Cognitive Evaluation of Examinees by Dynamic Question Set Generation based on Bloom's Taxonomy

Anjan Dutta[1], Punyasha Chatterjee[1], Nilanjan Dey[2], Pablo Moreno-Ger[3] and Soumya Sen[4]

[1]School of Mobile Computing and Communication, Jadavpur University, Kolkata, 700098, India; [2]Department of Computer Science and Engineering, Techno International NewTown, Kolkata, India; [3]School of Engineering, Universidad Internacional de La Rioja (UNIR), Logroño, Spain; [4]A.K.Choudhury School of Information Technology, Calcutta University, Kolkata, 700106, India

ABSTRACT

Educational data mining (EDM) is an emerging topic in recent years steered by data mining and machine learning techniques to enhance students' overall learning experience and academic progress. In recent years EDM techniques are frequently used to improve assessment systems but the evaluation procedure is majorly marks driven. Developing an evaluation system to distinguish candidates, based on their ability to answer cognitively difficult questions is a challenging task. In this study, a unique methodology is proposed to dynamically rank the candidates to develop an outcome-based online examination system that will properly evaluate a candidate's cognitive competencies. The questions are segmented into different cognitive groups based on classical Bloom's educational taxonomy. The Jenks Natural Breaks Optimization technique is used here to segment the questions and as a result, distinct question clusters based on different cognitive levels are obtained. Students are evaluated with different questions from these cognitive groups and ranking is done for individual candidates considering both the marks of the questions and his/her ability to solve questions from different difficulty levels.

KEYWORDS

Bloom's taxonomy; cognitive level; clustering; educational data mining (EDM); Jenks Natural Breaks Optimization; machine learning; ranking

1. INTRODUCTION

In the last decade a revolution in education was observed with the introduction of Artificial Intelligence (AI)-based techniques [1]. The future trend in the modern educational system is the deep integration of AI with the existing methodology to enhance the overall learning system. Every student is considered an independent, dynamic and continuously learning individual having immense development potential by emphasizing the concept of AI in the evaluation process [2]. Machine learning is an important part of AI that helps to find out the patterns from the data resulting in the discovery of new knowledge that has immense potential for the benefit of society [3]. An emerging multidisciplinary research field has been evolved through data mining and machine learning methods in education which is known as educational data mining (EDM) [4]. EDM applies machine learning methods to educational data to find out the hidden patterns to predict the students' behavioral changes and individual performances. The evaluation of a student is traditionally done by written examinations in educational institutes. In recent years researchers have proposed different approaches to generate examination questions to analyze the efficacy of the existing learning method [5]. Among different approaches, questions prepared by matching the intended learning outcomes (ILO) of a particular course module are proven to be effective to measure the overall learning progress of a student. Therefore, academic bodies are trying to follow a well-defined taxonomy to produce high-quality question papers that can judge over cognitive levels of a student. The learning outcomes, based on the course syllabi can be described through an educational taxonomy. Furthermore, an overview of the various levels of understanding about a particular learning topic is achieved by using these outcome-based taxonomies.

Bloom's taxonomy [6] and its modified version [7] are largely popular among the proposed taxonomies. Bloom's taxonomy is based on six cognitive learning levels (i) remembering (ii) understanding (iii) application (iv) analyzing (v) evaluate and (vi) create. At the remembering level the memorization ability of a student, regarding the facts and fundamental information, is evaluated e.g. "Label different phases of the life cycle of a process". Student's ability to understand different topics and concepts based on prior acquired knowledge is measured in the understanding level e.g. "Describe different types of stack operations". Implementation skill of the acquired knowledge is judged at the application level e.g. "Apply stack operations to implement a queue". The analysis level deals with assessing the student's ability to analyze the gained knowledge and ability to distinguish between facts and different options e.g. "Compare merge sort and quick sort by analyzing their time complexities". A student's ability to defend and evaluate based on some criteria is measured in the evaluation level e.g. "Assess the performance of linear search and binary search on a sorted array". In the create level,a student's ability to reorganize and reconstruct elements to synthesize new patterns or ideas is measured e.g. "Design a model of smart hospital management system". Different levels of Bloom's taxonomy with the action verbs are mentioned in [8]. The lower level of thinking skills is needed in the remembering and understanding level and the difficulty level increases gradually in the rest of the levels. The expected learning outcomes of a course are described by identifying particular action verbs. In 2009, the revised version of Bloom's taxonomy [9] was introduced where cognitive dimensions are mapped with the knowledge dimensions. The cognitive domain introduced in Bloom's taxonomy is a hierarchical structure consisting of the aforesaid six levels. In present days more and more importance is given to outcome-based learning systems [10–12] for the overall growth of a student and for the all-round development from understanding a topic and to the extent of applying the gathered knowledge to design a new model or concept. Nowadays, the online mode of education is gaining popularity for its effectiveness and recent research studies have demonstrated their encouraging effect on achieving learning outcomes [13]. Moreover, in this coronavirus disease (COVID-19) pandemic situation more and more countries are relying on online modes to manage and maintain academic activities. Nearly 1.2 billion students are affected due to school closures worldwide [14]. In a recent research by UNESCO [15] nearly twenty-three countries are depending on online assessments to monitor and evaluate students' academic progress. Learners' progress needs to be assessed to identify learning gaps and plan for the remedial classes once the schools reopen. Therefore, a proper online assessment system is necessary that can judge the competency of the student in every cognitive domain. In this outcome-based educational system, the existing evaluation method is facing challenges such as (i) whether the question paper has a proper mix of questions that can analyze all the cognitive levels of a particular candidate (ii) whether the evaluation system is competent enough to identify the candidates who are solving analytically more difficult questions compared to the other candidates. In recent years, several research studies have been carried out in the area of cognitive analysis in the educational domain. The majority of the study is focused on the analysis of question papers or course content to know its effectiveness to evaluate the cognitive achievements of the students. As to the best of our knowledge, no existing evaluation system has been developed to date where the candidates are ranked dynamically by analyzing their cognitive skill set. The question sets are generated dynamically based on the performance of the candidates reflect their cognitive levels. The novelty of this research work is the proposal of dynamic question set generation based on bloom's taxonomy that evaluates them based on their cognitive level. Henceforth, the proposed method is fully customized for every candidate who is appearing in this examination system. In the proposed framework the candidates are categorized based on his/her proficiency in answering the questions adhering to different learning levels of Bloom. The main contributions of this work are - (i) proposing a novel methodology to measure the difficulty level of a question by qualitative analysis and classifying into various difficulty groups per Bloom's outcome-based education taxonomy (ii) introduction of a dynamic candidate ranking system in the online examinations which not only considers the total marks obtained by a candidate but also gives importance to the cognitive skills achieved by the candidates. A novel algorithm to classify or segment one-dimensional data is Jenks Natural Breaks Optimization (JNBO) [16] which is largely popular in the Geographical Information System (GIS). In this work, the JNBO method is experimentally used to classify the questions into different difficulty groups.

The rest of the paper is organized as follows: In Section 2 recent studies regarding the application of Bloom's Taxonomy in the question paper generation are vividly discussed. The application of unsupervised learning in academics is also discussed here. The objective of this work is described in Section 3 and a brief discussion of the JNBO method is given in Section 4. The proposed automatic clustering of the questions according to Bloom's Taxonomy is explained in Section 5. In Section 6 the suggested approach to dynamically rank the candidates based on their competencies in the different Bloom's cognitive domains is presented. The implementation of the suggested method and results are discussed in Section 7. The conclusion and future scope of this work are mentioned in Section 8.

2. RELATED WORK

In recent years unsupervised learning methods are widely applied to analyze students' learning achievements. In [17] unsupervised learning is applied to analyze the effect of students' demographic characteristics and online learning engagements on overall learning achievements. One of the unsupervised learning techniques i.e. the k-Means clustering technique is used by several researchers to improve the overall education system [18]. k-Means clustering is applied to analyze the distribution of Indonesian high school teachers [19]. After preprocessing the selected dataset, k-Means clustering is applied to obtain clusters of three categories: less, adequate and excess teachers. The result of this research is helping to solve the issues arising from the uneven distribution of teachers across the country. In another study [20] the k-means clustering algorithm is used to improve the academic performance of the students in MOOCs. Different parameters such as the number of assignments completed, number of posts on the course discussion forum, course video view count, task completion percentage and the final score are considered to establish the relationship between online learning behavior and overall academic progress. K-Means clustering is used to perform predictive analysis [21] of students' performance by analyzing their answer scripts. A hybrid approach, based on a decision tree and data clustering, is proposed in [22] for the prediction of students' Grade Point Average (GPA) that acts as a reference for educators to improve the performance of the students. Cluster analysis has the great potential to understand students' learning behavior from hyperlinked knowledge sources by analyzing the log files generated by the web servers. In [23] k-Means clustering is applied to analyze students' learning behavior during real-time online problem-solving. An analysis is carried out in [24] relating the course evaluation by the university students and their corresponding results. Clustering techniques are used here to find the correlation between the course evaluation and average results made by the students. This study also helped to point out the regularities of the courses that emerge over the years. K-Prototype clustering is utilized here to form groups or clusters of students considering the demographic parameters and students' interaction in online learning platforms. The learning achievement of individual groups is investigated to know the performance of each group of students. Hence insight is obtained to design online courses that adapt to students' needs. Unsupervised learning methods [25] are applied to explore the textual data in the discussion forum of Massively Open Online Courses (MOOCs) online performance. Understanding the comments in the discussion forum helps to provide adaptive support based on the students' needs. Clustering methods are used here to group similar posts and these clusters are compared and analyzed by the MOOC researchers. Knowledge and skillset acquired by a group of people are largely affected by the nature of these groups. Therefore, the effectiveness of learning is greatly improved by forming the groups systemically and intelligently. An adaptive framework [26] is explored to form groups to optimize the intended performance criteria. Linear regression is used here to dynamically update the rules used for forming the groups. An unsupervised approach is used to analyze students' learning characteristics in an open-ended learning environment [27]. Students are grouped by their learning behaviors by combining feature selections and clustering methods. The learning behaviors of individual groups are analyzed and linked to the students' potentiality to develop accurate models. Clustering methods like k-Means clustering are efficient for multidimensional data, whereas for classifying 1-d data a popular and efficient method is the JNBO technique. Alhough to the best of our knowledge there is no existing work where JNBO is applied in the education domain, the Jenks method is applied frequently in Geographical Information Systems (GIS) [28]. In recent times several studies [29,30] have been reported where Jenks optimization is widely used to classify Geographical environment Units.

Bloom's taxonomy is widely studied in the literature for improving teaching and learning efficiency [31]. Disregarding the pyramid structure of Bloom was identified as the major cause of laboratory examination failure in engineering courses [32]. The medium-level and higher-level knowledge are disrupted due to the lack of lower-level knowledge that results in incomplete execution of the laboratory assignments. Bloom's taxonomy was used for creating well-structured assessments and evaluating the cognitive levels [33] of students of computer science courses. Text analysis is used here to automate the question-generation process based on Bloom's learning taxonomy. The authors have demonstrated 81.35% accuracy by implementing the proposed method. A distinction is made between higher-order and lower-order questions in [34] based on Bloom's taxonomy. This study aimed to assess whether the engineering academics were preparing questions to evaluate the analytical and problem-solving skills of the students. In [35] Bloom's taxonomy levels are discussed and interpreted individually with suitable exemplars. This is a useful reference for educators in the computer science domain to follow Bloom's taxonomy for building outcome-based programming assessments. Here the authors demonstrated that the quality of assessments in fundamental programming courses is greatly improved by designing the examination following Bloom's taxonomy. The basic level of programming skills of the students was measured by changing programming questions based on Bloom's taxonomy levels [36]. The developed framework has resulted in relating the student performance with the cognitive level of the questions. A systematic review of the automatic generation of the MCQ question-based assessment system is given in [37]. In this study the authors have presented a generic workflow to automate the MCQ question generation system.

A vivid literature review is done for each phase of the workflow. Different techniques to evaluate the MCQ-based automated question paper generation systems are also elaborated here. A new method of assigning weights and categorizing examination questions was suggested in [38]. A unique rule set is developed here using Natural Language Processing (NLP) and Wordnet similarity algorithms to distinguish categories and assign weights to the examination questions following Bloom's taxonomy. In another research work [39] a novel methodology to classify questions in the absence of verbs was proposed. Wordnet similarity algorithm and Cosine similarity are also used here to classify the questions according to Bloom's taxonomy levels. An automated question paper generation methodology was proposed using a genetic algorithm [40] to categorize the questions into six levels of Bloom's taxonomy. TF-IDF statistical feature [41] was used to classify the examination questions to maintain the cognitive domain of Bloom's taxonomy. Classification is done by using three popular machine learning methods, namely Support Vector Machine (SVM), K-Nearest Neighbor and Naïve Bayes. In [42] the authors presented a framework based on Bloom's taxonomy for mapping the questions with the intended learning outcomes. Here a prototype of an automated question generation tool is demonstrated which evaluates the achievement of the expected learning outcomes. The performance of SVM in Bloom's taxonomy-based classification of questions was analyzed in [43]. The effect of removing stop words and the frequency of the keywords were considered here to evaluate the performance of the SVM. A unique system for automatic identification of Bloom's cognitive levels of each question was proposed to aid educators in preparing the examination questions [44]. The shuffling algorithm is used here as a technique for randomization. Different modules e.g. user admiration, subject choice and specification of the related course outcomes, question paper generation and management were included in the system.

3. OBJECTIVE

The current evaluation system of the candidates in the examinations is based on the obtained score. No reward is given to the candidates who are solving more difficult questions to achieve the score compared to other candidates. A multi-criteria decision-making solution is adapted here to classify questions into different difficulty levels. Thereafter, a novel methodology is proposed here to dynamically rank the candidates based on their cognitive skillsets. Certain parameters are considered to measure the difficulty level or hardness of a particular question:

i) Marks assigned to the question
ii) Which Bloom's level it belongs
iii) How much it is related to a particular Bloom's level

Next, the ranking is done dynamically depending on the candidate's competency in solving questions belonging to different hardness groups.

4. JENKS NATURAL BREAKS OPTIMIZATION

Jenks Natural Breaks Optimization algorithm was introduced in 1977 [45] as a methodology for optimal data classification. The algorithm was primarily based on exact optimization techniques developed by Fischer in 1958 [46]. The JNBO method was mainly used for analyzing geographic data and over the year it evolved as a standard data classification algorithm in Geographical Information Systems (GIS). In an industry-leading GIS software package ArcGIS [47] the JNBO method is used as the default classification algorithm. Numerous research studies [48,49] primarily in the GIS domain have validated the reliability of this algorithm. The core principle of JNBO is to find the best possible arrangements in a series of data [50,51]. This property of Jenks is utilized in this study to segment the questions into different hardness groups based on Bloom's outcome-based learning taxonomy. Classic clustering techniques such as k-Means [52], DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [53] and Gaussian Mixture Model [54] are applied to multivariate data. Jenks Natural Breaks Optimization is similar to one-dimensional k-Means algorithm. In this study clustering algorithm is applied to one-dimensional data i.e. the hardness factor of the questions. Since the JNBO method is applied on single-dimensional data, hence it is used here as the clustering method. As a result, unnecessary mathematical calculations and complexities are avoided and accurate clusters of questions are obtained. Density Based clustering methods are not applied here because these methods are not effective for varying density clusters. In recent decade there were few works on Cognitive analysis by Deep Learning-based architectures. In [55] A deep learning-based model is suggested for the prediction of cognitive processes and knowledge dimensions. The fundamental goal of this paper is to create a relative study of the classification of the summative assessment based on Revised Bloom's taxonomy using the Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) of Deep Learning techniques, to attain significant accomplishment and elevated precision levels. In another work [56] transfer learning via bidirectional encoder representations from Transformers (BERT) was adopted for automatic and large-scale classification of

MOOCs based on their learning objectives and Bloom's taxonomy. The proposed method is not comparable with deep learning-based methods as the number of question papers taken as input is not high and also the number of parameters considered is low, rather only one, that is, the difficulty level of the question. Henceforth, multi-dimensional clustering techniques are not in the scope of this study, rather these techniques will increase the computational complexity without any improvement in performance.

In this study questions are clustered into different difficulty groups based on their calculated hardness.

This is described in detail in the next section.

5. AUTOMATIC CLUSTERING OF QUESTIONS

A dynamic procedure for the automatic clustering of the questions is proposed in this study. Bloom's taxonomy and the JNBO technique are applied here to classify a set of questions having different difficulty levels into different groups. The procedure is described as follows:

First, keywords related to each Bloom's level are identified and stored in an array named **Bloom's** Keyword Array (BKA []). Thereafter, all the individual levels of Bloom's taxonomy are provided with distinct weights according to their chronological difficulty levels in the range of 1–6. Since the difficulty levels in Bloom's taxonomy increases from the lower to the upper level, hence remembering level is assigned a weight of 1; the understanding level is assigned a weight 2 and so on. Each question is analyzed to obtain the number of matches with the keywords corresponding to each level. The questions are ranked by the specific parameters - marks, number of matches at each Bloom's level and the individual Bloom's level weight. Marks for a specific question are considered here because marks can be a criterion to calculate the hardness of the questions. A particular question carrying more marks is considered to be tougher than the same question carrying a lower mark because in the earlier case, a greater level of knowledge was required to answer. A particular question may fall under multiple levels of Bloom. The weightage scores for individual levels are stored in a question difficulty level (QDL) matrix. The total hardness of a question spanning across all the levels is stored in an array arrHardness[]. After that algorithm 2 is called where the JNBO technique is applied to the weighted question set to get k clusters of the intended difficulty level. k is dynamic here and decided by the user to select the number of difficulty levels. The algorithm for automatic question classification is mentioned in Algorithm 1.

Algorithm 1: AutoQuestionClustering(BKA[],Q)

Input: Number of question segments or clusters K, the total questions paper pool (Q), BKA [] storing keywords related to each Bloom's Taxonomy level. Keywords of level i are stored in BKA[i], where 0 < = i < = 5.

Output: K number of questions clusters with varying difficulty levels.

Begin
1. Define the array d[] = {1,2,3,4,5,6} where each element represents the hardness level in Bloom's taxonomy from understanding to creating the level.
/∗Store the weight of a question in a Question Difficulty Level (QDL) matrix. Each row of the //matrix represents an individual question q and each column represents each Bloom's taxonomy level.∗/
2. Declare QDL[Q][6] where Q is the total number of questions
3. For each question q in the set of questions Q
4. For each row j in BKA
5. matchedCount = 0
6. For each word m in q
7. For each keyword r in BKA[j]
8. If m matches with r are contained in BKA[j]
9. matchedCount=matchedCount+1
10. break;
11. End If
12. End For
13. End For
14. $$QDL[q][j] = \text{matchedCount} * \text{mark} * \frac{d[j]}{\sum_{k=0}^{5} d[k]} \quad (1)$$
15. End For
16. End For
17. Declare arrHardness[N] where N is the total number of questions
18. For each row i in the QDL matrix
19. $HF_q = 0$
20. For j in the range (0,5):
21. $$HF_q = HF_q + QDL[i][j] \quad (2)$$
22. End For
23. Put HF_q inside **arrHardness**[]
24. End For
25. Initialize the no of clusters K

26. call **JNBO(arrHardness,K)** //Algorithm 2
End

Algorithm 2: JNBO (arrHardness[N],K)
Input: Array having hardness factors (HF) of all N number of questions and number of classes K
Output: Clusters of questions of different difficulty levels
Begin
1. Sort the array arrHardness[] in an ascending order of the hardness or weight.
2. For each question difficulty in the arrHardness[] compute the sum of squared deviation (SSD) from the array mean(M).

$$\mathrm{SSD} = \sum_{i=0}^{N} (M - \mathrm{arrHardness}\ [i])^2 \tag{3}$$

3. All possible C combinations of K pair classes or hardness arrays are formed by combining different HF values
/∗SSDs based on class means (SSDCM) are calculated for each class (C) having n number of questions among the K number of classes∗/
4. Declare SSDCM[K]
For each k pair class among all combinations do
a. For each class C in k do
(i) Compute mean (m)

$$\text{(ii) } \mathrm{SSD}_C = \sum_{i=0}^{n} (m - \mathrm{C}[i])^2 \tag{4}$$

$$\text{(iii) } \mathrm{SSDCM}[C] = \mathrm{SSDCM} + \mathrm{SSD}_c \tag{5}$$

/∗ Goodness of Variance fit (GVF) is calculated for each of the K pair classes

$$\mathrm{GVF} = \frac{\text{SSD-SSDCM}}{\mathrm{SSD}} \text{where} 1 \leq \mathrm{GVF} \leq 0 \tag{6}$$

GVF = 1 means a perfect fit and GVF = 0 means poor fit.∗/
5. Declare GVF[K]

6. For each k pair of classes in all K combinations do

$$\mathrm{GVF}[k] = \frac{\mathrm{SSD} - \mathrm{SSDCM}[k]}{\mathrm{SSD}} \tag{7}$$

7. Return a combination of classes or groups of questions having the greatest GVF value.

End
After obtaining the question clusters according to their difficulty levels in the above-mentioned procedure the next step is to rank the candidates which is explained in the next section.

6. CANDIDATE RANKING SYSTEM

In this study an online competitive examination scheme is proposed where questions are presented one by one according to the chronological difficulty levels. The questions are introduced from the easiest level and gradually the difficulty level of the questions is increased based on the candidate's performance. If the candidate is successful to answer the questions from a particular difficulty level then only he/she can face the question belonging to the next higher difficulty level. If the candidate is unable to answer a question then the next question will be selected from the group, one level below the hierarchy. Once the examination is completed, a ranking methodology is applied which considers both the question marks and their corresponding difficulty level so that the cognitive skill of the student is properly evaluated. The algorithm for candidate ranking is given in Algorithm 3.

Algorithm 3: CandidateRanking (questionClusters, w, succAttmt[])
Input: K clusters or subgroups of questions from Algorithm AutoQuestionClustering, weight or total hardness (w) of each question, succAttmt[] containing successfully answered question numbers.
Output: Ranks of the candidates in a descending order

Begin

1. For each candidate appearing in the online competitive examination
For each question subgroup l in k subgroups
Declare Rank = 0
For each question q among all the questions Q in l do
i. Calculate mean hardness $h_l = \frac{\sum_{q=1}^{Q} w_q}{Q}$ where w_q is the total hardness of question q
End For
2. For each correctly answered question i do

$$\mathrm{Rank} = \mathrm{Rank} + \mathrm{Marks}_i * h_l \tag{8}$$

3. End For
4. End For
5. End For
6. Sort candidates according to the descending order of their rankings.

End

The block diagram of the overall candidate ranking system is given in Figure 1.

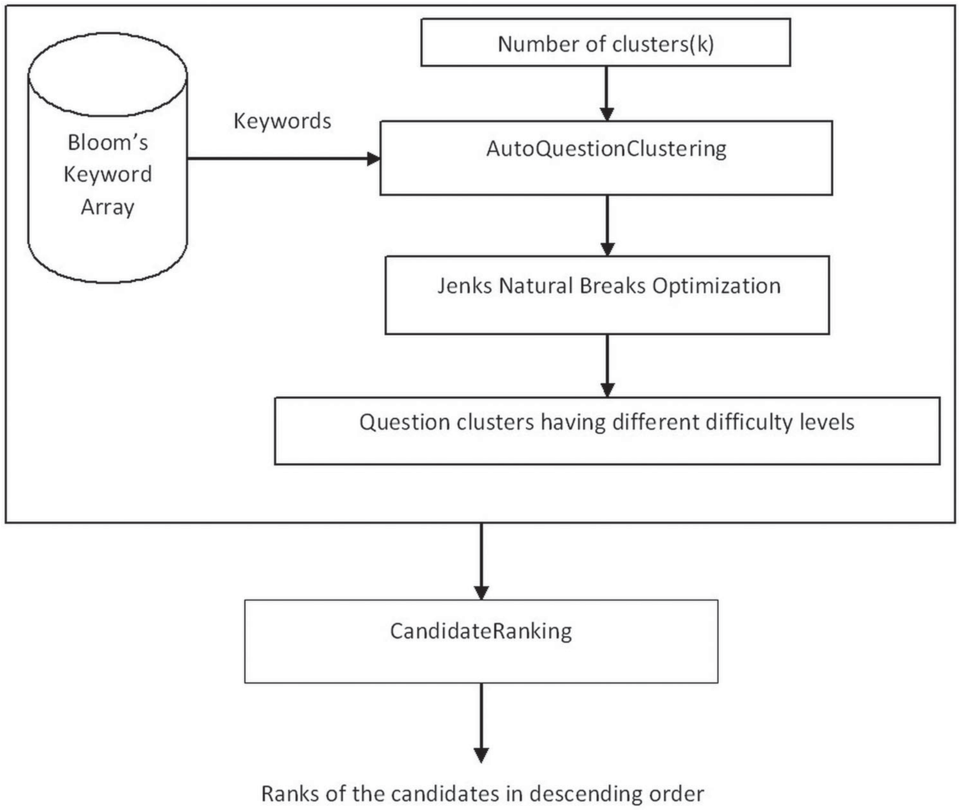

Figure 1: Block diagram of the candidate ranking system

7. RESULTS AND DISCUSSION

BCL's (Bloom's taxonomy Cognitive Level's) dataset is used here and the proposed methodology is applied to a dataset of nearly 600 questions introduced in [57,58]. This dataset contains questions from different domains since this is collected from several web sources. The questions in the data set were manually categorized into six learning levels of Bloom's. This dataset was used for automating the classification of questions according to Bloom's cognitive levels. Academic accreditation and regulatory organizations (e.g. NBA, NAAC, AICTE in India) throughout the world are giving significant importance to Outcome-Based Education (OBE) to improve the teaching and learning process. Outcome-based learning that will nurture the cognitive competency of the students is of foremost importance in the current academic periphery. Bloom's taxonomy is widely applied in the educational domains for outcome-based course design and especially it is used for assessment structuring for understanding the cognitive level achievements of the students. In this study Bloom's taxonomy is followed to develop the proposed model of cognitive ranking of the students. All the questions should have at least one Bloom's predefined keyword to a particular Bloom's cognitive level. Hence the questions having Bloom's predefined keywords are only considered here. The proposed method is applied to a subset of this dataset and has a combination of questions from different levels of Bloom's taxonomy. **Spyder 3.3.3** tool [59] from the Anaconda distribution is used here as a Python development environment. (Python version 3.7.3). Python's jenkspy library is used here to apply the JNBO procedure to the data distribution. The JNBO partitioning method is considered as the one-dimensional k-Means. Well-defined clusters of questions having different hardness levels are obtained through this method. In this work, the density-based clustering method is also analyzed along with partitioning-based clustering (JNBO) on the question paper database. The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method is applied here to the questions having varying difficulty levels. The resultant clusters of questions are shown in Figure 2. A significant presence of clusters having only one question is observed that

Indicates the DBSCAN method is not able to form well-defined clusters here. One probable reason behind this is that one of the major disadvantages of the DBSCAN method is, it fails in case of varying density clusters [60].

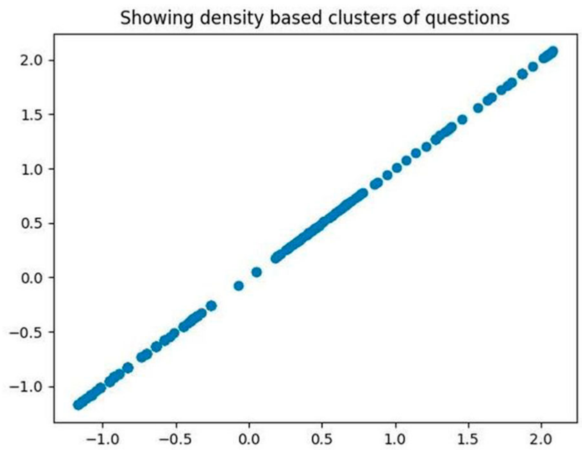

Figure 2: Different question clusters having varying difficulty levels

Table 1: Weightage given to Bloom's taxonomy levels

Blooms Levels	Keywords
0	cite, define, describe, identify … …
1	abstract, arrange, articulate, associate, …
2	apply, calculate, carry out, classify, … ..
3	analyze, arrange, breakdown … … …
4	appraise, apprise, argue, assess, … … ..
5	arrange, assemble, build, collect, … … .

In this study the JNBO method is used to partition the questions into different difficulty groups. First, weightage is given to the individual Bloom's taxonomy level, as shown in Table 1. Thereafter numbers of matched keywords are calculated, comparing the values in BKA (Table 2) and weights are assigned to the individual question by using Equation 1. A sample of the outcome is given in Table 3. After assigning weights to individual questions the Jenks Natural Breaks Optimization method is applied to find the breakpoints that segment the questions in different difficulty groups. Nine different groups of questions having increasing difficulty levels are created. The breakpoints obtained by this method are 0.048, 0.29, 0.57, 1.05, 1.71, 2.56, 3.26, 4.03, 4.55 and 4.98 respectively. So the ranges are [0.048 … … 0.29], [0.29 … … .0.57], [0.57 … … 1.05], [1.05 … .1.71], [1.71 … .2.56], [2.56 … .3.26], [3.26 … .. 4.03], [4.03 … .4.55] and [4.55 … 4.98], respectively. This is also shown in Figure 2. During the online examination candidates are given questions ranging from the easiest groups to the most difficult group. If a candidate can successfully answer the questions from a particular group then he/she is presented with the question from the next higher group in the ladder.

(B) **Case Study-Dynamic candidate ranking:**
The proposed methodology of the dynamic ranking of

Table 2: Bloom's Keywords Array (BKA)

Bloom's Taxonomy Level	Given Weight
Remember	1
Understand	2
Apply	3
Analyze	4
Evaluate	5
Create	6

Table 4: Showing rank calculation of candidate 1

Question No	Mean Weight of the group	Marks	Answer status	Weight ∗ Marks
1	0.12	3	Success	0.36
2	0.38	3	Success	1.14
3	0.81	5	Failure	-
4	0.38	4	Success	1.52
5	1.33	5	Success	6.65
Rank				13.72

the candidates in competitive examinations is demonstrated in Table 4. Whenever a candidate is answering a question successfully then first the corresponding question cluster is found by comparing its weight with the breakpoints. Then the marks of the question are multiplied with the mean weight or difficulty level of the group. Whenever a question is successfully answered the next question will come from the next higher difficulty question segment. As an example, after successfully answering question no 2 the next question the candidate facing, is question no 3 having higher difficulty level. If the candidate fails to answer a question from a certain difficulty level then the next question will come from a lower difficulty level. In Table 4 it is observed that the candidate fails to answer question no 3 and hence the next question he is facing is question no. 4 belonging to the lower difficulty level. The marks of all successfully answered questions and corresponding mean segment weight are multiplied to calculate the score of the candidate for answering a particular question. Thereafter, all these scores are added to obtain the final rank of the candidate. In this method both the marks of the question and the corresponding difficulty level are considered to evaluate the candidate.

In this study the JNBO procedure is applied to obtain nine tight clusters of questions having increasing difficulty levels.

In Figure 3 each group is identified by different colours. It is observed that the entire question sets are divided into three major divisions: easy, medium and hard questions and each section is further divided into three groups having increasing difficulty levels. By dividing the entire question pool into nine different question groups, flexibility is provided to choose questions considering the difficulty levels. The system gets enough scope to evaluate the cognitive skill set of the student by evaluating his/her

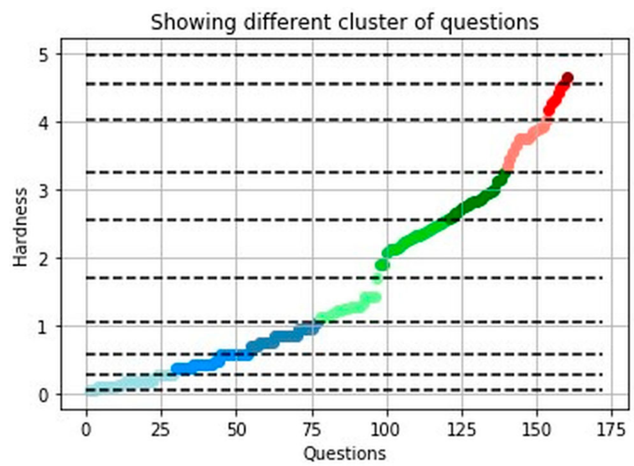

Figure 3: Different breakpoints indicating the groups of questions. Break points are shown by the dotted lines(Breakpoints: 0.048, 0.29, 0.57, 1.05, 1.71, 2.56, 3.26, 4.03,4.55 and 4.98)

efficiency in each level. The ranking method introduced in this study is provided with the capability to judge the students' performance based on these cognitive levels. One candidate can enter a particular difficulty level by completing the earlier level. Both marks of a particular question and the mean weight of the group it belongs to are considered to find the ranking of a candidate. Hence a candidate who has solved a greater number of harder questions is considered as possessing higher cognitive skills and the same is reflected in the acquired rank.

In this work the questions are identified as a member of the particular level of Bloom's taxonomy based on matching Bloom's keyword at each level. There are some questions that fall under particular Bloom's level semantically but Bloom's keyword is missing in them Se.g. "find the time complexity of the Merge sort algorithm". Semantically these questions falls under analysis level in Bloom's taxonomy though there is no Bloom's keyword. Moreover, sometimes hardness of the Moreover sometimes hardness of the questions falling in the same level of Bloom's differ. As an example, the hardness of the following two questions is different- (i) "find out the time complexity of Bubble Sort" and (ii) "find out the time complexity of merge sort". Both of these belong to analysis level but the second question is analytically more complex than the first one. Classifying and computing the hardness of these kinds of questions needs further analysis and research.

8. CONCLUSION

The proposed framework generates groups of questions automatically based on intended learning outcomes which act as a repository of questions indicating the corresponding hardness level. This creates an opportunity to conduct online examinations that will properly evaluate students' cognitive skills. The traditional evaluation system only considers the final marks obtained and sometimes the accuracy of the student. The proposed methodology considers the marks of a question and the mean hardness level of the cluster to which it belongs, to reflect the cognitive skills of students in the final result or rank list. In competitive examinations the suggested methodology will assist to find out the candidate with a higher cognitive level with the ability to solve complex analytical problems. This framework will help educational institutes to identify the students' weaknesses in different cognitive skills so that focused remedial measures can be taken to improve the individual academic performance. Moreover, in this COVID-19 pandemic situation schools and colleges are mostly continuing the educational activities by conducting online classes but

Table 3: Snapshot of the outcome after question weight assignment according to Blooms taxonomy

Questions	Remembering	Understand	Apply	Analyze	Evaluate	Create	Total Weightage
By comparing the map of the tectonic plates to the earthquake map, what inferences can you make?				1			0.761904762
Compare Calliope with Howie. Use the word bank.		1		1			0.857142857
Compare two of the characters in this book.				1	1		0.857142857
State and explain BFS algorithm.	1	1					0.857142857
State and explain DFS algorithm.	1	1					0.857142857
Create an equation to represent the solution to this problem.						1	0.857142857
Create plan of local environment by applying drawing around boxes.			1			1	0.857142857
Design a building according to given specifications						1	0.857142857
design a cost effective strategy to generate reliable data.						1	0.857142857
Appraise the speech's effectiveness based upon the class' criteria.					1		0.952380952
Critique an experimental design or a research proposal.					1		0.952380952

the physical classroom is needed especially for the academically less-proficient students. This work helps the academics to properly track the academic progress of the students so that adequate measures are taken to overcome the knowledge gap. Moreover the corporate can choose the suitable employees as per their job role by adopting the proposed framework. Even based on the ability to solve the difficult problems the organization can fix the salary of the selected candidates.

As discussed earlier some questions devoid of Bloom's defined keywords. Further study is required to analyze the semantics of these kind of questions to assign it to a particular Bloom's level. There are some questions that fall under the same level of Bloom's but their hardness can be different. Further investigation and analysis is required to accurately compute the hardness of these questions by understanding their semantics.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

REFERENCES

1. D. McArthur, M. Lewis, and M. Bishary, "The roles of artificial intelligence in education: current progress and future prospects," *J. Educ. Technol*, Vol. 1, no. 4, pp. 42–80. https://www.learntechlib.org/p/161310/.
2. S. Yang, and H. Bai, "The integration design of artificial intelligence and normal students' education," *J. Phys. Conf. Seri.*, Vol. 1453, pp. 012090, Jan. 2020. DOI: 10.1088/1742-6596/1453/1/01209.
3. M. Ciolacu, A. F. Tehrani, R. Beer, and H. Popp, "Education 4.0—Fostering student's performance with machine learning methods," in 2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME), Oct. 2017, pp. 438–43. DOI: 10.1109/SIITME.2017.8259941
4. A. Pardo, F. Han, and R. A. Ellis, "Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance," *IEEE Trans. Learn. Technol.*, Vol. 10, no. 1, pp. 82–92, 2016. DOI: 10.1109/TLT.2016.2639508
5. J. M. Azevedo, E. P. Oliveira, and P. D. Beites, "How Do mathematics teachers in higher education look at E-assessment with multiple-choice questions," in *CSEDU (2)*, 2017, pp. 137–45. DOI: 10.5220/0006324801370145
6. B. S. Bloom, "Taxonomy of educational objectives: The classification of educational goals," in *Cognitive Domain* the classification of educational goals by a committee of college and university examiners. Handbook I: Cognitive Domain. New York, NY: Longmans, Green, 1956.
7. D. R. Krathwohl, "A revision of Bloom's taxonomy: An overview," *Theory. Pract.*, Vol. 41, no. 4, pp. 212–8. DOI: 10.1207/s15430421tip4104_2.
8. L. W. Anderson, and B. S. Bloom, "*A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*," *Longman*, Vol. 83, no. 3, 2001.
9. U. R. Hodeghatta, and U. Nayak. *Business analytics using R-a practical approach*. Apress, 2016.
10. R. M. Crespo, et al., "Aligning assessment with learning outcomes in outcome-based education," in IEEE EDUCON 2010 Conference, Apr. 2010, pp. 1239–46. DOI: 10.1109/EDUCON.2010.5492385
11. J. Pérez, C. Vizcarro, J. García, A. Bermúdez, and R. Cobos, "Development of procedures to assess problem-solving competence in computing engineering," *IEEE Trans. Educ.*, Vol. 60, no. 1, pp. 22–28, 2016.
12. F. Camelia, T. L. Ferris, and M. B. Behrend, "The effectiveness of a systems engineering course in developing systems thinking," *IEEE Trans. Educ.*, Vol. 63, no. 1, pp. 10–16, 2019. DOI: 10.1109/TE.2019.2926054.
13. P. J. Martínez, F. J. Aguilar, and M. Ortiz, "Transitioning from face-to-face to blended and full online learning engineering master's program," *IEEE Trans. Educ.*, Vol. 63, no. 1, pp. 2–9, 2019. DOI: 10.1109/TE.2019.2925320.
14. COVID-19 Impact on Education. https://en.unesco.org/covid19/educationresponse (Last access date: 08/08/2020).
15. Exams and assessments in COVID-19 crisis: fairness at the centre. https://en.unesco.org/news/exams-and-assessments-covid-19-crisis-fairness-centre (last access date: 23/07/2020).
16. J. Chen, S. Yang, H. Li, B. Zhang, and J. Lv, "Research on geographical environment unit division based on the method of natural breaks (Jenks)," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, Vol. 3, pp. 47–50. DOI: 10.5194/isprsarchives-XL-4-W3-47-2013.
17. S. Liu, and M. d'Aquin, "Unsupervised learning for understanding student achievement in a distancelearningsetting," in 2017 IEEE Global Engineering Education Conference (EDUCON), Apr. 2017, pp. 1373–7. DOI: 10.1109/EDUCON.2017.7943026
18. Y. Chen, "Association analysis of online learning behavior in interactive education based on an intelligentconcept machine," *Int. J. Contin. Eng. Educ. Life Long Learn.*, Vol. 30, no. 2, pp. 161–75, 2020.DOI: 10.1504/IJCEELL.2020.106342
19. T. Widiyaningtyas, M. I. W. Prabowo, and M. A. M. Pratama, "Implementation of K-means clustering method to distribution of high school teachers," in 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), IEEE, Sep. 2017, pp. 1–6. DOI: 10.1109/EECSI.2017.8239083

20. J. Qin, Z. Jia, and P. Ma, "Analysing learning behaviours of advanced mathematics in MOOCs," *Int. J. Contin. Eng. Educ. Life Long Learn.*, Vol. 29, no. 1-2, pp. 113–28, 2019. DOI: 10.1504/IJCEELL.2019.099251.

21. P. De Morais, M. Alana, J. M. Araujo, and E. B. Costa, "Monitoring student performance using data clustering and predictive modeling," in 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, IEEE, Oct. 2014, pp. 1–8. DOI: 10.1109/FIE.2014.7044401

22. M. Shovon, H. Islam, and M. Haque, "An approach of improving students academic performance by using k means clustering algorithm and decision tree," *Int. J. Adv. Comp. Sci. Appl. (IJACSA)*, Vol. 3, no. 8, pp. 146–9, 2012. DOI: arXiv:1211.6340

23. P. D. Antonenko, S. Toy, and D. S. Niederhauser, "Using cluster analysis for data mining in educational technology research," *Educ. Technol. Res. Dev.*, Vol. 60, no. 3, pp. 383–98, 2012. DOI: 10.1007/s11423-012-9235-8.

24. R. Campagni, D. Merlini, and M. C. Verri, "Finding regularities in courses evaluation with K-means clustering," in CSEDU (2), 2, Apr. 2014, pp. 26–33. DOI: 10.5220/0004796000260033

25. A. Ezen-Can, K. E. Boyer, S. Kellogg, and S. Booth, "Unsupervised modeling for understanding MOOC discussion forums: a learning analytics approach," in Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, Mar. 2015, pp. 146–50. DOI: 10.1145/2723576.2723589.

26. A. Mujkanovic, D. Lowe, K. Willey, and C. Guetl, "Unsupervised learning algorithm for adaptive group formation: Collaborative learning support in remotely accessible laboratories," in International Conference on Information Society (i-Society 2012), IEEE, Jun. 2012, pp. 50–7.

27. N. Zhang, G. Biswas, and Y. Dong, "Characterizing students' learning behaviors using unsupervised learning methods," in International Conference on Artificial Intelligence in Education, Cham, Springer, Jun. 2017, pp. 430–41. DOI: 10.1007/978-3-319-61425-0_36

28. M. A. North, "A method for implementing a statistically significant number of data classes in the Jenks algorithm," in 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, vol.1, Aug. 2009, pp. 35–8. DOI: 10.1109/FSKD.2009.319

29. N. Khamis, T. C. Sin, and G. C. Hock, "Segmentation of residential customer load profile in peninsular Malaysia using Jenks Natural Breaks," in 2018 IEEE 7th International Conference on Power and Energy (PECon), IEEE, Dec. 2018, pp. 128–31. DOI: 10.1109/PECON.2018.8684113

30. M. do Carvalhal, R. Lourenço, V. Pereira, and H. G. Costa, "A multicriteria approach to the human development index classification," *Soc. Indic. Res.*, Vol. 136, no. 2, pp. 417–38, 2018. DOI: 10.1007/s11205-017-1556-x.

31. A. Manzoor, H. Aziz, M. Jahanzaib, A. Wasim, and S. Hussain, "Transformational model for engineering education from content-based to outcome-based education," *Int. J. Contin. Eng. Educ. Life Long Learn.*, Vol. 27, no. 4, pp. 266–86, 2017. DOI: 10.1504/IJCEELL.2017.087136.

32. J. Rutkowski, K. Moscinska, P. Jantos, and J. Rutkowski, "Application of Bloom's taxonomy for increasing teaching efficiency–case study," in Proc. of ICEE 2010, Jul. 2010.

33. S. F. Kusuma, R. Z. Alhamri, D. O. Siahaan, C. Fatichah, and M. F. Naufal, "Indonesian question generation based on Bloom's taxonomy using text analysis," in International Seminar on Intelligent Technology and Its Applications (ISITIA), IEEE, 2018, pp. 269–74. DOI: 10.1109/ISITIA.2018.8711015

34. A. J. Swart, "Evaluation of final examination papers in engineering: A case study using Bloom's taxonomy," *IEEE Trans. Educ., IEEE*, Vol. 53, no. 2, pp. 257–64, 2009. DOI: 10.1109/TE.2009.2014221.

35. E. Thompson, A. Luxton-Reilly, J. L. Whalley, M. Hu, and P. Robbins, "Bloom's taxonomy for CS assessment," in Proceedings of the Tenth Conference on Australasian Computing Education, vol. 78, Jan. 2008, pp. 155–61. DOI: 10.5555/1379249.1379265

36. J. L. Whalley, R. Lister, E. Thompson, T. Clear, P. Robbins, P. K. Kumar, and C. Prasad, "An Australasian study of Reading and comprehension skills in novice programmers, using the Bloom and SOLO taxonomies," in Proceedings of the 8th Australasian Conference on Computing Education, Australian Computer Society, Inc, vol. 52, pp. 243–52. DOI: 10.5555/1151869.1151901

37. D. R. Ch, and S. K. Saha, "Automatic multiple choice question generation from text: A survey," *IEEE Trans. Learn. Technol.*, Vol. 13, no. 1, Dec. 2018. DOI: 10.1109/TLT.2018.2889100.

38. K. Jayakodi, M. Bandara, and I. Perera, "An automatic classifier for exam questions in engineering: A process for Bloom's taxonomy," in 2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), IEEE, Dec. 2015, pp. 195–202. DOI: 10.1109/TALE.2015.7386043

39. K. Jayakodi, M. Bandara, and D. Meedeniya, "An automatic classifier for exam questions with WordNet and Cosine similarity," in 2016 Moratuwa Engineering Research Conference (MERCon), IEEE, Apr. 2016, pp. 12–17. DOI: 10.1109/MERCon.2016.7480108

40. A. Rahim, T. N. Tengku, Z. Abd Aziz, A. Rauf, R. Hafsah, and N. Shamsudin, "Automated exam question generator using genetic algorithm," in 2017 IEEE Conference on e-Learning, e-Management and e-Services (IC3e), IEEE, Nov. 2017, pp. 12–17. DOI: 10.1109/IC3e.2017.8409231

41. M. Mohammed, and N. Omar, "Question classification based on Bloom's taxonomy cognitive domain using

modified TF-IDF and word2vec," *PloS one*, Vol. 15, no. 3, 2020. DOI: 10.1371/journal.pone.0230442.

42. A. Amria, A. Ewais, and R. Hodrob, "A framework for automatic exam generation based on intended learning outcomes," in *CSEDU (1)*, 2, 2018, pp. 474–80. DOI: 10.5220/0006795104740480

43. A. A. Yahya, Z. Toukal, and A. Osman, "Bloom's taxonomy–based classification for item bank questions using support vector machines, in *Modern Advances in Intelligent Systems and Tools*, Berlin, Heidelberg: Springer, 2012, pp. 135–40. DOI: 10.1007/978-3-642-30732-4_17.

44. D. Naglot, S. Gaikwad, P. Gaikwad, A. Salvi, and S. Mutyal, "Keyword based shuffling algorithm for question paper generator," *Int. J. Comput. Appl.*, Vol. 179, no. 37, pp. 36–40, 2018. DOI: 10.5120/ijca2018916866

45. G. F. Jenks. Optimal data classification for choropleth maps. *Department of Geographiy, University of Kansas Occasional Paper*, 1977.

46. W. D. Fisher, "On grouping for maximum homogeneity," *J. Am. Stat. Assoc.*, Vol. 53, no. 284, pp. 789–98, 1958. DOI: 10.1080/01621459.1958.10501479.

47. Arcgis, Data classification methods. https://pro.arcgis.com/en/pro-app/help/mapping/layer-properties/data-classification-methods.htm (Last Access date: 23/07/2020).

48. C. A. Brewer, and L. Pickle, "Evaluation of methods for classifying epidemiological data on choropleth maps in series," *Ann. Asso. Am. Geogr.*, Vol. 2, no. 4, pp. 662–81, 2002. DOI: 10.1111/1467-8306.00310

49. M. Fairweather, and K. G. Fairweather. Choropleth Mapping: The Problems of Classification and Data Presentation, ERIC, 1984. Available: https://eric.ed.gov/?id=ED242628

50. ESRI, What is the Jenks optimization method. 2012. Available: https://support.esri.com/en/technical-article/000006743 (last access date: 23/07/2020).

51. R. McMaster, "In Memoriam: George F. Jenks (1916-1996)," *Cartogr. Geogr. Inf. Sys.*, Vol. 24, no. 1, pp. 56–9, 1997. DOI: 10.1559/152304097782438764.

52. S. Dong, D. Zhou, W. Ding, and J. Gong, "Flow cluster algorithm based on improved K-means method," *IETE. J. Res.*, Vol. 59, no. 4, pp. 326–33, 2013.

53. E. Maadi, A. & Djouadi, and M. S, "Using a light dbscan algorithm for visual surveillance of crowded traffic scenes," *IETE. J. Res.*, Vol. 61, no. 3, pp. 308–20, 2015.

54. K. Samunnisa, G. S. V. Kumar, and K. Madhavi, "Intrusion detection system in distributed cloud computing: Hybrid clustering and classification methods," *Meas. Sens.*, Vol. 25, pp. 100612, 2023.

55. M. D. Laddha, V. T. Lokare, A. W. Kiwelekar, and L. D. Netak, "Classifications of the summative assessment for revised Blooms taxonomy by using deep learning," *Int. J. Eng. Trends Technol.*, Vol. 69, no. 3, pp. 211–88, 2021.

56. H. Sebbaq, and N. E. El Faddouli, "Fine-tuned BERT model for large scale and cognitive classification of MOOCs," *Int. Rev. Res. Open Dis. Learn.*, Vol. 23, no. 2, pp. 170–90, 2022.

57. A. Yahya. Bloom's Taxonomy Cognitive Levels Data Set, 2011. DOI: 10.13140/RG.2.1.4932.3123

58. A. A. Yahya, and A. Osman. Automatic classification of questions into Bloom's cognitive levels using support vector machines, 2011.

59. K. S. Ajithabh, and P. K. Patro, "SigMT: An open-source Python package for magnetotelluric data processing," *Comput. Geosci.*, Vol. 171, pp. 105270, 2023.

60. A. Moreira, M. Y. Santos, and S. Carneiro. Density-based clustering algorithms–DBSCAN and SNN. University of Minho-Portugal, 1, 18, 2005.

AUTHORS

Anjan Dutta completed his bachelor of technology in information technology from RCC Institute of Information Technology, Kolkata, India and master of technology in information technology from Calcutta University, Kolkata, India. Currently, he is working as a lecturer in computer science and technology, Raja Ranjit Kishore Govt. Polytechnic, Jhargram, India. He is also pursuing PhD at Jadavpur University, Kolkata, India. His research area includes data mining, big data and machine learning.

Corresponding author. Email: anjan.dutta.edu@gmail.com

Punyasha Chatterjee received the B Tech, M Tech and PhD degrees in information technology from the University of Calcutta, Kolkata, India, in 2003, 2005 and 2018, respectively. She has been an assistant professor in the School of Mobile Computing and Communication, Jadavpur University, Kolkata, India, since 2012. Her area of interest includes adhoc networks, wireless sensor networks, Internet of Things and pervasive computing. She is a senior member of IEEE and a member of ACM.

Email: punyasha.chatterjee@gmail.com

Nilanjan Day was born in Kolkata, India, in 1984. He received his B Tech degree in information technology from West Bengal University of Technology in 2005, an M Tech in information technology in 2011 from the same University and a PhD in digital image processing in 2015 from Jadavpur University, India. In 2011, he was appointed as an assistant professor in the Department of Information Technology at JIS College of Engineering, Kalyani, India followed by Bengal College of Engineering College, Durgapur, India in 2014. He is now employed as an associate professor in the department of Computer Science and Engineering, Techno International New Town, Kolkata, India. His research topic is signal processing, machine learning and information security.

Email: neelanjan.dey@gmail.com

Pablo Moreno-Ger was born in Madrid, Spain, in 1981. He received a degree in computer engineering from Universidad Complutense de Madrid, in 2004, and a PhD degree in computer science from the Department of Software Engineering and Artificial Intelligence, Universidad Complutense de Madrid, in 2007. He is currently the director of the School of Engineering and Technology, UNIR. He was with the Universidad Complutense de Madrid, where he was a member of the e-UCM Research Group. He was also the Vice Dean for Innovation with the School of Computer Engineering, Universidad Complutense de Madrid. He has a strong research record in technology enhanced learning, AI, machine learning and learning analytics. He has published more than 150 academic articles in these topics. He holds the IBM Chair on Data Science in Education at UNIR.

Email: pablo.moreno@unir.net

Soumya Sen received a PhD in computer science & engineering from the University of Calcutta in 2016. He received his M Tech in computer science & engineering in 2007 and MSc in computer and information science in 2005 also from University of Calcutta. He joined the A K Choudhury School of Information Technology at the University of Calcutta in 2009. Dr. Sen has around 80 research publications in International Journals and conferences. He has 3 international patents and also published 2 books. Dr. Sen is a TPC member of many conferences across the world and serves as a reviewer for many International journals. He is a member of IEEE and ACM. His current research area interests are data warehouse & OLAP tools, data mining and distributed database.

Email: iamsoumyasen@gmail.com