Smart Automated Grading System Using Machine Learning Algorithm for Short Answers Questions

Doaa I. Hasan*[1], Alaa Fareed Abdulateef[2], Yasser F.Hassan[4], and Wagdy Gomaa[3]
[1,3]The Faculty of Science's Department of Mathematics and Computer Science, Alexandria University
[2]Universiti Utara Malaysia UUM, School of Computing
[4]Faculty of Computing and Data Science, Alexandria University

Abstract:
An automated grading system was developed in this work to help teachers quickly grade multiple assignments at the same time, with fewer efforts, and especially with the Pandemic of COVID-19 and the challenges facing e-learning, in terms of correcting essay-answer tests. The use of automatic grading systems saves time and effort for educational institutions in achieving high-efficiency educational goals. considering that the number of students participating in online learning has increased dramatically, it is expected to reach more than 1.38 billion by 2023. Thus, it has become necessary to prepare a system for analysing and classifying students' responses to various types of electronic exams automatically. The k-Means method has been employed in many research fields. However, it has flaws due to the large number of parameters that affect its performance and the difficulty in managing the results. For that reason, the OPTICS clustering algorithm was employed in this work, as well as the African Buffalo Improvement technique, to get optimal results. The OPTICS algorithm clearly outperformed the K-Means algorithm, according to the results. The findings were greatly improved by using OPTICS and incorporating the African buffalo optimization technique.

Keywords: Automatic grading, machine learning, OPTICS algorithm, African Buffalo Optimization, clustering.
DOI: 10.24297/j.cims.2023.1.2

1. Introduction

An automated grading system has become one of the essential tools in every educational institute, especially with the COVID-19 pandemic[1]. It is the system's capacity to "learn" the correct and wrong answers and store them for future use [2]. The development of partially or fully automated grading systems utilizing computational approaches has become a key study subject, as technology has quickly gotten more powerful in scoring examinations and essays, particularly since the 1990s. By grading the tests, the elements of success can be detected, and students' strengths and weaknesses can be detected to improve and develop the educational level, as well as the effectiveness of the teaching strategies used and the educational resources used in the education process[2]. Besides, holding exams and grading answers to put marks on them and monitoring students' results is done manually, as this method has become incompatible and ineffective with the e-learning system[3].

Moreover, when the coronavirus outbreak occurred, it threw a shadow over the education sector since it forced educational institutions to lock their doors in order to decrease the likelihood of its spread. This generated widespread anxiety among people associated with the education sector.[4][5].

All of this has prompted educational institutions to consider converting to electronic education, which has been debated for some time [6] as an alternative. For a long time, there has been controversy regarding whether or not it should be integrated into the educational process. This was particularly true once the educational process was immediately impacted by the industry's automation and technological growth[7]. Those mentioned above led to recent years' education development and have become reliant on information and communication technology. This is one of the general pillars on which the modern e-learning concept is built, and it is important to note that the use of the Internet in educational science did not begin in the year 2000 but rather occurred prior to that year, as most universities today make use of so-called education management systems[8].

The system provided the faculty member with the opportunity to hold the test on the website, which facilitates the teacher in grading the test answers and monitoring students' results automatically [8]. The research, itis talking about combining machine efficiency and human experience to give the best way to evaluate students' answers. Since the outbreak of the COVID-19 virus, there has been significant growth in the number of students engaging in online learning, which is expected to reach more than 1.38 billion users by 2023.Figure 1 shows the fast growth of online learning users since the beginning of 2020.It has become necessary to establish a system for evaluating and grading students' answers to online tests [3][4].

The K-Means algorithm has been commonly used by many researchers for grading systems. However, this algorithm suffers from These are the limitations which is a **research gap**[11]:
• Predicting the number of clusters (K-Value) is difficult. The initial seeds have a significant influence on the outcome.
• The order in which the data is entered has an influence on the findings.

These limitations make it not a good choice to use as it has affected the accuracy of the results.
To overcome the limitations of previous studies, this study proposes using the Ordering Points To Identify the Clustering Structure (OPTICS) method to identify the clustering structure [12]. An optimization technique is required to determine the optimal value for the ε (Epsilon) variable to obtain the best possible outcomes. One of the promising optimization algorithms that can be used to select the optimal value for the OPTICS algorithm parameter is the African Buffalo Optimization algorithm [13].

The significance of the scientific contribution of this paper is as follows:

1. A new model has been proposed for an automated student grading system based on a clustering technique in machine learning.
2. Optimal parameter tuning for the clustering algorithm's sensitive parameter.
3. A new hybrid model combines clustering and optimization algorithms.
4. Increase the depth and breadth of the literature by adding new topics and method options.

2. Related Work

The importance of automatic grading has been highlighted by many reports in the literature, where the automatic grading system was applied by using clustering algorithms, and the issues of the automatic grading system were successfully addressed with good results. Despite their capability in handling automatic text grading, the proposed algorithms cannot be directly adopted due to some limitations.

As a result, for performing efficient and fault-tolerant automatic text grading, several modifications are required. One of the most important research questions was proposed by [14], where grouping responses were involved by using the K-Means algorithm based on their similarity to the question's common words and the model answer's similarity to the student's response. The study was followed by [15], which proposed a method for scoring short responses based on semantic similarity and the related grades. In this work, the clustering approach was K-Means. In addition, in another interesting work in the literature, 148 essay answer data from 37 students were employed, and eight dummy answers (irrelevant answers) were added. Moreover, the feasibility [16] to infer grammatical and semantic linkages from both the student response and the reference answer has been also demonstrated, proving that both were correct.

As a consequence of this quality,Long Short-Term Memory (LSTM) has the ability to develop the concept of "long-term reliance." The different methods of feature selection have been also investigated and the method of choosing features has been proposed. The [16] use of semi-supervised clustering approaches for short response scoring has been also introduced, and this was supported by the findings.

The most significant contribution of this work was the use of metric pairwise constrained MPCKM (Metric Pairwise Constrained K-Means) for clustering, which is considered an extension of the K-Means algorithm and represents a breakthrough. This work, which was followed by [17], recommended the goal of a binary multi-verse optimization method to address the issue of choosing text functions in text clustering by minimizing the number of possible outcomes. The K-Means clustering algorithm considers features as entries, allowing for an evaluation of the feature selection technique to take place.

Furthermore, document preparation and feature selection are two more steps in this study's proposed model [18], including clustering algorithm phases. Clustering methods, such as K-Means and incremental K-Means were used, while a collection of unlabeled documents was allocated to predetermined groups in the unsupervised methodology. In addition to that, [19]proposed a method for short text clustering based on a multi-step process. More specifically, first, the embedding was transformed into a latent space, and then a soft clustering was performed (Smooth Inverse Frequency (S.I.F.)) by using k-Means on the latent space.

In another work [20], link-based particle swarm optimization (LBPSO) approaches for unsupervised text clustering were utilized for feature selection. The k-Means clustering method then evaluated these salient features to optimize the proposed algorithm's performance and at the same time minimize computing time. As can be ascertained from the most frequently cited related work, the K-mean approach is the most frequently utilized clustering algorithm, despite its shortcomings. Under this direction, the goal of this work is to close this gap by introducing a novel combination of clustering and optimization methods for achieving the best outcomes possible for an automated grade system.

3. Automated Grading System Based Machine Learning

To create an automatic grading form, several steps should be applied that start with collecting data from the typical answer that the teacher creates and the students' answers. These data need to be preprocessed and the text data represented with numeric vectors to be subsequently used as the dataset for the OPTICS algorithm that works with numeric data. A schematic illustration of the automated grading system is presented in Figure 2.

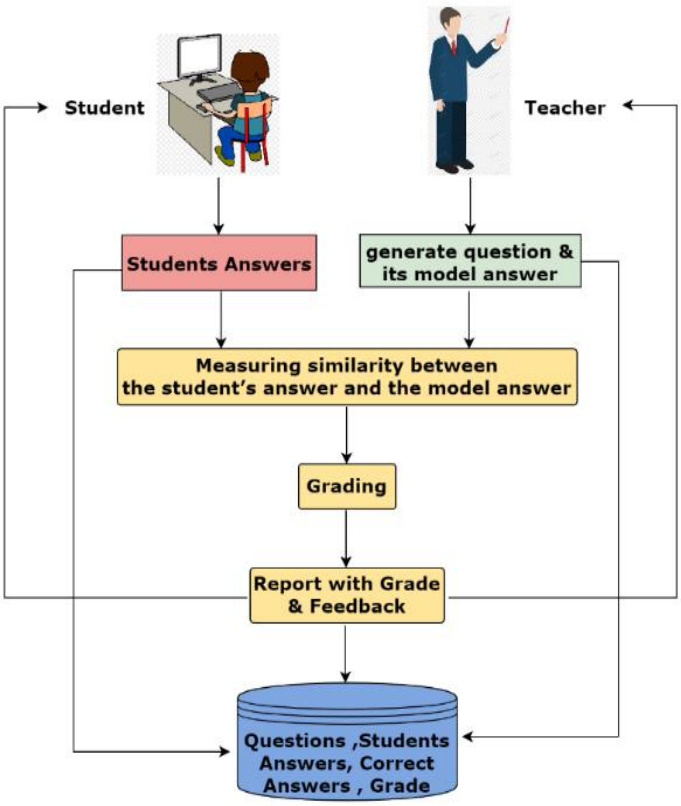

Figure 1: Depiction of the Automatic Grading System Architecture

3.1. Preprocessing and Data Representation

By using text data preprocessing techniques, text data can be cleaned up for their use in the proposed model. The preprocessing of data is regarded as a critical step in the development of a machine learning model since the outcomes are strongly dependent on how well the data are preprocessed [21]. Text data include noise in the form of emotions, punctuation, numeric value, special characters, and text in various cases, as is shown in Figure 3.

The text data processing process begins after collecting the answers from all students to all the questions in addition to the teacher's model answers.

- Tokenization: The majority of text mining research comprises words or sentences, which must be divided word by word before being further processed (Splitting the sentence into words) [22]. This is an example of tokenization:
  Before tokenization: This essay is arranged into 3 sections.
  After tokenization: [This] [essay] [is] [arranged] [into] [three] [sections].
- Stop Word Removal: In the information retrieval and text mining process, many of the most commonly used words in English are meaningless. Stop-words are common terms that contain no information and are language-specific functional words. 'The,' 'of,' 'and,' and 'to' are examples of such terms [23]. Figure 4 illustrates how the number of words is reduced when the stop words were eliminated.
- Lemmatizationis the process of breaking down complex words into their most basic components. When affixes are removed from features, the result is that the inflected

words are reduced to their stems. The stem does not have to be directly related to the word's original morphological root, although it is usually linked since words map to stems that are similar to one another. When many different kinds of features are combined into a single feature, this process is also used to minimize the number of features in the feature space and boost the clustering effect of the clustering algorithm in the question [23].

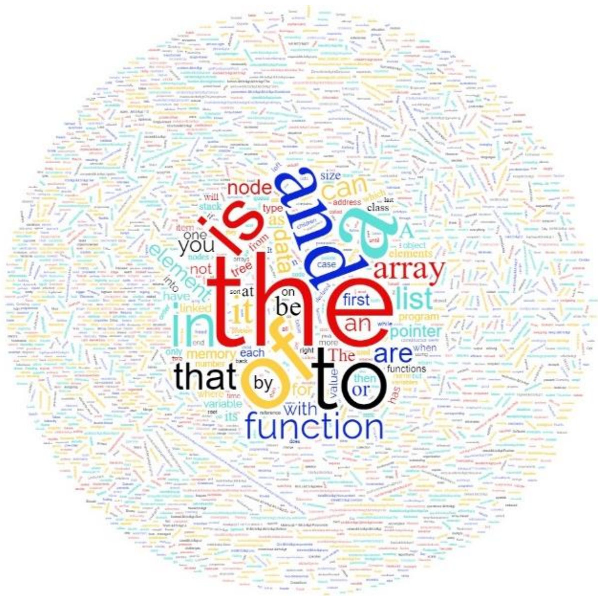

Figure 2: Depiction of the word cloud from all the answers to all 87 questions from all the students. There are a total of 1500+ words. The size of large terms indicates the higher frequency and if the words are smaller, a lower frequency is indicated.

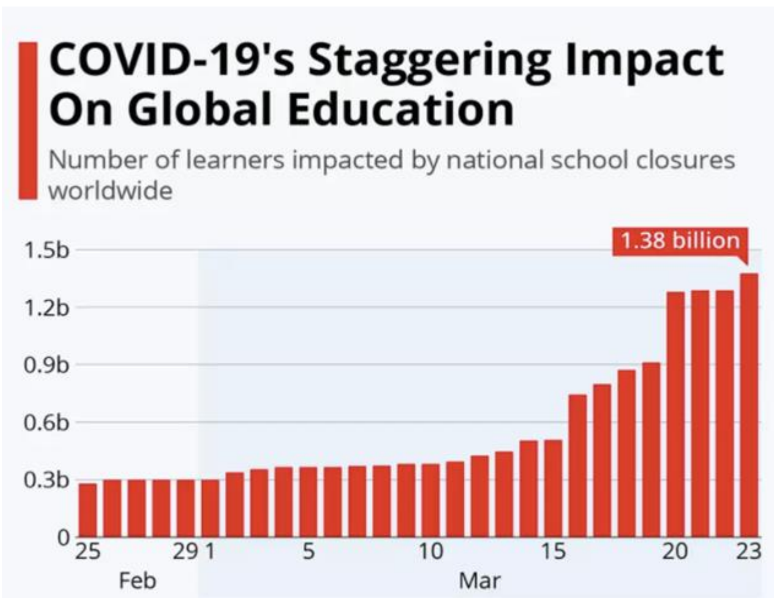

Figure 3: This graph shows the number of learners affected by the closure of national schools around the world [9]

While traditional data analysis requires the development of a model from primary data and knowledgeable opinion to establish anassociation between variables, machine learning is a promising technique that has been successfully applied by numerous researchers in various application domains, including automated grading systems[10]. The importance of ML in the automatic grading system reduces time and effort for educational institutions to achieve educational goals with high efficiency [8], and provides rich sources of information, as it is one of the characteristics of electronic evaluation of tests that are considered research motives [6][10]. Previous studies have shown that It is one of the most difficult domains to grade short responses automatically since it is strongly reliant on semantic similarity in meaning, which is defined as the degree to which two phrases are similar in meaning [13] when evaluating short answers.

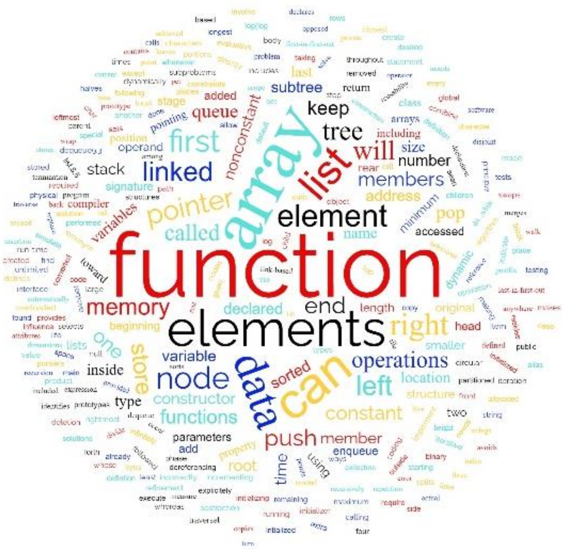

Figure 4: Depiction of the word cloud after removing stop words. This step is considered data cleaning

Figure 5: Depiction of the Vocabulary word cloud for the answers to question 1.1. The model answer to question 1.1 can be seen in Table 1.

- By using the Bag-Of-Words (B.O.W.) technique, the text has been turned into a vector of numbers. A word bag is a fundamental technique for vectorising texts. More specifically, it bypasses word order and grammatical rules in favour of constructing a two-dimensional matrix, where the first dimension represents the word and the second dimension represents the bunch of times the word appears in the corpus of texts [24].

3.2 Model Development

Before the OPTICS algorithm is applied to cluster the text data, it is reprocessed and converted into numerical data. When the OPTICS algorithm is started, the African Buffalo algorithm is utilized to determine the value of the OPTICS algorithm's important parameter (ε) that is given the alpha values. The optics algorithm will employ all of the alpha band values, whereas the accuracy of each value in the chosen alpha range will be determined. A pool of values with their associated accuracy values will be then produced and compared with the value that possesses the highest accuracy. The latter will be picked as the value for the optics algorithm's fundamental parameter.

3.2.1 OPTICS Algorithm

OPTICS, or ordering points to identify the clustering structure, The method is used in data science to cluster density-based, and common-related data. It is also used for clustering high-noise data since is a method for data mining in machine learning [25]. The optics algorithm requires the implementation of two basic parameters, namely the Epsilon ε (radius), which describes the maximum distance (is defined as the maximum search distance) to be observed and whose value is determined by using the African Buffalo algorithm, and $MinPts$, which describes the number of points required to form a cluster. The value of this parameter was taken from previous studies. From the experimental outcomes, it can be argued that $MinPts$ uses values between 10 and 20 to always get good results [26][12][27]. For each point, both the core distance and reachability distance are stored [12]. In the OPTICS algorithm, the parameters (ε) and ($MinPts$), directly define density-reachable, density-connected, cluster, and noise, as is described in [12]. Instead of recording cluster membership from the beginning, OPTICS stores the order in which items are clustered. Interestingly, the values of $core{-}distance$ and $reachability{-}distance$ endow the clusters with additional information about OPTICS ordered dataset [12].

The $\boldsymbol{core-distance}$ parameter of a sample ρ is the least value at which ρ becomes the core point. If ρ is not the core point, the core distance of ρ is indeterminate (is defined as the minimum search distance needed to include $MinPts$ cases)[12].

$$core-dist_{\varepsilon,MinPts}(\rho)=\begin{cases} UNDEFINEDif\,|N\varepsilon(\rho)|<MinPts \\ MinPts-thsmallestdistanceinN\varepsilon(\rho)otherwise \end{cases} \quad (1)$$

$N\varepsilon(\rho)$: (epsilon-neighborhood) points within a radius of εfrom a sampleρ.

The $\boldsymbol{reachability-distance}$ parameter of a point q for another point q is the larger value of the *core-distance* of ρ and the Euclidean distance midst ρ and q. Whether ρ is not the core point, the *reachability distance* between ρ and q is indeterminate (is defined as the maximum search distance). The OPTICS algorithm uses also Euclidean distance as a measure of similarity[28][12][29][30][31].

$$reachability-distance_{\varepsilon,MinPts}(\text{q},\rho)=\begin{cases} UNDEFINEDif\,|N\varepsilon(\text{q})|<MinPts \\ Max\left(core-dist_{\varepsilon,MinPts}(\text{q}),dist(\text{q},\rho)\right)otherwise \end{cases} \quad (2)$$

Algorithm1: the main function of OPTICS

Step1: Initiating the reachability distance of the given point (specified answer) for each point (answer) that has not been processed from the responses.
Step2: Getting the neighborhood of the selected point $N\varepsilon(\rho)$ according to the definitions of epsilon ε and a minimum of points ($MinPts$).
Step3: Marked ρ as handled.
Step4: Output ρ at the ordered list.
Step5: Checking if the core distance of ρ is not undefined (selected point ρ is not noise), otherwise, move to the next unprocessed point.
Step6: The priority queue is being set up in order to acquire the most nearby information, in terms of reachability distance.
Step7: Invoke the function that updates the data (this will be explained).
Step8: Repeating the process for the next closest point for each following q in Seeds.
Step9: Mark q as handled.
Step10: Output q toward the ordered list.
Step11: Invoking the update function.

An update function is invoked that changes the seed series, or in different words, the priority queue is altered in two places above (in steps 7 and 11). Updating occurs because the series is empty (at either the beginning of the algorithm or when all close points have been handled but unprocessed points remain) or when fresh points are identified to increase the cluster.
The following presents the steps of the update function [12][29]:

***Algorithm2**: the update function*

Step1: Calculating the core distance for the given point ρ.
Step2: Updating the reachability distance for each
the neighborhood of ρ for each point in $N\varepsilon(\rho)$ if (the
Point is not processed).
Step3: Checking if the reachability distance for each
Neighborhood of ρ is in seeds (it's undefined).
Step4: Insert neighborhood of ρ in the priority queue
Precisely at reachability distance.
Step5: The updating step, in other words,
continuously changes the instruction in the
priority queue, where the points with a
minor reachability distance (the points nearer to
the cluster's center points) are placed.

Clustering is attained by using the OPTICS method and sorting the datasets' entities, resulting in the so-named "reachability graphic." The reachability procedures are built by processing the dataset's entities sequentially and assigning a value called "reachability-distance" to each one. The algorithm always chooses the object that can be reached with the fewest possible steps while maintaining the lower limit denoted by MinPts, which is roughly translated to "most dense"[28].

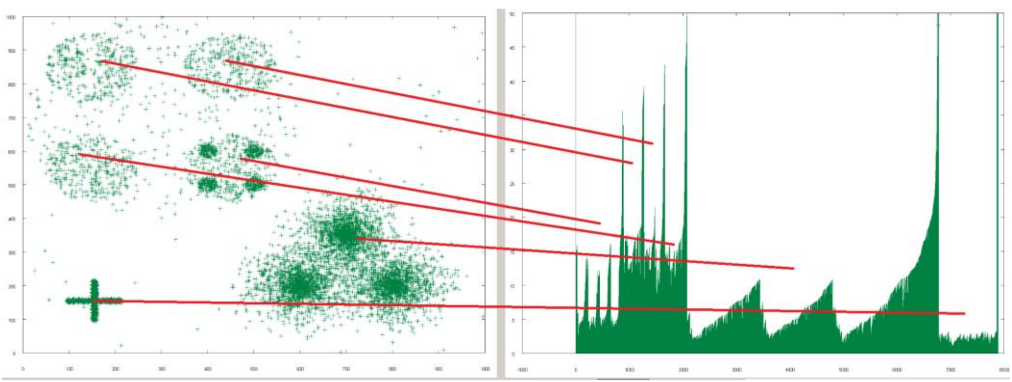

Figure 6: OPTICS reachability plot of an example dataset (right) with clusters of varying shape, size, and density (left). The red lines indicate the correspondence: clusters correlate to concave areas on the right (every concavity means a cluster)[27][28][12].

3.2.2 African Buffalo Optimization (ABO) algorithm

Inspired by the movement of African buffaloes, the African buffalo optimization search algorithm (ABO) is a freshly designed optimization search algorithm [32]. African buffaloes are massive herbivores who roam to track rich green fields to fulfill their insatiable appetite. In addition, voting is a mechanism through which African buffaloes may utilize the collective intellect of their herd, which is regarded as one of the most amazing traits of these species. In order for buffalos to move from a hungry area to a rewarding one, they must use the "waaa" cries, which request that the buffalos go on to a safer and/or more rewarding places, and the "muaa" vocalizations, which request that the buffalos come to a safe place and lush spot to graze. Overall, optimization is associated with decreasing wasteful input, boosting speed, and maximizing profitably produced goods and services [32]. The goal of this algorithm was to improve the performance of the OPTICS algorithm by enhancing the accuracy of clustering. The functionality of the proposed algorithm was also exploited with the performance of the OPTICS algorithm by controlling the parameter value epsilon(ε), which is considered one of the basic parameters of the optics algorithm, as was previously explained. The African Buffalo algorithm determines the parameter value (ε) by the alpha values rang.

The African Buffalo Optimization method essentially models the three main properties. m_k represents the ‘maaa' sound of buffalo k (k=1, 2, 3,..., n), while w_k stands for the ‘waaa' sound. The movement of the buffalos is determined mathematically by Equation (3). More specifically, the memory portion (m_{k+1}) of Equation (3) indicates that the animals are aware that they have shifted from their previous place (m_k) to a new one. This effect demonstrates their vast memory capacity, which is an important skill in their nomadic lifestyle. As far as the animals' cooperative qualities are concerned, they are represented in the second part ($bg_{max.k} - w_k$). The buffalos are considered also superb communicators, and they can track down the best buffalo in each iteration. The third element shows the animals' ability to work together ($bp_{max.k} - w_k$). Since buffalos are superb communicators, they can track down the best buffalo in each iteration.

$m_{k+1} = m_k + lp1(bg_{max.k} - w_k) - lp2(bp_{max.k} - w_k)(3)$

The movement of the buffalos throughout the alpha values range is controlled by the two major equations. Equations (3) and (4) are the ones in question (refer to algorithm3). The democratic Equation (3) serves as a theoretical framework for the animals' movement or lack thereof. Given the two competing parties, the “waaa” update allows for the accurate adjustment of the herd movement (“waaa” and “maaa” calls). As a result, the animals have a new position. The first Equation has two critical parameters, namely the global greatest (bg_{max}) and the personal greatest ($bg_{max.k}$) locations, both of which have a direct impact on the animal's decisions. On top of that, the algorithm subtracts the “waaa” element (w_k) from the maximum vector ((bg_{max} and$bp_{max.k}$) and multiplies it by the learning parameters ($lp1$,$lp2$), which their typical values lies between 0.1 and 0.6. The learning parameters 0.1 to 0.6 have proven to help achieve quick convergence thus far [33]. The aggregate of these items is then counted to the “maaa” (m_k) elements for the given dimension (instructing the animals to remain in the region to utilise it).

Then, the output from this step is sent into Equation (3), which determines whether the buffalos move or not in a given iteration.

$$w_{k+1} = \frac{w_k + m_k}{\lambda} \qquad (4)$$

λ: signifies the exploitation driver, which is chosen at random from 0 to 1.
w_{k+1} : denotes the shifting toward a current position.

***Algorithm3**: function of the African Buffalo Optimization* [34]

Step1: Objective function $f(x) \quad x = (x1.x2, \ldots, xn)^t$ Step2: Initialization: randomly place buffalos value to nodes at alpha values range. Step3: Update the buffalos' fitness values (best accuracy) Using Equation (3). Step4: Update the place of buffalo$k$-value ($bp_{max.k}$ and $bg_{max.k}$) using Equation (4). Step5: Is bg_{max}updating. Yes, go to 6. No, go to 2. Step6: Confirm that the stopping criteria have been met, go to 7; otherwise, Back to Step 3. Step7: The output is the value that gives the best accuracy (Epsilon Value =The bg_{max} value).

3.2.3 Steps to Implement the Proposed Model

The steps are shown in Figure 8.

1. Teachers should put questions with their typical answers.
2. Answers are collected from the students.
3. The Bag-Of-Words (B.O.W.) technique is used, which was explained in Part 3.1, to process the data set's texts, including questions and model answers, with students' answers.
4. With the processor text and the initial value of the basic parameterε of the algorithm, the Optics algorithm is called.
5. The value of the parameter M i nPts is determined based on previous research (values range between 10 and 20).
6. In the beginning, a random sample from the dataset is used as a selected sample (ρ).
7. The neighborhood of the selected point $N\varepsilon(\rho)$ is derived according to the definitions of epsilon ε and a minimum of points ($M\ i\,nPts$) (sets the core-distance and places the reachability-distance of sample ρ) by using Equation 3, Equation 4.
8. If sample ρ is a core point, at that time for each sample q in the E-neighborhood of ρ,

- ✓ OPTICS updates q reachability-distance as of ρ.
- ✓ Insets q into OrderSeeds if q has not yet stood processed.
- ✓ Sorts the points in the OrderSeeds in ascending order of their reachability-distance.

The iteration continues up until the input is fully depleted and OrderSeeds is empty.

9. OPTICS merely goes on to the next item in the OrderSeeds list if p is not a core point for the entity (or to the input dataset if OrderSeeds is empty).
10. Check if all points in the data set are visited.
11. To measure the accuracy of the model performance, the Mean Squared Error (M.S.E.) equation is used:

$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$(5)

where N: number of answers.

y_i :The actual grade is determined by the teacher.

$\hat{y}_i$: The predicted degree is determined by the model.

12. Check if the clustering meets the accuracy criteria (>2).
13. If the accuracy criteria are not reached, the African Buffalo optimization algorithm is called to generate an alpha values range from values to select an optimal parameter value (ε).
14. The algorithm begins by producing a buffalos' population (step1 in algorithm3). This is accomplished by assigning each buffalo (κ) in N-dimensional space a random position.
    κ:represent a particular buffalo.
15. The fitness value of each buffalo is found according to its position in the alpha values range by using Equation (3).
16. To arrive at a choice on whether to remain or continue to search for other opportunities in accordance with Equation (4).
17. It evaluates whether or not the most effective buffalobg_{max} is being modified. Therefore, if the global best fitness fulfills the exit conditions, the run is ended, and the vector's location is proposed as a solution to the issue.
18. However, if the location bg_{max} of the leading (best) the buffalo's condition does not improve after a specified number of repetitions and the termination requirements are not satisfied, the whole herd is re-established.

The employed procedure for awarding grades is based on a scale of 1 to 5, with 1 being the lowest and 5 being the greatest possible grade. Depending on their distance from the model answer, the OPTICS algorithm generates 5 clusters of comparable student responses. In a cluster, each response receives the same grade and comment. Then, a vector of the right answer is created, and its distance from each of the five clusters is determined, with all of the answers in the closest cluster receiving the highest grade and those in the farthest cluster receiving the lowest. The distance between each cluster and the model answer determines the degree. By using the Euclidean equation, a vector of the correct solution is used by the following principle: where a distance closer to 0 implies the existence of a bigger degree, and vice versa, while the degree is decreased as the distance is expanded.

4. Application and Results

4.1 Data Sets

The dataset used in this research is the same as the one used in [14].According to Table 1, this dataset contains answers from 29 students. Moreover, there were ten assignments, each consisting of a couple of questions and two exams. For instance, at the University of North Texas, an introductory computer science course was taken. The total number of individual questions was 87. In addition, the dataset contains the actual marking of the answers to 87 questions of 29 students, whereas not all students have answered all the questions on the investigation. i.e., for some questions, there were less than 29 answers. This dataset also contains the original and correct answers to every 87 questions. These 87 answers will be used as a reference to compare the student's answers. Two human judges graded the answers by using a scale of 0 (totally incorrect) to 5 (perfectly correct) (excellent answer). The data set designers selected the average grade of the two assessors as the gold standard for analysing the computerised grading task.

Table 1: Sample questions and answers for varying lengths of the data

	The question, Model Answer, and Student answers
Question:	"What is the role of a prototype program in problem-solving?
Model Answer :	To simulate the behavior of portions of the desired software product".
Student answer 1:	"High-risk problems are addressed in the prototype program to make sure that the program is feasible. A prototype may also be used to show a company that the software can be possibly programmed."
Student answer 2:	"it simulates the behavior of portions of the desired software product."
Question:	"What are the main advantages associated with object-oriented programming?
Model Answer :	Abstraction and reusability."
Student answer 1:	"They make it easier to reuse and adapt previously written code, and they separate complex programs into smaller, easier-to-understand classes."
Student answer 2:	"Reusable components, Extensibility, Maintainability, reduce large problems into smaller, more manageable problems."

4.2. Results and Discussion

A model for automatic classification and evaluation of the brief answer questions was created in this work for providing helpful comments to the students. This part evaluates the model based on the distance between an answer from the student and the model answer, as well as the accuracy in detecting similarities between the manual evaluation of grades by teachers and the automatic evaluation of the proposed model. The experimental studies are presented to show how the proposed model works with automatic recording and evaluation of responses. We have to underline that the model works well as the typical answer vocabulary can be clearly defined. The model reflects the human assessment and knows nothing about correct classification. This means that the teacher's grades can be forecast by the distance with increased accuracy. This effect indicates that the amount of words used correctly in a student's answer has a greater impact on the grade evaluation. Hence, identifying the important words in a student's response gives teachers an idea of the grades and students' misunderstandings. This method saves time recording grades and gives quick comments to students by testing the critical words used in the model answer.

To measure the model's accuracy, the M.S.E. was used to check how close the automatic evaluation is to the actual manual evaluation of teachers. When the M.S.E. value is relatively low, the closer to the actual one is expected. This is used as a typical rating scale for regression models since a lower value indicates a better fit. To ensure the correctness of obtaining the highest accuracy in the performance of the model, Equation (5) (M.S.E.) has been applied to the model by using the K-mean algorithm once, the OPTICS algorithm again, and the proposed model last. The acquired results are displayed in Table 2.

Table 2: The M.S.E. of the model when it worked with (K-means) algorithm once and with (OPTICS) again and finally with the proposed model (OPTICS + ABO)

Methods	M.S.E.
k-means	4.6152
OPTICS	**1.9081**
OPTICS + ABO	**1.6596**

When the means square error is taken into account, the model was developed by using OPTICS + ABO, where clustering fared the best outcomes among the other approaches. According to Table 2, the implementation of the OPTICS algorithm, as well as OPTICS with varying ε values, the performance is improved because the OPTICS algorithm is not sensitive to parameters, as it does not require defining the number of clusters on the basis of which the data will be clustered, but rather depends on the value of the parameter ε in the formation of clusters, and the latter is specified its best value is using the ABO algorithm.The proposed OPTICS + ABO model has a lower mean squared error than the k-means model because the algorithm is very sensitive to the value of the K parameter that determines the number of groups for which the data is clustered. This suggests that the proposed methodology provided automatic grading that was similar to the manual grading by teachers.Table 3 presents the values of the optimal epsilon (with the least mean squared error) chosen by ABO for clustering the answers to each question by using the OPTICS algorithm.

Table 3: The best epsilon values Selected by ABO to cluster the answers to each question

Questions	Best value for epsilon	Questions	Best value for epsilon
Questions 1	2.39	Questions 7	2.79
Questions 2	2.38	Questions 8	2.67
Questions 3	3.34	Questions 9	2.68
Questions 4	2.704	Questions 10	2.93
Questions 5	2.702	Questions 11	2.91
Questions 6	3.68	Questions 12	2.89

5. Conclusion

By considering that electronic learning is an imperative need, it has a huge impact on the educational activity. As a result of the rapid advancements in knowledge and technology, as well as crises and natural disasters, society's organizations and people have been forced to search for the most effective ways and methods of assisting learners in learning and to provide an interactive learning environment that meets the needs of learners in the twenty-first century. The goal is to help them develop their abilities so that they can deal with the variables of this age. One of the conditions for e-learning is the use of an automated grading system for exams and assignments. Although autonomous grading systems based on clustering and the k-Means method have been reported, this approach has a number of flaws that must be addressed. Under this perspective, by using the OPTICS and the ABO algorithms, an automated grading system based on cluster aggregation was proposed in this work, with the fundamental parameter of the system being adjusted using the ABO algorithm. Some intriguing findings were found, while the suggested model achieved a mean squared error of (1.6596), whereas the model utilizing the k-Means method achieved a significantly higher mean squared error of (4.6152). Our work paves the way for developing an automated scoring model to work on clustering questions that involve answering by using mathematical equations, diagrams,,geometric shape drawings, or geographical maps , as well as how to reduce the real time by using a computer-assisted manner.

Clustering is attained by using the OPTICS method and sorting the datasets' entities, resulting in the so-named "reachability graphic." The reachability procedures are built by processing the dataset's entities sequentially and assigning a value called "reachability-distance" to each one. The algorithm always chooses the object that can be reached with the fewest possible steps while maintaining the lower limit denoted by MinPts, which is roughly translated to "most dense"[28].

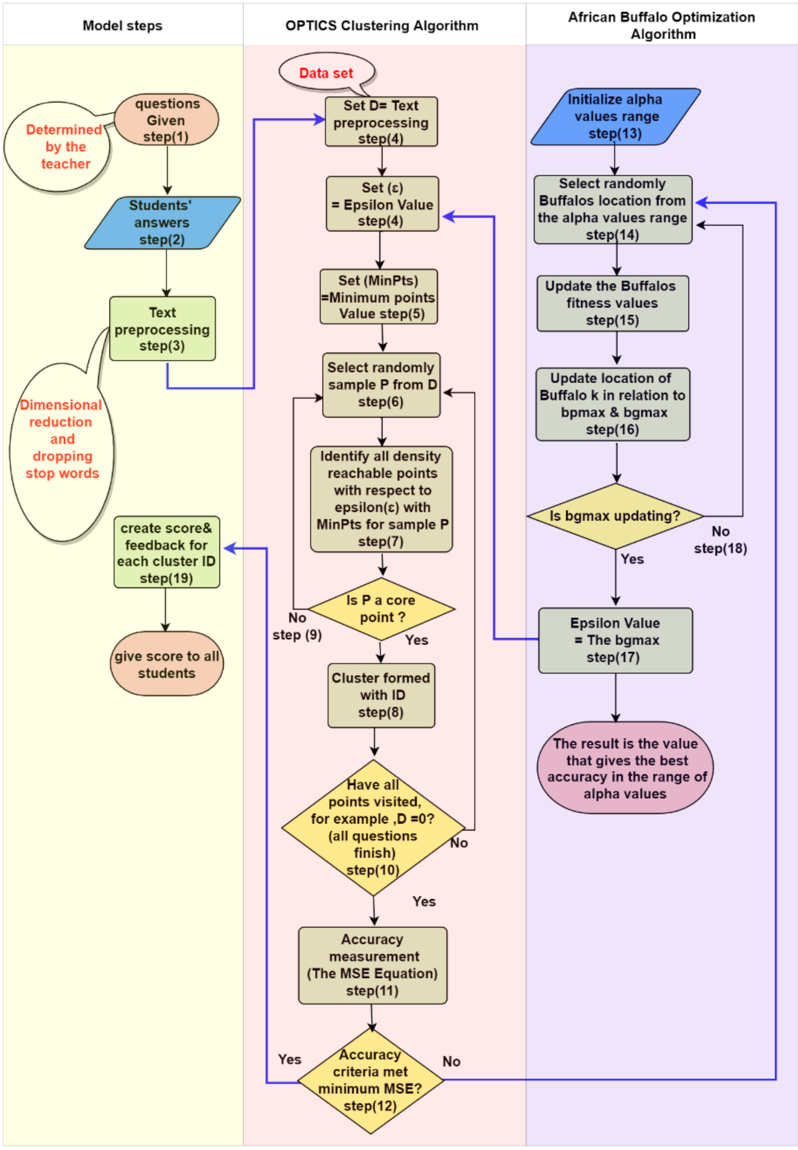

Figure 7: The proposed model

References

1. M. Mohler, R. Bunescu, and R. Mihalcea, "Learning to grade short answer questions using semantic similarity measures and dependency graph alignments," ACL-HLT 2011 - Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol., vol. 1, 2011.
2. A. Qoiriah, R. Harimurti, A. I. Nurhidayat, and Asmunin, "Application of K-Means Algorithm for Clustering Student's Computer Programming Performance in Automatic Programming Assessment Tool," 2020.

3. T. Meirovitz, S. Russak, and A. Zur, “Heliyon English as a foreign language teachers ’ perceptions regarding their pedagogical-technological knowledge and its implementation in distance learning during COVID-19,” Heliyon, vol. 8, 2022, [Online]. Available: https://doi.org/10.1016/j.heliyon.2022.e09175.
4. K. A. A. Gamage, E. K. de Silva, and N. Gunawardhana, “Online delivery and assessment during COVID-19: Safeguarding academic integrity,” Educ. Sci., 2020.
5. M. A. Zamora-Antuñano, J. Rodríguez-Reséndiz, M. A. Cruz-Pérez, H. R. Reséndíz, W. J. Paredes-García, and J. A. G. Díaz, “Teachers’ perception in selecting virtual learning platforms: A case of mexican higher education during the COVID-19 crisis,” Sustain., vol. 14, 2022.
6. H. Ilgaz, “Providing online exams for online learners : Does it really matter for them ?,” 2020.
7. S. Kisanjara, “Factors Influencing E-Learning Implementation in Tanzanian Universities,” Online J. Distance Educ. e-Learning, 2020.
8. V. Kumar and D. Boulanger, “Explainable Automated Essay Scoring: Deep Learning Really Has Pedagogical Value,” Front. Educ., 2020.
9. UNESCO’s, “COVID-19 Staggering impact on clobal education,” https://www.statista.com, 2020.
10. V. David and U. M. Enrekang, “The Interpretable Machine Learning among Students and Lectures during the COVID-19 Pandemic,” J. Iqra’ ..., vol. 6, 2021, [Online]. Available: https://journal.iaimnumetrolampung.ac.id/index.php/ji/article/view/1312.
11. And X. S. Wang, Juntao, “An improved K-means Clustering Algorithm,” 2011.
12. M. Ankerst, M. M. Breunig, H. P. Kriegel, and J. Sander, “OPTICS: Ordering Points to Identify the Clustering Structure,” SIGMOD Rec. (ACM Spec. Interes. Gr. Manag. Data), vol. 28, 1999.
13. M. A. K. Azrag, T. A. Abdul Kadir, J. B. Odili, and M. H. A Essam, “a Global African Buffalo Optimization,” Int. J. Softw. Eng. Comput. Syst., vol. 3, 2017.
14. N. Süzen, A. N. Gorban, J. Levesley, and E. M. Mirkes, “Automatic short answer grading and feedback using text mining methods,” Procedia Comput. Sci., vol. 169, no. 2019, 2020, [Online]. Available: https://doi.org/10.1016/j.procs.2020.02.171.
15. H. Sadr and M. N. Solimandarabi, “Presentation of an Efficient Automatic Short Answer Grading Model Based on Combination of Pseudo Relevance Feedback and Semantic Relatedness Measures,” J. Adv. Comput. Res. 10(2), 17-30, vol. 10, 2019.
16. T. S. Walia, G. S. Josan, and A. Singh, “An efficient automated answer scoring system for Punjabi language,” Egypt. Informatics J., vol. 20, 2019, [Online]. Available: https://doi.org/10.1016/j.eij.2018.11.001.
17. A. Horbach and M. Pinkal, “Semi-supervised clustering for short answer scoring,” Lr. 2018 - 11th Int. Conf. Lang. Resour. Eval., 2019.
18. A. K. Sangaiah, A. E. Fakhry, M. Abdel-Basset, and I. El-henawy, “Arabic text clustering using improved clustering algorithms with dimensionality reduction,” Cluster Comput., vol. 22, 2019, [Online]. Available: https://doi.org/10.1007/s10586-018-2084-4.

19. A. Hadifar, L. Sterckx, T. Demeester, and C. Develder, "A self-training approach for short text clustering," ACL 2019 - 4th Work. Represent. Learn. NLP, RepL4NLP 2019 - Proc. Work., 2019.
20. N. Kushwaha and M. Pant, "Link based BPSO for feature selection in big data text clustering," Futur. Gener. Comput. Syst., vol. 82, 2018, [Online]. Available: https://doi.org/10.1016/j.future.2017.12.005.
21. A. I. Kadhim, Y. N. Cheah, and N. H. Ahamed, "Text Document Preprocessing and Dimension Reduction Techniques for Text Document Clustering," Proc. - 2014 4th Int. Conf. Artif. Intell. with Appl. Eng. Technol. ICAIET 2014, 2015.
22. S. Qaiser and R. Ali, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents," Int. J. Comput. Appl., vol. 181, 2018.
23. J. Leveling, "On the effect of stopword removal for SMS-based FAQ retrieval," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, vol. 7337 LNCS, doi: 10.1007/978-3-642-31178-9_12.
24. D. M. El-Din, "Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis," Int. J. Adv. Comput. Sci. Appl., vol. 7, 2016.
25. V. Lytvynenko, I. Lurie, J. Krejci, M. Voronenko, N. Savina, and M. A. Taif, "Two step density-based object-inductive clustering algorithm," CEUR Workshop Proc., vol. 2386, 2019.
26. B. T. S. V. Sharma, B. T. S. S., Ahmad, B. T. S. K., & Singh, "Clustering Approaches to Offer Business Insights," Int. J. Comput. Appl. 181(1), 25-29, vol. 7, 2021.
27. G. Redinger and M. Hunner, "Visualization of the Optics Algorithm," Univ. Wien, 2017.
28. A. Ali et al., "Systematic Review: A State of Art ML Based Clustering Algorithms for Data Mining," Proc. - 2020 23rd IEEE Int. Multi-Topic Conf. INMIC 2020, 2020.
29. N. Verma, M., Srivastava, M., Chack, N., Diswar, A. K., & Gupta, "A Comparative Study of Various Clustering Algorithms in Data Mining," Int. J. Eng. Res. Appl., vol. 2, 2012.
30. G. Iván and V. Grolmusz, "On dimension reduction of clustering results in structural bioinformatics," Biochim. Biophys. Acta - Proteins Proteomics, vol. 1844, 2014.
31. R. J. G. B. Campello, P. Kröger, J. Sander, and A. Zimek, "Density-based clustering," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 10, 2020.
32. G. Wang et al., "Reducing Annotation Efforts in Supervised Short Answer Scoring," Proc. Tenth Work. Innov. Use NLP Build. Educ. Appl. (pp. 124-132), 2015.
33. H. Jafari, M. Nazari, and S. Shamshirband, "Optimization of energy consumption in wireless sensor networks using density-based clustering algorithm," Int. J. Comput. Appl., vol. 43, 2021, [Online]. Available: https://doi.org/1206212X.2018.1497117.
34. N. K. Amandeep Kaur Mann, "Review Paper on Clustering Techniques," Glob. J. Comput. Sci. Technol., vol. 13, 2013, [Online]. Available: http://globaljournals.org/papers/file.php?file=3417.

35. M. M. A. Patwary, D. Palsetia, A. Agrawal, W. K. Liao, F. Manne, and A. Choudhary, "Scalable parallel OPTICS data clustering using graph algorithmic techniques," Int. Conf. High Perform. Comput. Networking, Storage Anal. SC, 2013.
36. O. J.B, K. M.N.M, A. S, and A. M, "Tutorials on African Buffalo Optimization for Solving the Travelling Salesman Problem," Int. J. Softw. Eng. Comput. Syst., vol. 3, 2017.
37. J. B. Odili, A. Noraziah, and M. Zarina, "A Comparative Performance Analysis of Computational Intelligence Techniques to Solve the Asymmetric Travelling Salesman Problem," Comput. Intell. Neurosci., vol. 2021, 2021.
38. A. E. Babalola, B. A. Ojokoh, and J. B. Odili, "A Review of Population-Based Optimization Algorithms," 2020 Int. Conf. Math. Comput. Eng. Comput. Sci. ICMCECS 2020, 2020.
39. J. B. Odili, M. N. M. Kahar, and S. Anwar, "African Buffalo Optimization: A Swarm-Intelligence Technique," Procedia Comput. Sci., vol. 76, 2015.
40. J. B. Odili, M. N. M. Kahar, S. Anwar, and M. A. K. Azrag, "A comparative study of African Buffalo Optimization and Randomized Insertion Algorithm for asymmetric Travelling Salesman's Problem," 2015 4th Int. Conf. Softw. Eng. Comput. Syst. ICSECS 2015 Virtuous Softw. Solut. Big Data, 2015.
41. P. Branchet et al., "Machine learning applications for building structural design and performance assessment: State-of-the-art review," J. Build. Eng., vol. 33, 2021.
42. Y. Mahmood, N. Kama, A. Azmi, A. S. Khan, and M. Ali, "Software effort estimation accuracy prediction of machine learning techniques: A systematic performance evaluation," Softw. - Pract. Exp., vol. 52, 2022.