (cognitive status) varies. Their ability to learn may also differ. Fitting teaching materials may vary from one student to another based on these factors. Consequently, there is the need for the selection of suitable learning objects appropriate to a learner's knowledge level. Hence, this study focusses on developing an adaptability model for e-Learning system that is suitable for the selection of appropriate learning materials, based on the cognitive status of a student.

The rest of this paper is organised as follows. Section 2.0 addresses related works on selection of learning materials. Section 3.0 focuses on data collection and methodology used while Section 4.0 focusses on the evaluation of the proposed system.

2. RELATED WORKS
There are many works on personalization of learning objects for learners. Reference [4] worked on an e-learning model for the personalization of courses based on both a student's needs and capabilities and on the teacher's profile. Personalized learning paths in the courses were modelled using the graph theory. [5] also proposed an automatic tool, based on the students' learning performance and communication preferences, for the generation and discovery of simple student models with the ultimate goal of creating a personalized education environment. The approach was based on the Projective adaptive resonance theory (PART) algorithm, which produces rules from pruned partial decision trees. Reference [6] described a tool that could help trace deficiencies in students' understanding. It resorted to a tree abstract data type (ADT), built from the concepts covered in a laboratory, lecture, or course. Once the tree Abstract data type (ADT) is created, each node can be associated with different entities such as student performance, class performance, or laboratory development. Using this tool, a teacher could help students by discovering concepts that needed additional

coverage while students might discover concepts for which they would need to spend additional working time.

[7] proposed a personalized web-based learning system by applying web usage mining techniques to personalized recommendation services. The approach is based on a web page classification method, which uses attribute-oriented induction according to related domain knowledge shown by a concept hierarchy tree. A clustering method was introduced by [8], where a network-based testing and diagnostic system was implemented. It entails a multiple-criteria test sheet-generating problem and a dynamic programming approach to generate test sheets. The proposed approach employs fuzzy logic theory to determine the difficulty levels of test items according to the learning status and personal features of each student, and then applies Fuzzy Adaptive Resonance Theory (Fuzzy ART) model to cluster the test items into groups, as well as dynamic programming for test sheet construction.

[9] presented the usability of Kohonen's Self-Organizing Maps (SOM) for the evaluation of students in a tutorial supervisor (TS) system, as well as the ability of fuzzy TS to adapt question difficulty in the evaluation process. An investigation on how data Mining techniques could be successfully incorporated to e-learning environments, and how this could improve the learning processes was presented by [10]. Data clustering is suggested as a means to promote group-based collaborative learning and to provide incremental student diagnosis. [11] presented user actions associated to students' Web usage as part of a Data Mining process. The Expectation-Maximization (EM) algorithm was used to group the users into clusters according to their behaviours. These results were used by teachers to provide specialized advice to students belonging to each cluster. The simplifying assumption that students belonging to each cluster should share web usage behaviour makes personalization strategies more scalable. Clustering was proposed by [12] to group similar learning documents based on their topics and similarities. A Document Index Graph (DIG) for document representation was introduced, and some classical clustering algorithms (Hierarchical Agglomerative Clustering, Single Pass Clustering and k-NN) were implemented. Different variants of the Generative Topographic Mapping (GTM) model, a probabilistic alternative to SOM, were also presented by [13] for the clustering and visualization of multivariate data concerning the behaviour of the students of a virtual course.

[14] explored the determination of personality characteristics, formal description and mining algorithms of learners to propose a construction method of learner model in adaptive learning system. The work does not specifically analyze the building of learner knowledge model. [15] presented an ontology-based approach to develop an adaptive e-learning system based on the design of semantic content, learner and domain models to tailor the teaching process for individual learner's needs. The student has a control over the learning path and can start anywhere in the system.

3. METHODOLODY
In order to develop the proposed learning material adaptability model, lecturers of Osun State College of Education, Ilesa were interacted with, on the understanding of the curriculum for a course offered in the Institution - Introduction to FORTRAN Programming Language, which is the selected course for this study. Following this interaction, a number of variables were identified for the selection of learning objects.

These are student model, formative tests, performance metrics, decision model and the respective learning objects.

3.1 Variables for the Selection of Learning Objects
The various variables for the selection of learning objects are considered in the following sections.

3.1.1 Student model
This is the demographic information about the students taking the course. It can also be seen as student profile. The profile is created with the need of uniquely identifying the personal trait of each student using certain characteristics that define the students. The student profile used is as shown in Table 1. For the purpose of this study, student's age, ethnicity, gender and department were used. These are explained as follows.

i. Student age: The students' ages for this study were classified according to age groups defined within intervals - 15–20 years, 21–25 years and 26–30 years based on the distribution of students taking the course. This was necessary in order to eliminate problems with too many labels assigned to a variable.

ii. Ethnicity: The students were also classified according to their respective ethnic groups in Nigeria to which they belong. The ethnicity was classified as: Yoruba, Hausa, Ibo and others.

iii. Sex/Gender: The students were classified according to their gender - either male or female, as it was observed from the students taking the course; and

iv. Department: The students were classified according to their respective departments as it was observed that students from other departments took the course as required by their respective departments.

Table 1. Student Profile for Learning Material Adaptive Model

S/N	Variable	Labels
1.	Name	Character String
2.	Age	15 -20
3.	Ethnicity	Yoruba
4.	Sex/Gender	Female
5.	Department	Computer Science Education

3.1.2 Formative tests
For the purpose of this study, multiple choice questions were used to develop the formative tests taken by the students. Formative tests are used to evaluate the performance of the students and the students' weakness so that the necessary provisions are provided. The formative tests are made up of eight (8) objective type questions labelled a – d (4 options) out of which only one of the options is correct. This is as shown in Table 2. Each formative test represents a topic in "Introduction to FORTRAN programming" with a total score of 100%. It is assumed that there are five (5) topics in the proposed course. There are five (5) formative tests with forty (40) questions to be answered by the student. Each formative test contains three (3) compulsory (pre-requisite questions) which are tagged to a concept or sub-topic under the formative test in view. Each concept is tagged to a learning material which contains information on the respective concept. A student can then read the learning object provided before he moves to the next formative tests.

Table 2. Formative Tests' Contents for Learning Material Adaptive Model

Formative Test	Topic/Learning Objective	Questions (Score)
1	Number Systems and Computer Organization	8 (100%)
2	Flowcharts, Algorithms and Programming Design	8 (100%)
3	FORTRAN Program Structure	8 (100%)
4	FORTRAN Data Structures	8 (100%)
5	FORTRAN Control and Conditional Statements	8 (100%)
	TOTAL	40 Questions

3.1.3 Performance metrics

Performance metrics form a part of the student profile needed for developing the student model for the learning material adaptability model proposed for this study and this is a classification of the scores of the students for every formative test taken. This is as shown in Table 3. This implies that a new performance metric is defined for every test taken by the student and this is used to classify the student's performance or cognitive ability for each formative test. Each formative tests is composed of eight (8) questions, and allocated a total score of 100%. This implies that each question carries a score of 12.5%. The pass mark for each formative test is 70%.

Table 3. Performance Metric for Student Profile

S/N	Performance Metric	Score Interval (%)
1	Excellent	70 – 100
2	Good	60 – 69
3	Fair	50 – 59
4	Poor	40 – 49

3.1.4 Decision Model

The decision model is the part of the system that is solely responsible for the selection of learning objects (material) to a student based on his cognitive ability. The model, basically, maps the student model alongside the performance metric and the questions failed in the formative test, to determine which learning material (object) is to be recommended to the student. The e-learning system considered provides a formative test to a student after which the score is determined, with the questions failed stored in the system's database. The decision model checks the performance metric of the student to determine if the student scores up to the necessary threshold provided. For the proposed e-learning system developed, the threshold is 70%. If a student scores above the threshold and passes all the pre- requisites, the student is allowed to progress to the next formative test; else, the student is requested to retake the formative test failed. If the student satisfies the requirement, which in this case, is passing the threshold, the questions that were failed in the formative tests are considered next. If a student fails a compulsory (pre-requisite) question, the student is provided the learning material, which is mapped to the concept covering the compulsory question failed. This information was used to personalize the system for the student such that, the required learning object is proposed to each student using the e-

learning system while exploring their respective areas of weakness.

3.1.5 Learning objects

The learning objects are the learning materials which are recommended to a student using the adaptive e-learning system. The learning objects are expected to enhance learning by addressing the areas of weakness observed in the student's performance. Each learning material is therefore a portable document format (.pdf file) which is uploaded at the end of every formative tests where a compulsory question is failed. For each formative test, three (3) out of the total eight (8) questions are mapped compulsory and assigned a learning object respectively labelled 1, 2 and 3. If it is observed that a student failed the tests and is told to retake the tests, a learning object covering all aspects of the topic is selected and is assigned a label 4. Therefore, there are four learning objects for each question; each representing a compulsory concept required to be understood by the student before proceeding to the next formative test as shown in Table 4.

Table 4. Learning Objects for Adaptive Learning Model

S/N	Formative Test	Concept	Learning Object
1	Number System and Computer Organization	Base 2 Number System	1
		Hexadecimal Number System	2
		Computer Organization	3
2	Flowcharts, Algorithms and Program Design	Flowcharts	1
		Algorithms	2
		Top-Down Programming Design	3
3	FORTRAN Programme Structure	Program Structure	1
		Declarative Statements	2
		Programming Errors	3
4	FORTRAN Data Structures	Variable data types	1
		BEDMAS	2
		Logical data types	3
5	FORTRAN Control and Conditional Statements	Conditional Statements	1
		Arrays	2
		Subroutines	3

3.2 Proposed Learning Material Adaptability Model

The proposed model is as shown in Figure 1. The formative tests were first presented to the student in order to determine how much the student knows about the topics covered by the formative tests. The student by this time had provided his/her personal information which was used to develop a student model, which in turn, is used to develop a profile for the student. This part of the adaptive model is called the information gathering layer of the adaptive model. After the information gathering layer, the formative test is presented to the student. This stage involves the identification of the areas of weakness which the student faces in the presented topic covered in the formative test. Whenever a student fails a course, learning object 4 is recommended, which provides information needed for the review of the whole topic covered

in the formative test. This was used as a form of feedback to the adaptive model in order to determine the cognitive level of the student before presenting the information to the decision layer. The performance scores of the formative tests were used to evaluate the knowledge level of the student. Thus, students who scored more than 70% are allowed to progress to the next formative test.

The decision layer is that part of the model that holds the decision model. The decision model is the part of the learning material adaptability model that is solely responsible for assigning learning objects to the students based on the students' performance in compulsory (pre-requisite) questions during formative tests.

The decision model for this study uses the k-means clustering algorithm to assign learning objects to the students based on the information provided on the test taken, questions failed,

status of the test (pass or fail), the concept related to the question etc. The clustering algorithm on the other hand was developed by using the historical data collected from the student's interaction with an e-learning system based on the use of formative tests provided to the student. Unsupervised machine learning was used to create the different clusters existing in the dataset such that each cluster represents a learning object for each formative test.

The next stage is the assessment layer which evaluates the way in which the progress of the formative tests is performed. This stage is responsible for the determination of progress or failure depending on the threshold set for the student for every formative tests taken by the student. The formative tests are divided into five (5) different frames consisting of Test 1 to Test 5, each comprising of eight (8) questions carrying equal marks. Each formative test is built on the rules of programmed instructions:
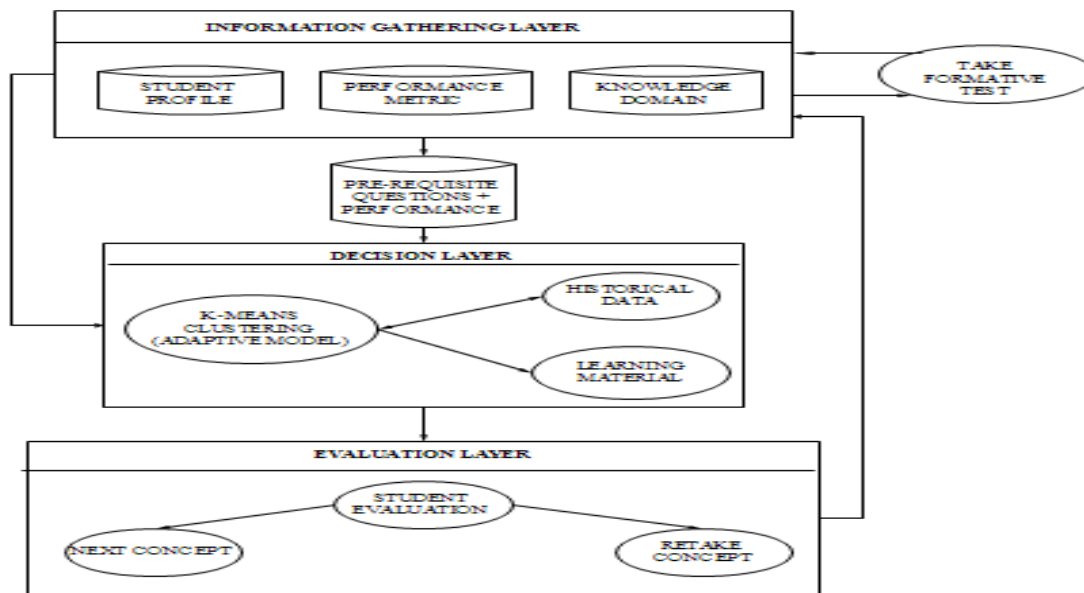
Figure 1 Proposed Learning Material Adaptability Model

i.   The content of each test is presented in small chunks;

ii. The contents are organized and presented in a

     simple-to-complex chunk (from Test 1 for foundational all the way to Test 5);

iii. The learner responds and receives feed-back;

iv. Learner can set his/her own pace; and

v.   The path of instruction is linear.

The learner responds by correctly, receiving feedback and moving forward. If the response is correct, the learner progresses, as long as there is no mistake, thus allowing the learner to set his own pace. The formative tests provided are linear with no path diversion from the directed instruction. The formative tests contain mostly knowledge level questions because knowledge level tests are used to confirm whether the student can recall a fact or not. Nevertheless, a student should not be allowed to move to the next formative test as long as he has not passed the required threshold of 70% required for each formative test.

3.3 Learning Material Adaptability Clustering Algorithm

Clustering algorithms are a type of unsupervised learning techniques. Clustering is a task of grouping a set of objects or attributes in such a way that objects in the same group (called a cluster) are more similar (in one sense or the other) to each other than to those in other groups (clusters). For the purpose of the development of the proposed learning object adaptability model, which selects learning objects based on the performance of students in their formative tests, the clusters were used in grouping the student model, performance metrics and the prior knowledge of the student into any of the appropriate learning material or clusters. This is as shown in Figure 2.
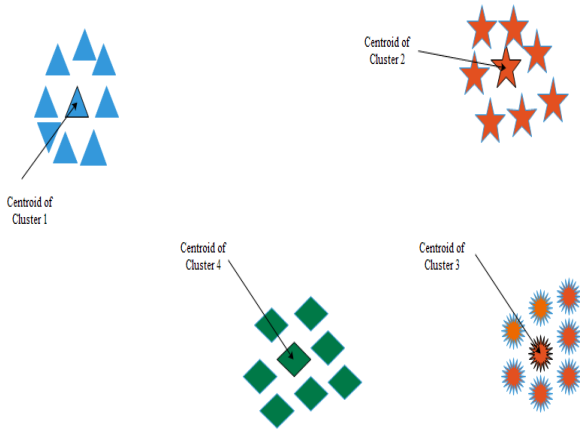
Figure 2 Set of Data Partitioned into Four Clusters

The figure shows how clusters are created for a given number of data sets. The datasets for the study comprise of the student profile, performance metrics and proper knowledge used in determining the path of learning by the e-learning system's decision model.

The proposed learning material adaptability model is required to classify each student into a particular learning object cluster based on its ability to portion each student into a particular cluster based on the chosen criteria of the cluster's centre. For the purpose of this study, the mean was chosen as the criteria of selecting the cluster's centre. The k-means clustering algorithm was the selected clustering algorithm used to partition the students into their respective learning material (object) cluster.

In the k-means clustering algorithm, the first parameter that needed to be specified is the value of K i.e. the number of clusters. After the value was determined, these K points were chosen as cluster centres. All instances (student and knowledge information) that need to be classified are assigned to their closest cluster centre according to the Euclidean distance metric. After this, the centroid or the mean of all instances in each cluster were calculated. These centres or mean values were taken to be the new centre values for their respective clusters. The process was repeated iteratively until the same points are assigned to cluster centres in consecutive rounds, at which stage, the cluster centres are stabilised and did not change after this point. Figure 3 shows the flowchart for the k-means clustering algorithm proposed for clustering the learning materials.

K–means clustering algorithm's objective is to minimize the average squared Euclidean distance of instances (student profiles, performance metrics and prior knowledge) from their cluster centres. This is defined as the mean or centroid, μ of the students in a cluster ω:

$$\bar{\mu}(\omega) = \frac{1}{|\omega|} \sum_{\bar{x} \in \omega} \bar{x} \qquad (1)$$

The definition of k-means clustering algorithm assumes that instances are represented as length normalized vectors (each record contains the student profile, performance metrics and prior knowledge) in a real valued space in the

familiar way. The ideal clustering k-means is a sphere with the centroid as its centre of gravity. Ideally, the clusters should not overlap. Following is the pseudo-code for k-means clustering algorithm:

K-MEANS$(\{\overline{x_1}, \ \overline{x_2}, \ldots \ldots, \overline{x_N}\}, K)$

1     $(\overline{s_1}, \ \overline{s_2}, \ldots \ldots, \overline{s_N})$ ←
      SelectRandomSeeds$(\{\overline{x_1}, \overline{x_2}, \ldots \ldots, \overline{x_N}\}, K)$

2   **for** k ← 1 **to** 4

3   **do** $\overline{\mu_k} \leftarrow \overline{s_k}$

4   **while** stopping criterion has not been met

5   **do for** k ← 1 **to** 4

6   **do** $\boldsymbol{\omega_k} \leftarrow \{\}$

7   **for** n ← 1 **to** N

8   **do** j ← arg min$_j$, $\left|\overline{\mu_j} - \overline{x_n}\right|$

9           $\omega_j \leftarrow \omega_j \cup \{\overline{x_n}\}$
    (reassignment of vectors)

10   **for** k ← 1 **to** K

11   **do** $\overline{\mu_k} \leftarrow \frac{1}{|\omega_k|} \sum_{\bar{x} \in \omega_k} \bar{x}$ (re-
    computation of centroids)

12   **return** $\{\overline{\mu_1}, \ \overline{\mu_2}, \ldots \ldots, \overline{\mu_N}\}$,

The measure of how well the centroids represent the members of their clusters is the residual sum of squares or RSS - the squared distance of each vector residual sum of squares from its centroids summed over all vectors, as shown in Equation 2 and Equation 3.
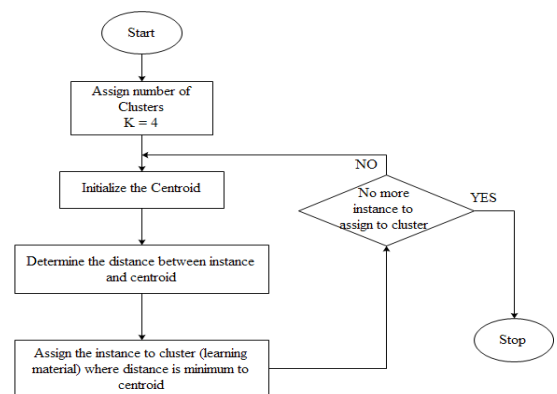

Figure 3 Flowchart for K-Means Clustering Algorithm

$$RSS_k = \sum_{\bar{x} \in \omega_k} |\bar{x} - \bar{\mu}(\omega_k)|^2 \qquad (2)$$

$$RSS = \sum_{k=1}^{4} RSS_k \qquad (3)$$

RSS is the objective function in the K-means and the aim of this study is to minimize this value. This is because N is fixed (number of instances for the study), RSS equivalent to minimizing the average squared distance, a measure of how well centroids represent the students.

3.4 Historical Data for Training Learning Material Adaptability Model

For the purpose of the development of the k-means clustering algorithm needed for developing the decision model for the learning material adaptability model proposed for the e-learning system, historical data containing the records of students who were selected for the course via a simple random sample selection method were used. This is as shown in Figure 4 which shows a sample of the training data for the proposed system.

Figure 4. Sample of the Training Data for the Proposed System

The instances contains information about the student's profile, performance metrics and the knowledge attained by the student alongside the respective learning object recommended by the decision-based model on the developed e-learning system. The goal of using the historical data is to develop an adaptive model suited for effective learning material (object) selection based on the information stored in the historical data.

The attribute relationship file format *(.arff)* is the recommended file format required by the WEKA software for the process of necessary data mining tasks to be performed. The file format requires that the name of the file be defined as the relation name, this is followed by the set of attributes used to define the instance for each student consisting of student profile, performance metric and knowledge acquired. The historical data were collected by recording every question activity performed by the student for every test performed by the student.

The attributes of the historical data used in this study are Student ID, Sex, age group, ethnicity, test number, attempt, test score, metric, question failed, pre-requisite, status and learning object. Learning object was used to define the output class or the appropriate learning material/object recommended to the student. If the student is discovered to fail a pre-requisite question, appropriate learning object is recommended based on the formative test written by the student. If the pre-requisite failed is 1, 2 or 3 and the student passes the formative test (student's score > 70%) then the respective learning material is obj_1, obj_2 or obj_3 respectively. If the student fails the test, the student is assigned learning object obj_4 irrespective of the fact that he passes or fails a pre-requisite question. Finally, if the student passes the course and does not fail a pre-requisite question then a learning object obj_0 is assigned to the student.

The historical data collected consist of 165 processed questions (pre-requisite or not) for 20 students who took all 5 formative tests for Introduction to FORTRAN programming in Osun State College of Education, Ilesha in Osun State, Nigeria. The historical data contain 165 instances with 11 inputs and 1 output variable.

3.5 Attribute Selection

Before using the historical data to perform the necessary training needed for the development of the k-means clustering algorithm needed for the development of the learning material adaptability model, it is important to evaluate the relevance of the attributes in the dataset collected.

The essence of reducing the number of attributes is to encourage the following:

- Target function/learning objects would be more compact;
- Speed-up in the running of actual learning algorithm; and
- Having more comprehensive results as a result of the compatibility.

By selecting appropriate attribute selection algorithms and search methods, the simulation environment is able to determine which of the attributes has more prevalence in determining the output variable of the dataset provided. The final result of selected attributes is thus a subset of the initial selected attributes provided. These attributes were then used to perform the training needed for the development of the proposed clustering algorithm.

The WEKA software chosen for simulation has a number of attribute evaluator algorithms and search method algorithms. For the purpose of this study, the chosen attribute evaluator was the Gain Ratio Attribute Evaluator which evaluates the worth of an attribute by measuring the gain ratio with respect to the output class (learning objects). The selected search method chosen by default for the Gain Ratio Attribute Evaluator is the Ranker which ranks all attributes by their individual evaluations.

The mathematical formula for the attribute evaluator is as defined in Equation 4.

$$GainR(Class, Attribute) = \frac{(H(Class) - H(Class|Attribute))}{H(Attribute)} \qquad (4)$$

where

$$H(Y) = -\sum_{y \epsilon Y} p(y) \log_2(P(y)),$$ and

$$H(Y|X) = -\sum_{x \epsilon X} p(x) \sum_{y \epsilon Y} p(y|x) \log_2(P(y|x))$$

After using the Gain Ratio Attribute Evaluator and the ranker search method to select attributes, the results in Table 5 shows the rank of the attribute selected for the study.

Table 5 The Rank of Attribute Selected for Clustering Algorithm's Training Dataset

Rank	Ranked Value	Attribute
1	0.946	Status
2	0.8253	Metric
3	0.7573	Pre-requisite
4	0.4099	Test Score
5	0.3169	Question Failed
6	0.1567	Attempt
7	0.0524	Test Number
8	0.0207	Ethnicity
9	0.0182	Age group
10	0.0127	Sex
11	0	Student ID

3.5.1 Training Data

Following the attribute selection, the numbers of instances were reduced from 165 to 66 by considering only the pre-requisite questions that were failed by the students. These were used to train the k-means clustering learning material adaptability model. The number of attributes of each instance of data was also reduced to fit the attribute selection rank provided by the gain ratio attribute evaluator. This is as shown in Figure 5. The attributes that were considered were then used to develop the *.arff* file format for the training data used for developing the k-means clustering model for the learning material adaptability model proposed for the study. The selected attributes used for the training of the k-means learning material adaptability clustering algorithm are status, metric, pre-requisite, test score, question failed, attempt and test number.

Figure 5 Section of Pre–processed Training Data Used for Proposed Model

4. DISCUSSION OF RESULTS

From the results of the simulation, it was discovered that the k-means clustering model made 22 incorrect classifications out of 66 instances provided which showed an accuracy of 66.67%. Figure 6 shows a graphical plot of the clusters that were allocated by the k-means clustering algorithm for learning material adaptability model. The correct and incorrect classification of the training and testing data can be clearly observed from Figure 4.3. Recall that cluster 0, cluster 1, cluster 2 and cluster 3 represent object 2 (dark blue), object 3 (red), object 4 (green) and object 1 (light blue) respectively.

Correct classifications are represented by star symbols while incorrect classifications are represented using square symbols. From Figure 6, which shows the results of the clusters allocated to the instances in the training data, out of the total 66 instances, there were 22 misclassified clusters. Out of 7 instances assigned to obj_1, the k-means cluster misclassified 2 to cluster 1 (obj_3) and the remaining 5 were correctly classified to cluster 3 (obj_1). Out of 24 instances assigned to obj_2, the k-means cluster correctly classified 16 to cluster 0 (obj_2) and the remaining 8 were misclassified to cluster 2 (obj_4). Out of 18 instances assigned to obj_3 all were correctly assigned to cluster 1 (obj_3). Out of 17 instances assigned to obj_4; 2 were misclassified to cluster 0 (obj_2), 7 misclassified to cluster 1 (obj_3), 3 were misclassified to cluster 3 (obj_1) and 5 were correctly classified as cluster 2 (obj_4). The confusion matrix was used to evaluate the performance of the learning material adaptability model. This is as shown in Figure 7.
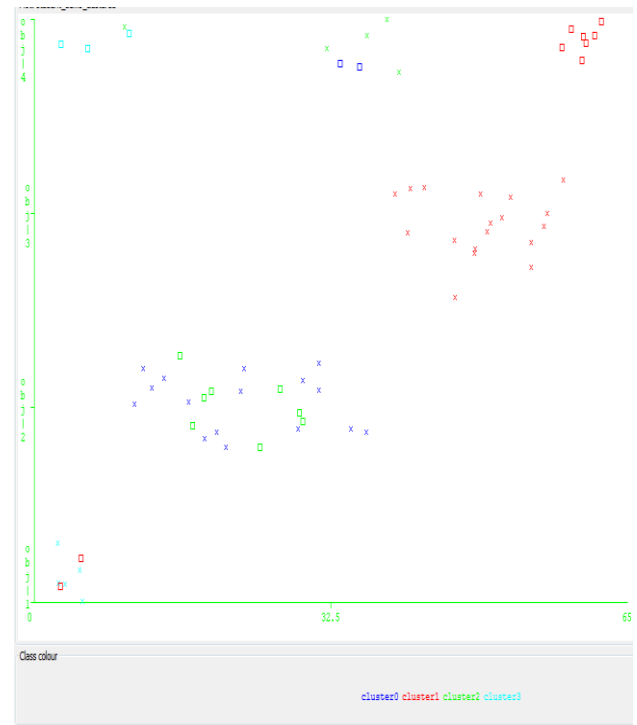

Figure 6 Graphical Plot of the Clusters Allocated to the Instances

4.1 Performance Evaluation of K-means Clustering Learning Object Adaptability Model

Precision and Recall were used to evaluate the performance of the k-means clustering learning material adaptability model developed for the correct selection of learning material for e-

learning systems. Precision is defined as the proportion of correct classifications out of the total classification made by the clustering algorithm. Recall is defined as the proportion of actual classes that were correctly classified by clusters.

$$Recall = \frac{TP}{TP + FN} \qquad (5)$$

$$Precision = \frac{TP}{TP + FP} \qquad (6)$$

Recall results for the learning objects, as shown in Figure 8, are 0.6250, 0.8889, 0.6667 and 0.3846 for learning object 1 learning object 2, learning object 3 and learning object 4 respectively. Precision results are 0.7143, 0.6667, 1.0000 and 0.2941 learning object 1 learning object 2, learning object 3 and learning object 4 respectively. The results show that the proposed learning material adaptability model has the likelihood of recommending 88% of the time learning object 2 to students followed by object 3 and object 1. The low true positive rate observed for learning object is believed to be due to the fact that students who fail the tests and are recommended learning object 4 are most likely to be misclassified as either learning object 1, 2 or 3 depending on the pre-requisite questions failed. The precision results of the learning objects show that the learning material adaptability model has more likelihood of correctly recommending learning object 3 followed by learning object 1 and learning object 2.

Figure 8. Performance Evaluation of the Adaptability Model

This work provided a means of providing a learning material adaptability model using K-Means clustering algorithm. The performance evaluation showed that the model has an accuracy of 66.67%.

Figure 7 Confusion Matrix of the K-means Cluster

Legend:
obj_2 is Cluster 0
obj_3 is Cluster 1
obj_4 is Cluster 2
obj_1 is Cluster 3

5. CONCLUSION

It is a fact that students may come from different backgrounds and their needs of study and goals may also be different; above all their abilities to learn may not be similar. Therefore, teaching style and fitting teaching materials may differ from a student to another. A learning material adaptability model (decision model) for e-learning that assesses the requirements, goal and capability of a student and dynamically sets a path for study was proposed. The instruction materials were dynamically selected as per the student's level of understanding from a given set of instruction materials. The system monitors the student and changes the path of study automatically as per the performance of the student.
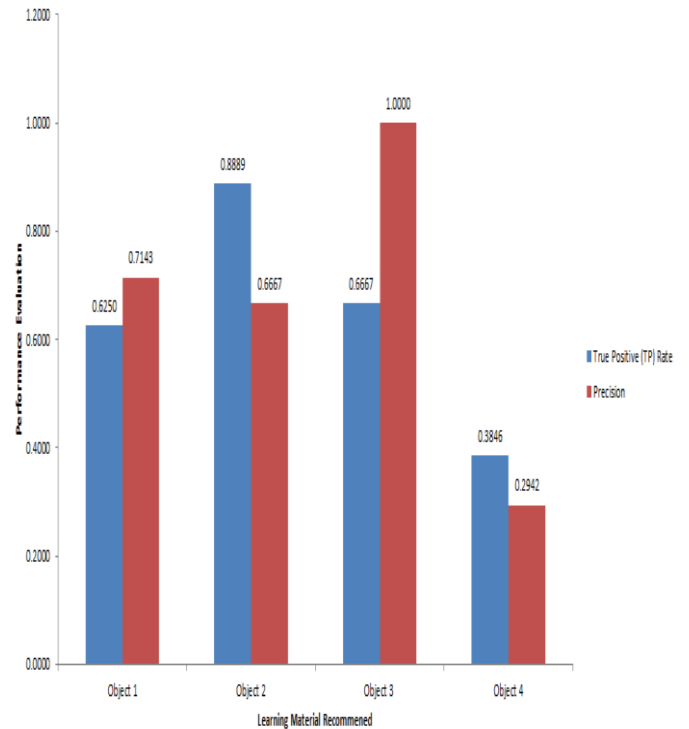
6. REFERENCES

[1] Como, L., and Snow, E. R. (1986). Adapting Teaching to Individual Differences among Learners. *In* M. C. Wittrock (Ed.), Handbook of Research on Teaching (3rd ed.). New York: Macmillan.

[2] Wang, M. and Lindvall, C. M. (1984). Individual Differences and School Learning Environments. *Review of Research in Education,* 11, 161–225.

[3] Glaser, R. (1977). Adaptive education: Individual, Diversity and Learning. New York: Holt.

[4] Carchiolo, V., Longheu, A., Malgeri, M., and Mangioni, G. (2003). Courses Personalization in an e- Learning Environment. *In Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies, ICALT'03*. July 9-11, 252-253.

[5] Licchelli, O., Basile, T.M., Di Mauro, N. and Esposito, F. (2004). Machine Learning Approaches for Inducing Student Models. *In: 17th International Conference on Innovations in Applied Artificial Intelligence, IEA/AIE 2004. LNAI* Vol. 3029. Springer-Verlag, Berlin Heidelberg New York, pp. 935-944.

[6] Yoo, J., Yoo, S., Lance, C. and Hankins, J. (2006). Student Progress Monitoring Tool Using Treeview. *In the 37th Technical Symposium on Computer Science Education, SIGCSE'06.* ACM Press. March 1-5, Houston, USA, pp. 373-377.

[7] Grieser, G., Klaus, P.J. and Lange, S. (2002). Consistency Queries in Information Extraction. In the *Proceedings of the 13th International Conference on*

Algorithmic Learning Theory. Lecture Notes in Artificial Intelligence, Vol. 2533. Springer-Verlag, Berlin Heidelberg New York, 173-187.

[8] Hwang, G.J. (2003). A Test-Sheet-Generating Algorithm for Multiple Assessment Requirements. *IEEE Transactions on Education,* 46(3), 329-337.

[9] Mullier, D., Moore, D. and Hobbs, D. (2001). A Neural-Network System for Automatically Assessing Students. In: Kommers, P., Richards, G. (eds.): *World Conference on Educational Multimedia, Hypermedia and Telecommunications,* 1366-1371.

[10] Tang, T.Y. and McCalla, G. (2005). Smart Recommendation for an Evolving e-Learning System: Architecture and Experiment. *International Journal on e-Learning,* 4(1), 105-129.

[11] Teng, C., Lin, C., Cheng, S and Heh, J (2004). Analyzing User Behavior Distribution on e-Learning Platform with Techniques of Clustering. *In the Society for Information Technology and Teacher Education International Conference,* pp. 3052-3058.

[12] Hammouda, K. and Kamel, M. (2005). Data Mining in e-Learning. In: Pierre, S. (ed.): e-Learning Networked Environments and Architectures: A Knowledge Processing Perspective. Springer-Verlag, Berlin Heidelberg New York.

[13] Castro, F., Vellido, A., Nebot, A. and Minguillón, J. (2005). Detecting a Typical Student Behaviour on an e-Learning System. *In: VI Congreso Nacional de Informática Educativa, Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación, SINTICE'2005.* September 14-16, Granada, Spain, 153-160.

[14] Qing L., Shaochun Z., Peng W., Xiaozhuo G. and Xiaolin Q. (2010). Learner Model in Adaptive Learning System. *Journal of Information and Computational Science*, 7(5), pp 1137-1145.

[15] Maryam Y., Hossein J. and Abdel R. (2013). A Personalized Adaptive e-learning Approach Based on Semantic Web Technology. *Webology*, 10(2). Article 110, Available at: http://www.webology.org/2013/v10n2/a110.pdf