

# Adaptive MOOCs Based on Intended Learning Outcomes Using Naïve Bayesian Technique

<https://doi.org/10.3991/ijet.v15i04.11420>

Ahmed Ewais (✉)

Arab American University, Jenin, Palestine  
Vrije Universiteit Brussel, Brussel, Belgium  
aewais@aaup.edu

Duaa Abu Samra

Arab American University, Jenin, Palestine

**Abstract**—Widespread adoption of MOOCs got researchers interest to support learners in their learning process. However, most of provided courses are teacher-centered approach rather than learner-centered approach. One of the possible solutions to enhance the learning process is to enable learner to learn a course that achieve a number of Intended Learning Outcomes (ILOs). Therefore, the main goal of this research work is to propose an approach for adapting MOOCs learning materials based on ILOs using classification algorithm namely Naïve Bayesian algorithm. Furthermore, the proposed approach considered the pedagogical aspects by generating a learning path based on the pedagogical relationship between learning concepts which are mapped to learning materials. As a result, the learner will be able to follow a course generated automatically based on selected ILOs and pedagogical relationships. To validate the proposed approach, a prototype has been developed and the effectiveness of the adopted technique has been validated using a precision-recall indicator. The results were promising as the precision-recall indicators provided interesting results in the classification process.

**Keywords**—MOOCs, Bayesian Network Classifier, Adaptive MOOCs, Intended Learning Outcomes.

## 1 Introduction

A number of the most famous universities like Stanford, MIT, and Harvard universities started offering a series of courses using open and free web-based platforms. The offered courses are so-called Massive Open Online Courses (MOOCs). Currently, a number of famous MOOCs platforms such as edX, Coursera, Udacity and FutureLearn are offering a wide range of courses related to different domains. According to a MOOCs report [1], the number available online courses has increased from around 100 MOOCs in 2012 to almost 6,850 MOOCs in 2016 from over 700 universities. This interest in MOOCs by high ranked university is due to a number of features [2] such as Scalability: where MOOCs are open access to a large number of

learners around the world at the same time, Accessibility: so that learners can access learning materials in flexible way, and Openness: which means that MOOC provides courses to learners without being limited to the place and time.

Despite the widespread use of MOOCs in different universities, there are number of aspects that still need enhancements and improvements [3]. Among the aspects that needs improvements, MOOCs have a wide range of learning materials with different formats such as video, PDF, slideshows, etc. and different levels (beginner, intermediate, advanced). Such richness of learning resources can lead to overwhelming the learners and make it difficult to them to focus on a specific course [4]. Furthermore, there are a large number of learners who have different knowledge level, backgrounds, and preferences. Therefore, the solution of one size fits all can be difficult to meet the learners' needs and expectations [5]. Also, it is highlighted in [6] that the low completion rate is considered as an important issue that get researchers' interest. This low rate is because of the students are less engaged with the learning materials of the provided MOOCs. That can be because either the learning materials are so easy or too difficult for learners to follow the course. Accordingly, the students become demotivated and they have negative attitude to complete the whole course in MOOCs.

Previous drawbacks have led the researchers to propose different solutions. Among the proposed solutions and a way to diminish previous obstacles is to personalize learning materials and deliver them adaptively based on learner's knowledge level, preferences, learning outcomes [7], [8]. However, most of the proposed approaches and frameworks in literature body are depending on supporting learner with learning materials based on learners' characteristics, progress or based on learning styles. There are a limited number of studies related to considering learning outcomes in the process of selecting and delivering learning materials to MOOCs' learners [9].

The aforementioned limitations are considered as a motivation for conducting this research work. Therefore, this study aims at proposing a solution by considering both classification techniques from data mining to provide suitable learning materials for the learners in the context of MOOCs and the adaptation techniques from hypermedia. In particular, this research proposes a novel hybrid approach for MOOCs to adapt course contents and materials based on Intended Learning Outcomes (ILOs). This is done by considering two aspects. First, it classifies learning materials by exploiting the Naïve Bayesian classifier technique to match them with ILOs. Second, it delivers learning materials using adaptation techniques from adaptive hypermedia domain to generate a learning path to be followed by the learner in a specific course.

This research is considered important from different perspectives. First, learner will be able to follow customized course's learning materials so that they will be able to achieve specific learning outcomes instead of following too difficult or too easy courses. Another important aspect is that different companies will be able to upskills their employees by enabling them to follow courses based on learning outcomes, goals and objectives [10].

The rest of the paper is organized as follows: Section 2 present a state-of-the-art related work that covers topics related to this research such as learning outcomes and adaptation in MOOCs. Section 3 presents the compiled research questions and the

research methodology. After that, section 4 describes the general overview of proposed system as well as the different phases to provide adaptive MOOC. Also, it presents the developed prototype of the proposed approach. Section 5 illustrates and discusses the experiment that have been conducted to evaluate the proposed solution using precision and recall. Finally, section 6 concludes the article, presents a number of limitations related to this work, and directs researchers to future work pointers.

## **2 Related Work**

Learning Outcome (LO) is a statement that are used to clarify what learner should know and understand by the end of a course. There are different definitions for learning outcomes. For instance, the European Qualification Framework<sup>1</sup> defined it as “statements of what a learner knows, understands and is able to do on completion of a learning process, which are defined in terms of knowledge, skills, and competence”. Other definition considered learning outcomes as a set of sentences that determine what the learner wants to achieve after success completing study the course content [11].

Recently, a limited number of efforts have been made to exploit the learning outcomes in the process of retrieving learning materials in the context of online open learning resources [9]. For instance, researchers [7], proposed a prototype called “moo crank” to enable the learner exploring a number of recommended MOOCs based on the learning outcomes. However, the proposed solution recommends a number of courses to be followed based on ILOs rather than generating a course with learning materials that help the learner to achieve the ILOs. Another interesting work proposed by [12] which provided adaptive assessments in MOOCs based on learner objectives. However, the proposed work is focusing on exams and questions rather than delivering learning materials of a course adaptively.

Concerning adaptation in MOOCs, researchers [13] indicated that adaptive MOOCs can be considered as intelligent systems capable of adapting content and presentation to each learner according to their needs, objectives and interests. The adaptation process is normally performed based on the learner decisions and adaptive engine which adapts the learning materials and resources according to learner model. In this context, it is necessary to mention here that the machine learning techniques are one of the most important techniques applied in intelligent systems. Furthermore, a number of machine learning algorithms are applied in the online adaptive MOOCs domain to analyze vast amounts of complex data as well as extract information in order to enable learner to acquire knowledge. More comprehensive reviewed work

---

<sup>1</sup> [http://ec.europa.eu/eqf/terms\\_en.htm](http://ec.europa.eu/eqf/terms_en.htm)

related to providing adaptation in MOOCs using machine learning algorithms can be found in [14]–[17].

Among the different machine learning techniques, Naïve Bayesian algorithm is considered as one of the common techniques that has been explored in different domains. The reason behind the use of Naive Bayesian Classifier technique in this research is that it is easy to construct, relatively strong, fast and far from using complex repetitive variables which making it a beneficial model for large datasets [18]. Moreover, the simplicity and high precision distinguish it from other techniques. Furthermore, Naïve Bayesian technique is capable of performing the classification process in a way that surpasses most classification models such as boosted trees and random forests techniques. This makes it as an adopted algorithm in various contexts. For instance, it is used for document organization into one or more classes as email spam or not spam [19], text classification into categories [20], sentiment detection [21][22], management systems and medical diagnosis [23].

There is already a number of studies considered Naïve Bayesian algorithm in classification process in MOOCs. For instance, researchers in [18] tried to classify MOOCs video based on their metadata features such as length of video, rate, comments, etc. Two algorithms Decision tree J48 and Naive Bayesian algorithms were used to perform the classification process. The results of both algorithms are analyzed and compared, and the Naive Bayesian algorithm was found more effective than Decision tree J48 algorithm to classify MOOCs videos based on their metadata. Other researchers [24] used the Naive Bayes classifier technique to identify and adapt knowledge level of learner's in online test systems, and then determine the learner's actual ability and their competence based on performance analysis in online courses learning process. However, to the best of our knowledge, there is no work conducted to utilize Naïve Bayesian algorithm in classifying learning materials based on ILOs in MOOCs.

### **3 Research Methodology**

This section presents the research goal and the research question of our research work. Next, the different research steps that have been conducted in this research work is explained in the research methodology subsection.

#### **3.1 Research purpose and questions**

The purpose of this research was to investigate the possibility of applying data mining algorithm in classifying learning materials based on related learning outcomes. Therefore, the identified research questions are as follow: the main research question is “How to facilitate the adaptation process in MOOCs for delivering suitable learning materials?” Accordingly, there are two sub research questions

“How to utilize the Naïve Bayesian Algorithm in delivering learning materials based on intended learning outcomes?”

“How to consider pedagogical relationships between learning concepts in delivering adaptive learning materials in the context of MOOCs based on intended learning outcomes?”

To answer the formulated research questions, a research methodology was adopted. Next subsection explains the different research steps that have been conducted.

### **3.2 Research methodology**

This research work emphasizes on the improvement of learning experience in MOOCs by delivering learning materials that are suitable for learners. The difficulty of delivering learning materials that fit learners need, knowledge level, etc. in MOOCs can be considered as one of the major barriers to enhance the learning process. Therefore, this research conducted a number of scientific research steps to propose a valid solution for the formulated research problem.

Basically, the research approach is based on the Design Science Research Methodology (DSRM) [25]. Following DSRM, the research methodology includes five steps:

Problem identification and motivation which are realized after explorative study  
Delineate the objectives of a solution. The objectives are established by first understanding the current attempts in providing adaptation to the learning materials in MOOCs using data mining algorithms and specifying how different adaptation techniques were incorporated in such learning environments  
Design and development of the solution. After conducting the previous two steps, insight on how to cooperate the learning outcomes in the adaptation process was realized. Accordingly, a conceptual framework as well as prototype for validating the proposed framework were implemented  
Demonstration was conducted by collecting a dataset which is used by a prototype of the proposed framework to show results of classification using Naïve Bayesian classifier technique  
Evaluation step of the effectiveness of the proposed system was evaluated based on how the classification process was accurate in mapping learning materials and ILOs. After that, obtained results are communicated with peers for sharing purposes in international conferences and journals. Part of our work has been published in [4]. Furthermore, an iterative process to satisfy the ultimate objectives and goals is considered.

## **4 Overview of Proposed Adaptive MOOC**

This section presents a general overview of the adaptive MOOC wherein adaptation is applied based on ILOs. After that, adaptation flow in the proposed framework is explained.

#### **4.1 Adaptive MOOCs approach**

To be able to deliver adaptive MOOCs, a number of conceptual models are considered in the proposed solution similar to adaptive hypermedia and hypertext [15], [26]. Accordingly, the different models that are required are as follow: the domain model, learner model, pedagogical model, and adaptation model. Moreover, a number of repositories are also required to store different information related to learning resources and learning outcomes. Finally, adaptive engine is required to realize the different actions and tasks such as selecting learning resources, defining the learning path, etc. Next is explanation about the role of the different models, repositories and adaptive engine in the proposed adaptation approach.

The domain model is mainly used to maintain different information related to learning concepts that will be covered in specific course. Each learning concept has a relationship like is-a or part-of. Such relationships can be used to specify the different learning materials. Learning materials are stored in a specific repository. Also, each learning concept is mapped to at least one ILO stored in the corresponding repository. Each ILO has one to many relationships with other learning concepts. The main goal of mapping learning concepts and ILOs is twofold. First, it will be used to annotate an ILO as completed if the learner was able to acquire the required knowledge about learning concepts related to the ILO. Second, this will be used by the Naïve Bayesian algorithm for classification process to map ILOs with corresponding learning materials.

The learner model contains information about learner such as learner's background, knowledge level, etc. Information maintained in the learner model can be categorized into two forms:

- Static data such as name, gender, age, etc., which can be filled in when the user create his profile
- Dynamic data such as learner preferences, skills, knowledge which will be updated during the learning process

Completed learning outcome is also one of the key elements of the learner model in our approach as it is required for providing adaptivity. However, generating learner model is out of the scope of this research work. For more details about learner modelling can be found in [27].

Pedagogical model is used to maintain the pedagogical relationships between the different learning concepts defined in domain model. For instance, prerequisite pedagogical relationship is used to prevent the learner from learning a new learning concepts until he acquires the required knowledge level of prerequisite learning concept. In general, pedagogical relationships define the learning path which is the sequence on which the learning concepts should be mastered in a course.

Adaptation model is used to realize the adaptation process based on defined adaptation rules in a form of "IF-THEN" rules. Therefore, based on satisfying conditional part of the rule, action part will be performed using the adaptive engine. Similar approach of adaptation can be found in [12], [28]. The adaptation rules are realized in two phases:

**First phase** uses adaptation rules to identify the different learning concepts that are related to an ILO. For instance, a rule can be defined as IF ILO-X is selected THEN add all related learning concepts to the list of learning concepts that should be acquired during the learning process of a course.

**Second phase** uses adaptation rules to select resources that are related to learning concepts using Bayesian Knowledge Tracing algorithm. The output of this phase is delivering learning resources, guidance towards related learning resources and arrange learning resources according to the appropriateness for the learner.

Finally, the adaptive engine is used to execute the predefined adaptation rules in the adaptation model. As a result of executing the adaptation rules, the adaptive engine is responsible for send the required updates such as inserting new learning material, update learning path, inserting quizzes, etc. to the client side. At the same time, a monitoring component (in the client side) is used to record the learner's activities such as exploring learning materials and exercises results. Recorded data will be sent to the adaptive engine. As a result, the adaptive engine will update the learner model based on the predefined pedagogical relationships between learning concepts (from pedagogical model). Furthermore, the adaptive engine will allow the learner to move to next learning materials depending on pedagogical relationships between learning concepts.

In conclusion, the proposed conceptual model framework expresses different aspects to perform adaptivity to course's learning materials, learning path, navigation and presentation. More details, about the proposed conceptual framework and its requirements and characteristics can be found in [4].

Next subsection explains the adaptation workflow for the proposed solution. The case study, which is presented in the different steps, considers a student who is trying to study an algorithm course.

## **4.2 Adaptation workflow**

This subsection gives an overview of the different steps performed by the developed prototype to enable the learner to follow adaptive MOOC course. To be able to understand the adaptation flow in the proposed approach, a learning scenario explains how adaptation can happen inside adaptive MOOC course. Therefore, each step in the proposed solution will be explained with examples.

However, before start explaining the different steps in the proposed approach, it is important to mention that since our work is mainly depending on supporting learner-oriented approach, the ILOs, which are selected by the learners themselves for learning a specific topic or course, are considered as a criterion for automatically generating course with corresponding learning materials. Another important aspect to be mentioned is that both adaptation rules and pedagogical relationships between the different learning concepts are defined by the expert too. However, this issue is discussed in the limitation section.

Figure 1 shows the sequence of the steps that take actions in the proposed prototype to support learner with adaptive learning materials based on selected ILOs

in MOOCs. Each step was given a number starting from 1 till 13. The explanation of steps is distributed over the following subsections.

**Selecting learning style and ILOs:** As a starting point, the learner will be able to select his learning style such as reading from slides and PDFs, viewing videos, etc. and the course subject that he would like to learn about (*step 1 and step 2*). After that a list of related ILOs, that are stored in the ILOs repository and related to the course subject, will be displayed to be selected.

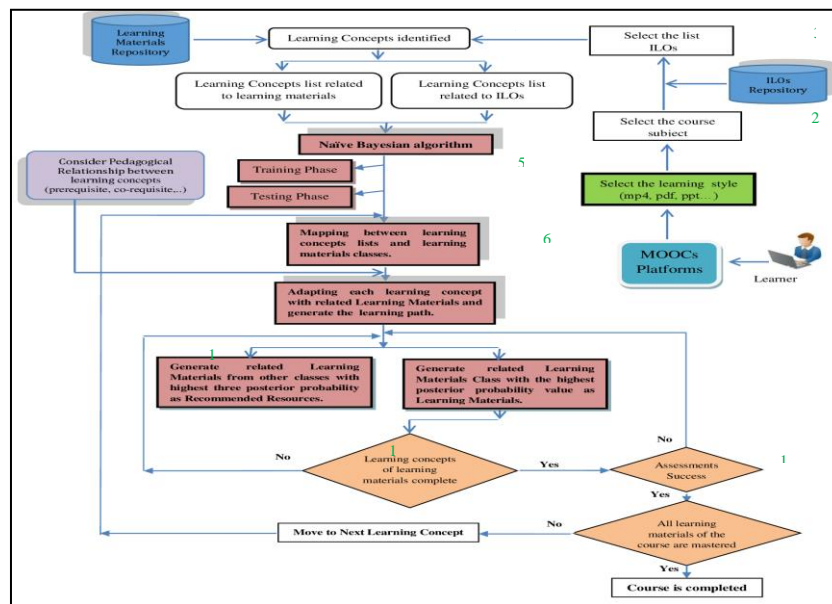


Fig. 1. Framework for adaptive MOOC.

Then, the learner will be able to select ILOs that he is intended to achieve by following the selected course (*Step 3*). Accordingly, an identification and extraction processes will be realized to match learning concepts of the selected ILOs with the similar learning concepts corresponding to the learning materials (*Step 4*). For instance, a rule can be defined as IF ILO “*Explain the major algorithms for spanning trees including Prime’s Minimum and Kruskal’s Minimum spanning tree algorithm*” is selected THEN Add all associated learning concepts (learning\_concept\_1: *Prims minimum spanning tree algorithm* and learning\_concept\_2: *Kruskals minimum spanning tree algorithm*) to the list of learning concepts that should be acquired during the learning of the generated course. As mentioned earlier, the generated course for a specific learner can be different than other learner depending on the selected ILOs.

**Automatic mapping between ILOs and learning materials:** After identifying all learning concepts and their associated learning materials, the utilization of the Naïve Bayesian algorithm is started (*Step 5*) by using two lists: the first list is related to the learning concepts that are identified out of the ILOs which are stored in the ILO



repository and the second list is related to the ILOs that are identified out of the learning materials that are stored in learning materials repository. After that the Naïve Bayesian algorithm classifies all available learning materials into classes during the training phase. Then, it calculates all posterior probabilities for each learning concept in each class during the testing phase. As a result, the learning materials class with the highest posterior probability value for each learning concept will be returned so that the learning materials in this class are presented for the learner to be explored.

To understand how the adopted algorithm works, a pseudo code is presented. After that an explanation of each step in the pseudo code is shown with example.

The following pseudo code shows the Naïve Bayesian classifier algorithm [29]:

- Given training dataset X which consists of learning concepts belonging to different classes A and B.
- Calculate the prior probability of class A= number of learning concepts of class A / total number of learning concepts.
  - Calculate the prior probability of class B= number of learning concepts of class B / total number of learning concepts.

Find  $n_i$ , the total number of learning concepts frequency of each class.

- $n_a$ = the total number of learning concept frequency of class A.
- $n_b$ = the total number of learning concept frequency of class B.

Find conditional probability of learning concept occurrence given a class.

- $P(X_1 / \text{class A}) = \text{learning-concept-count} / n_i (A)$
- $P(X_1 / \text{class B}) = \text{learning concept count} / n_i (B)$
- $P(X_2 / \text{class A}) = \text{learning concept count} / n_i (A)$
- $P(X_2 / \text{class B}) = \text{learning concept count} / n_i (B)$
- .....
- $P(X_n / \text{class B}) = \text{learning concept count} / n_i (B)$

Avoid zero frequency problems by applying uniform distribution.

Classify a new learning concept X based on the probability  $P(X/C)$ .

- Find  $P(\text{Class A} / X) = P(\text{Class A}) * P(X_1 / \text{class A}) * P(X_2 / \text{class A}) * \dots * P(X_n / \text{class A})$ .
- Find  $P(\text{Class B} / X) = P(\text{Class B}) * P(X_1 / \text{class B}) * P(X_2 / \text{class B}) * \dots * P(X_n / \text{class B})$ .

Assign learning concept to the class that has higher probability.

In general, given the learning concept classes  $C_K$  and a set of learning material  $X_i$  are represented by a vector  $X_i = (x_1, x_2, \dots, x_n)$  to a number n of learning materials. By using the Baye's theorem, the equation of conditional probability can be formulated as follow:

$$P(C_K|x_i) = \frac{P(C_K)P(x_i|C_K)}{P(x_i)} \quad (1)$$

Where:

- $P(C_K|x_i)$  is the posterior probability of learning concept class (target) given learning concept (feature).

- $P(C_K)$  is the prior probability of learning concept class.
- $P(x_i|C_K)$  is the likelihood which is the probability of learning concept given learning concept class.
- $P(x_i)$  is the prior probability of learning concept.

The following example illustrates the Naive Bayesian classifier algorithm. As indicated, given the learning materials classes (*Introduction Class, MergeSort Class, QuickSort Class, Heaps Class, Hash Tables Class, Bloom Filters Class, etc.*) and a set of learning concepts (*MergeSort, QuickSort*) will be classified. A list of learning concepts, which are related to the ILOs, are considered as an input for the algorithm. These inputs are divided into tokens in a process called tokenization process so that each token refers to a single word. After that, the resulting tokens are entered into the training phase of Naive Bayesian algorithm with the learning materials classes. In the Naive Bayesian testing phase, to determine any learning materials class associated with learning concept "MergeSort", initially, the frequency of learning concepts in each learning materials class are calculated as shown in Table 1.

**Table 1.** Learning Concepts Frequencies

		$x_1$	$x_2$
		MergeSort	QuickSort
$c_1$	Introduction Class	5 times	3 times
$c_2$	MergeSort Class	20 times	6 times
$c_3$	QuickSort Class	10 times	15 times
$c_4$	Heaps Class	3 times	0 times
		38 times	24 times

The likelihood of the learning concepts in each learning material class are calculated as shown in the Table 2.

**Table 2.** Likelihood Table for Two Learning Concepts

		$x_1$	$x_2$	
		MergeSort	QuickSort	
$c_1$	Introduction Class	5 / 38	3/24	8/62
$c_2$	MergeSort Class	20 / 38	6 / 24	26/62
$c_3$	QuickSort Class	10 / 38	15 / 24	25/62
$c_4$	Heaps Class	3 / 38	0	3/62
		38 / 62	24 / 62	

The prior probability of learning concept “MergeSort” is evaluated as follow:  $P(x_1) = P(MergeSort) = 38/62 = 0.6129$ . The posterior probability of learning concept class “IntroductionClass” given learning concept “MergeSort” is evaluated as follow:  $P(x_1|c_1) = P(MergeSort | IntroductionClass) = 5/38 = 0.132$ . Finally, the prior probability of learning concept class “IntroductionClass” is evaluated as follow:  $P(c_1) = P(IntroductionClass) = 8/62 = 0.129$ .

By using Baye’s theorem equation, the posterior probability of learning concept “MergeSort”to first learning materials class “IntroductionClass”is calculated as follows:  $P(c_1|x_1) = P(IntroductionClass | MergeSort) = ( P(IntroductionClass) P(MergeSort | IntroductionClass) ) / P(MergeSort) = (0.129 * 0.132) / 0.6129 = 0.02778$ .

Accordingly, all previous calculations are repeated to calculate all posterior probabilities for each learning concept in each learning materials class. After that, the equation of the Naïve Bayesian Classifier is applied to return the Maximum posterior probability value, which refers to the learning materials class that are related to the learning concept (Step 6).

**Learning path generation:** After assigning learning materials that will be covered in the course, a learning path is generated based on the defined pedagogical relationships between the different learning concepts (Step 7, Step 8). Accordingly, the generated learning path defines the sequence in which the learner can start exploring the learning materials of the learning concepts. For instance, *MergeSort* is prerequisite for *QuickSort*. From pedagogical point of view, this means that the learner will not be able to learn *QuickSort* learning concept till he acquired knowledge about *MergeSort* learning concept. From implementation point of view, this means the learning materials of the *QuickSort* learning concept will not be accessible until all learning materials of *MergeSort* learning concept have been completed and the learner passes the assigned quiz to *MergeSort* learning concept. Figure 2 shows a screenshot of the developed prototype to display the generated learning path that will be followed by the learner and it also shows how the prototype prevent learner from exploring learning materials that are related to *MergSort* learning concept because of the prerequisite pedagogical relationship with *QuickSort* learning concept.

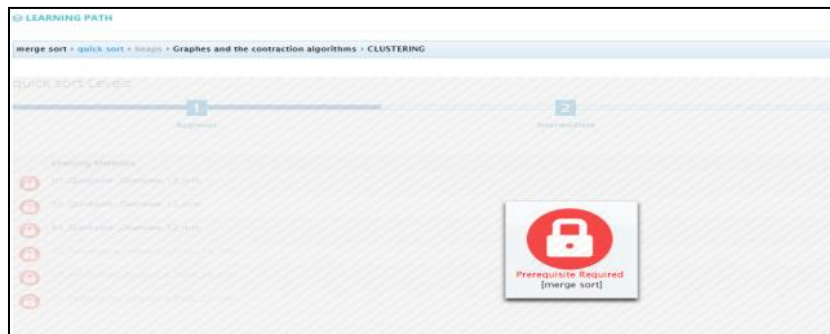


Fig. 2. Generated learning path for following the course.

**Recommending learning materials:** As shown in flow diagram (Figure 1), the output of utilizing the Naïve Bayesian algorithm is twofold. The first output is related to matching learning concepts from ILOs with learning materials (Step 9). The process of generating the first output is explained earlier in subsection 3. The second output is related to recommend three new learning concepts that can be used to increase the learner knowledge about the current learning concept (Step 10). This is

realized by updating Naïve Bayesian Algorithm to provide the learner with a list of three recommended learning materials that are also related to the learning concept which is currently explored by the learner.

Updates on Naïve Bayesian algorithm is done as follow. First, removing the learning materials class with the highest posterior probability value from the learning materials classes in training phase of the algorithm. Consequently, the algorithm does not retrieve the same learning materials that are related to the learning concept as a recommended resource. Second, the algorithm is recalled again with the updated classes. But here, the inputs to the algorithm will be the learning concept name and the name of the learning material that has been selected for reading by the learner.

**Assessment aspects:** To complete the learning process, the system offers assessments to evaluate the learner understanding of each learning concept after completing all learning materials that are associated with it (*Step 11*). Consequently, if the learner does not exceed the assessments threshold score, then the system re-presents the same learning materials that are related to the learning concept to be learned again (see Figure 3). It is important to mention here that the mapping between the assessments and the learning concepts is also done manually.

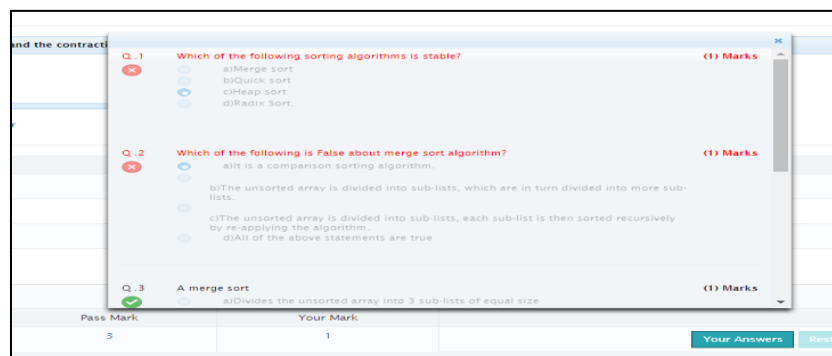


Fig. 3. Generated learning path for following the course

After completing assessments that are related to the currently explored learning concept successfully (*Step 12*), the system will check if the learner managed to master all learning concepts that are related to all selected ILOs (*Step 13*). If the learner completed all learning concepts, then he will be able to finish the course. Otherwise, the learner will be directed to explore the learning materials that are related to next learning concept. This is an iterative process which will be finished upon exploring all learning materials that are related to the list of learning concepts.

## 5 Evaluation Results and Discussion

This section describes the experiment that was performed to evaluate the effectiveness of the proposed system. The effectiveness of the proposed system is

evaluated based on accuracy aspects in specifying of relevance scores between (ILOs) and corresponding learning materials.

### 5.1 Experiment set-up

To test the effectiveness of the constructed classification model, a prototype was implemented using Java programming language 8, WEKA library, Net Beans IDE 8.0.1, SQL Navigator 6.2, Web application, using Java Server Faces (JSF), Oracle Database, Web server Glassfish 4. After developing the prototype, it was tested using a PC with core i3 CPU (2.5GHz) and (4 GB) RAM. The operating system is Windows 7.

Concerning datasets, a dataset of learning materials for a course were collected from Coursera MOOC platform. The course is called “Algorithms Design and Analysis”. The collected learning resources are composed of different learning materials formats such as video, PDF and slides, textual explanations, quizzes, exercises, etc. Concerning ILOs, a dataset consists of 48 different (ILOs) obtained from different courses syllabus related to algorithms design and analysis course. The datasets include also four courses covering two different subjects (data mining and algorithm design) with (1518) learning materials in different formats such as (.pdf, .mp4, .ppt). This is done for the purpose of having different learning resources to be classified by the developed prototype.

Then, after collecting required learning resources and ILOs, an expert reviewed the collected ILOs and learning materials to create a training dataset which includes a list ILOs and a list of their corresponding learning materials. Furthermore, the list of ILOs is used by the developed prototype to enable learners to select their ILOs.

### 5.2 Results and discussion

The performance of the classification model was measured in terms of number of both Precision (P) and Recall (R). Both are considered as indicators utilized to measure the quality of the obtained outcomes [30]:

$$P = \frac{|relevant\ LM \cap retrieved\ LM|}{|retrieved\ LM|} \quad (2)$$

$$R = \frac{|relevant\ LM \cap retrieved\ LM|}{|relevant\ LM|} \quad (3)$$

For instance, to calculate the Precision/Recall (P/R) indicators for the ILO “*Explain the most important graph-processing problems*”, the following steps were conducted:

- **First**, it calculates the number of learning resources that are relevant to the learning concept “*Graphs and the contraction algorithms*”, and their number was (39 learning materials).

- **Second**, it calculates the number of learning resources that are retrieved to the same learning concept “*Graphs and the contraction algorithms*” by the system during the learning process, and their number was (44 learning materials).
- **Third**, it calculates the number of learning materials relevant to the learning concept “*Graphs and the contraction algorithms*” and have been retrieved by the system during the learning process, the number of learning materials was (32 learning materials).

By applying the equations (2,3) mentioned above, the results of the indicators were as follow:

$$P = (32 / 44) * 100\% = 0.727 * 100 \% \cong 73 \% .$$

$$R = (32 / 39) * 100\% = 0.820 * 100 \% \cong 82 \% .$$

Similarly, the process is repeated to calculate the Precision and Recall(P/R) indicators for the remaining ILO's. The following table clarifies results for three ILOs:

**Table 3.** Likelihood Table for two Learning Concepts

ILOs	P	R
Explain the most important graph-processing problems	73%	82%
Explain the major algorithms for spanning trees including Prime's Minimum and Kruskal's Minimum spanning tree algorithm	80%	88%
Study and implement the Minimum spanning tree as an application to Clustering	71%	86%

As seen in Table 3, it can be noted that the precision indicator results range between 71% and 80% percent while the results of the recall indicator range between 82% and 88% percent. This means that the Naive Bayesian is able to do its work in terms of retrieving the learning materials and recommending other learning resources based on selected (ILOs) almost effectively.

There are many previous studies that adopted the use of Naive Bayesian Classifier algorithm for the classification texts and documents. Also, the studies' results indicated that obtained classified materials have a high accuracy and the algorithm had the ability to work in a way that surpasses the most advanced classification models such as boosted trees or random forests models (e.g.,[20], [19]-[31]) classifying learning materials based on Decision Tree algorithm [32]. However, a comparison between our work and previous mentioned research is not valid as they have been used in different contexts.

## 6 Conclusion and Limitations

This paper presents an approach to deliver adaptive MOOCs. The proposed solution consists of mapping learning materials with ILOs based on common learning concepts which are classified by utilizing the Naïve Bayesian algorithm. As a proof of

concept, a prototype has been developed to classify learning materials using Naïve Bayesian algorithm and generate learning path based on pedagogical relationships between the different learning concepts. The experiment showed an appropriate classification between learning materials and ILOs according to the identified learning concepts. It revealed that a course can be generated with learning materials that are related to learning outcomes selected by a learner. This is done by classifying, using data mining algorithm called Naïve Bayesian, both learning materials and learning outcomes based on so-called learning concepts. After classification phase is done, assigning learning materials with highest percentage to different learning concepts was validated to create the course's contents. After specifying which learning materials should be included in the course, a learning path is generated automatically by considering so called pedagogical relationships between the different learning concepts. For instance, learners will be able to explore some learning materials related to a learning concept after exploring learning materials that are related to the prerequisite learning concepts.

Concerning the limitations, it is important to mention that our proposed approach for adaptivity is an author driven approach. This means that the course's author needs to define all possible adaptation techniques that can be applied to the course contents, presentation and navigation explicitly. Also, he needs to specify the pedagogical relationship manually. This can be considered as a disadvantage. However, supporting authors with usable and appropriate tools can lessen the workload required by the authors. It is also important to notice that this work does not consider the creation process of the learning resources and course contents as there are many available platforms and tools in the context of Open Educational Resources (OER) [33].

A future work is related to investigating the effectiveness of the proposed approach from pedagogical point of view by involving learners in the evaluation phase of the developed prototype. Since the adopted algorithm evaluated the learning materials, not the learner, it is quite important to consider the learners' scores after exploring learning materials and conducting exams and quizzes. Scores can reflect the comprehension level of the learning concepts. In general, having good scores means the learner was able to understand the learning concepts and it proves the relevance between the classified learning materials and the learning concepts using our proposed solution. This can be considered as a motivation for the learner to continue the course and complete it. Accordingly, it will affect the dropout rate issue in MOOCs. Another future work will apply the proposed solution to other courses to determine the level of its compatibility and students' trust [34].

## **7 References**

- [1] Dhawal Shah, "By the Numbers: MOOCs in 2016," 2016. [Online]. Available: <https://www.classcentral.com/report/mooc-stats-2016/>. [Accessed: 11-Apr-2019].
- [2] A. Kahaei, "Design of Personalization of Massive Open Online," p. 87, 2015.
- [3] S. F. A. Fesol, "Towards MOOC for Technical Courses: A Blended Learning Empirical Analysis," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, pp. 1141–1147, 2016.

- [4] A. Ewais and Duaa Abu Samra, “Adaptive MOOCs: A Framework for Adaptive Course based on Intended Learning Outcomes,” in IEEE International Conference on Knowledge Engineering and Applications (ICKEA 2017), 2017. <https://doi.org/10.1109/ickea.2017.8169930>
- [5] D. F. O. Onah and J. E. Sinclair, “Design Science MOOC: A Framework of Good Practice Pedagogy in a Novel E-Learning Platform eLDa,” in 28th Annual World Conference on Educational Media and Technology EDMEDIA2016, 2016, no. June, pp. 1–7.
- [6] S. F. A. Fesol, S. Salam, and N. Bakar, “Wearable Technology in Education to Enhance Technical MOOCs,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 5, pp. 1873–1881, 2018. <https://doi.org/10.18517/ijaseit.8.5.3929>
- [7] I. Gutiérrez-Rojas, C. Alario-Hoyos, M. Pérez-Sanagustín, D. Leony, and C. Delgado-Kloos, “Towards an Outcome-based Discovery and Filtering of MOOCs using moocrank,” *Proc. Eur. MOOC Stakehold. Summit 2014*, pp. 50–57, 2014.
- [8] N. A. Abdullah, S. White, H. C. Davis, C. Knowledge, and G. Siemens, “Personalisation of MOOCs,” *7th Int. Conf. Comput. Support. Educ.*, pp. 88–97, 2015.
- [9] O. Pilli and W. Admiraal, “Students’ Learning Outcomes in Massive Open Online Courses (MOOCs) : Some Suggestions for Course Design.,” *J. Higher Educ.*, vol. 7, no. 1, pp. 46–71, 2017. <https://doi.org/10.2399/yod.17.001>
- [10] D. Ong and M. Jambulingam, “Reducing employee learning and development costs: the use of massive open online courses (MOOC),” *Dev. Learn. Organ.*, vol. 30, no. 5, pp. 18–21, 2016. <https://doi.org/10.1108/dlo-08-2015-0066>
- [11] S. G. Yildirim and S. W. Baur, “Development of Learning Taxonomy for an Undergraduate Course in Architectural Engineering Program,” pp. 1–10, 2016.
- [12] Y. Rosen, I. Rushkin, A. Ang, C. Federicks, D. Tingley, and M. J. Blink, “Designing Adaptive Assessments in MOOCs,” in *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale - L@S ’17*, 2017, pp. 233–236. <https://doi.org/10.1145/3051457.3053993>
- [13] A. Teixeira, J. Mota, A. García-Cabot, E. García-Lopéz, and L. De-Marcos, “A new competence-based approach for personalizing MOOCs in a mobile collaborative and networked environment Un nuevo enfoque basado en competencias para la personalización de MOOCs en un entorno móvil colaborativo en red,” *Rev. Iberoam. Educ. a Distancia*, vol. 19, no. 1, pp. 143–160, 2016. <https://doi.org/10.5944/ried.19.1.14578>
- [14] S. Ardchir, M. A. Talhaoui, and M. Azzouazi, “Towards an Adaptive Learning Framework for MOOCs,” in *E-Technologies: Embracing the Internet of Things: 7th International Conference, MCETECH 2017, Ottawa, ON, Canada, May 17-19, 2017, Proceedings*, E. Aïmeur, U. Ruhi, and M. Weiss, Eds. Cham: Springer International Publishing, 2017, pp. 236–251. <https://doi.org/10.1007/978-3-319-59041-7>
- [15] L. Leka, A. Kika, and S. Greca, “Adaptivity in e-learning systems,” in *CEUR Workshop Proceedings*, 2016.
- [16] D. Lers, M. L. Sein-Echaluce, M. Hernandez, and C. Bueno, “Validation of Indicators for Implementing an Adaptive Platform for MOOCs,” *Comput. Hum. Behav.*, vol. 72, no. C, pp. 783–795, 2017. <https://doi.org/10.1016/j.chb.2016.07.054>
- [17] G. Sun, T. Cui, W. Guo, G. Beydoun, D. Xu, and J. Shen, “Micro Learning Adaptation in MOOC: A Software as a Service and a Personalized Learner Model,” in *Advances in Web-Based Learning -- ICWL 2015: 14th International Conference, Guangzhou, China, November 5-8, 2015, Proceedings*, F. W. B. Li, R. Klamma, M. Laanpere, J. Zhang, B. F. Manjón, and R. W. H. Lau, Eds. Cham: Springer International Publishing, 2015, pp. 174–184. [https://doi.org/10.1007/978-3-319-25515-6\\_184](https://doi.org/10.1007/978-3-319-25515-6_184)



- [18] E. L. H. Othman, G. Abderrahim, and E. L. B. Jaber, “Moocs Video Mining Using Decision Tree J48 and Naive Bayesian Classification Models,” vol. 1, no. 1, pp. 26–32, 2017.
- [19] V. Metsis, I. Androutsopoulos, and G. Paliouras, “Spam filtering with naive bayes-which naive bayes?,” *Ceas*, p. 9, 2006.
- [20] R. P. Rajeswari and K. Juliet, “Text Classification for Student Data Set using Naive Bayes Classifier and KNN Classifier,” *Int. J. Comput. Trends Technol.*, vol. 43, no. 1, pp. 8–12, 2017. <https://doi.org/10.14445/22312803/ijctt-v43p103>
- [21] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” *Proc. 23th Int. Conf. Mach. Learn.*, pp. 161–168, 2006. <https://doi.org/10.1145/1143844.1143865>
- [22] T. Zhu, W. Wang, W. Zhao, and R. Zhang, “Participation Prediction and Opinion Formation in MOOC Discussion Forum,” *Int. J. Inf. Educ. Technol.*, vol. 7, no. 6, pp. 417–423, 2017.
- [23] I. Rish, “An empirical study of the naive Bayes classifier,” *IJCAI 2001 Work. Empir. methods Artif. Intel.*, vol. 22230, pp. 41–46, 2001.
- [24] S. Agarwal, N. Jain, and S. Dholay, “Adaptive testing and performance analysis using naive bayes classifier,” *Procedia Comput. Sci.*, vol. 45, no. C, pp. 70–75, 2015. <https://doi.org/10.1016/j.procs.2015.03.088>
- [25] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A Design Science Research Methodology for Information Systems Research,” *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, 2007. <https://doi.org/10.2753/mis0742-1222240302>
- [26] M. Alshammari, R. Anane, and R. J. Hendley, “Adaptivity in E-learning systems,” in *Proceedings - 2014 8th International Conference on Complex, Intelligent and Software Intensive Systems, CISIS 2014, 2014*. <https://doi.org/10.1109/cisis.2014.12>
- [27] Y. Long and V. Aleven, “Enhancing learning outcomes through self-regulated learning support with an Open Learner Model,” *User Model. User-adapt. Interact.*, vol. 27, no. 1, pp. 55–88, 2017. <https://doi.org/10.1007/s11257-016-9186-6>
- [28] L. Nguyen and P. Do, “Learner Model in Adaptive Learning,” *World Acad. Sci. Eng. Technol.*, vol. 45, no. November, pp. 395–400, 2008.
- [29] S. Sathyadevan, D. M. S., and S. G. S., “Crime analysis and prediction using data mining,” *2014 First Int. Conf. Networks Soft Comput.*, no. August, pp. 406–412, 2014.
- [30] C. D. Manning, P. Ragahvan, and H. Schutze, “An Introduction to Information Retrieval,” *Inf. Retr. Boston*, no. c, pp. 1–18, 2009.
- [31] J. Yoo, M. Lee, G. Aloyce, and D. Yang, “Creating a Naïve Bayes Document Classification Scheme Using an Apriori Algorithm,” vol. 133, pp. 34–38, 2016. <https://doi.org/10.14257/astl.2016.133.07>
- [32] F. Elghibari, R. Elouahbi, and F. El Khoukhi, “Data Mining for Detecting E-learning Courses Anomalies: An Application of Decision Tree Algorithm,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 3, pp. 980–987, 2018. <https://doi.org/10.18517/ijaseit.8.3.2756>
- [33] Y. Park and Y. Kim, “A Design and Development of micro-Learning Content in e-Learning System,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 1, pp. 56–61, 2018.
- [34] Eamon Costello, James Brunton, Mark Brown, & Laurence Daly. “In MOOCs we Trust: Learner
- [35] Perceptions of MOOC Quality via Trust and Credibility”. *International Journal of Emerging Technologies in Learning (IJET)*, 13(6). 2018. pp.214-222. <https://doi.org/10.3991/ijet.v13i06.8447>

## 8 Authors

**Ahmed Ewais** is assistant professor in Computer Science in Arab American University and Postdoc researcher in Web & Information System Engineering (WISE) laboratory in Vrije Universiteit Brussel, Belgium. He earned his bachelor degree in computer science from Arab American University (AAUP) Palestine in 2004, master in computer science from the Vrije Universiteit Brussel, Belgium in 2008, where he also obtained his PhD in 2014. His research is situated in the context of E-Learning, Massive Open Online Courses (MOOCs), Virtual Reality, Augmented Reality, Human Computer Interaction, UX&UI, Usability Evaluation. (E-mail: [awais@aaup.edu](mailto:awais@aaup.edu), [awais@vub.ac.be](mailto:awais@vub.ac.be), [awais@gmail.com](mailto:awais@gmail.com))

**Duaa Abu Samra** was born in Jenin, Palestine, in 1986. She received her master's Degree in Computer Science from the Arab American University, Palestine, 2018. During her career, she worked in several administrative Offices and different schools in Palestine. Her main research interests include evaluating the performance of Machine Learning Algorithms in Massive Open Online Courses (MOOCs), E-Learning and educational technology. (E-mail: [duaa.abuSamrah@aaup.edu](mailto:duaa.abuSamrah@aaup.edu))

Article submitted 2019-08-01. Resubmitted 2019-11-05. Final acceptance 2019-11-21. Final version published as submitted by the authors.